

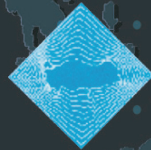
İSTKON



# ULUSLARARASI İSTATİSTİK KONGRESİ

27-30 Ekim 2013

Vogue Hotel Avantgarde  
Kemer - Antalya



TÜRK İSTATİSTİK  
DERNEĞİ  
[www.turkistatistik.org](http://www.turkistatistik.org)

BİLDİRİ KİTABI

## 8. ULUSLARARASI İSTATİSTİK KONGRESİ

27-30 EKİM 2013

## İÇİNDEKİLER

DÜZENLEME KURULU	iv
DANIŞMA KURULU	v
TÜRK İSTATİSTİK DERNEĞİ YÖNETİM KURULU	vi
BİLİMSEL PROGRAM	vii
BİLDİRİ OTURUMLARI 1	1
BİLDİRİ OTURUMLARI 2	77
BİLDİRİ OTURUMLARI 3	128
BİLDİRİ OTURUMLARI 4	182
BİLDİRİ OTURUMLARI 5	258
POSTER BİLDİRİLER	318

## DÜZENLEME KURULU

Ayşen APAYDIN (Başkan)	Türk İstatistik Derneği
Çağdaş Hakan ALADAĞ	Türk İstatistik Derneği
Furkan BAŞER	Türk İstatistik Derneği
Tolga BERBER	Karadeniz Teknik Üniversitesi
Türkan ERBAY DALKILIÇ	Karadeniz Teknik Üniversitesi
Süzülay HAZAR	Türk İstatistik Derneği
Tülay KESEMEN	Karadeniz Teknik Üniversitesi
Orhan KESEMEN	Karadeniz Teknik Üniversitesi
A. Sevtap KESTEL	Türk İstatistik Derneği
Zafer KÜÇÜK	Karadeniz Teknik Üniversitesi
Halil İbrahim ŞAHİN	Karadeniz Teknik Üniversitesi
Uğur ŞEVİK	Karadeniz Teknik Üniversitesi

#### DANIŞMA KURULU

Süleyman DÜNDAR	AFYON KOCATEPE ÜNİVERSİTESİ
Memmedağa MEMMEDLİ	ANADOLU ÜNİVERSİTESİ
Fahrettin ARSLAN	ANKARA ÜNİVERSİTESİ
İsmail ERDEM	BAŞKENT ÜNİVERSİTESİ
Selahattin KAÇIRANLAR	ÇUKUROVA ÜNİVERSİTESİ
Cengiz ÇELİKOĞLU	DOKUZ EYLÜL ÜNİVERSİTESİ
Şanslı ŞENOL	EGE ÜNİVERSİTESİ
Zeki YILDIZ	ESKİŞEHİR OSMANGAZİ ÜNİVERSİTESİ
Sinan ÇALIK	FIRAT ÜNİVERSİTESİ
Hülya BAYRAK	GAZİ ÜNİVERSİTESİ
Ufuk YOLCU	GİRESUN ÜNİVERSİTESİ
Hülya ÇINGİ	HACETTEPE ÜNİVERSİTESİ
Adnan MAZMANOĞLU	İSTANBUL AYDIN ÜNİVERSİTESİ
Esra AKDENİZ DURAN	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Münevver TURANLI	İSTANBUL TİCARET ÜNİVERSİTESİ
Türkan ERBAY DALKILIÇ	KARADENİZ TEKNİK ÜNİVERSİTESİ
Sevgi YURT ÖNCEL	KIRIKKALE ÜNİVERSİTESİ
Müjgan TEZ	MARMARA ÜNİVERSİTESİ
Nalan CİNEMRE	MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
Dursun AYDIN	MUĞLA SITKI KOÇMAN ÜNİVERSİTESİ
İnci BATMAZ	ORTA DOĞU TEKNİK ÜNİVERSİTESİ
Faruk ALPASLAN	ONDOKUZ MAYIS ÜNİVERSİTESİ
Aşır GENÇ	SELÇUK ÜNİVERSİTESİ
Emel ÇANKAYA	SİNOP ÜNİVERSİTESİ
Cenap ERDEMİR	UFUK ÜNİVERSİTESİ
Şaban EREN	YAŞAR ÜNİVERSİTESİ
Ali Hakan BÜYÜKLÜ	YILDIZ TEKNİK ÜNİVERSİTESİ

#### TÜRK İSTATİSTİK DERNEĞİ YÖNETİM KURULU

Ayşen APAYDIN	Başkan
Sevtap KESTEL	Başkan Yardımcısı
Furkan BAŞER	Genel Sekreter
İsmet TEMEL	Genel Sayman
Süzülay HAZAR	Üye
Çağdaş Hakan ALADAĞ	Üye
Şerafettin DEMİRSOY	Üye
Banu YAĞCI	Üye
Alişan CANSEVER	Üye

#### DESTEKLEYEN KURULUŞLAR



**Uluslararası 8. İstatistik Kongresi,  
27-30 EKİM 2013 Kemer-ANTALYA**

**BİLİMSSEL PROGRAM**

**28 EKİM / OCTOBER 2013 PAZARTESİ / MONDAY**

Salon / Hall	SALON 1	SALON 2	SALON 3	SALON 4	SALON 5	SALON 6
08:30 - 09:30	KAYIT - REGISTRATION					
09:30 - 10:30	<b>AÇILIŞ - OPENING</b> Doç. Dr. Türkhan Erbay Bahliç, Karadeniz Teknik Üniversitesi, Ronald Wasserstein, Head of American Statistical Association (Video Bağlantısı), Prof. Dr. Ayşen Ayyıldız, Türk İstatistik Derneği Başkanı					
10:30 - 11:30	<b>DAVETLİ - INVITED SPEAKER</b> Oturum Başkanı - Chair Person: Orhan GÜVENEN Konuşmacı - Speaker: Prof. Dr. David Wikie - U.K					
11:30 - 12:30	<b>DAVETLİ - INVITED SPEAKER</b> Oturum Başkanı - Chair Person: Fethi YILDIZ Konuşmacı - Speaker: Prof. Dr. Orhan Güvenen					
12:30 - 14:00	YEMEK ARASI - LUNCH					
Salon / Hall	SALON 1	SALON 2	SALON 3	SALON 4	SALON 5	SALON 6
<b>BİLDİRİ OTURUMLARI 1 SESSION 1</b>	<b>İSTATİSTİK TEORİSİ 1</b> Oturum Başkanı - Chair Person: Sınan ÇALIK	<b>UYGULAMALI İSTATİSTİK 1</b> Oturum Başkanı - Chair Person: Adnan MAZMANOĞLU	<b>FINANS, SİGORTACILIK, RİSK YÖNETİMİ</b> Oturum Başkanı - Chair Person: Murat BÜYÜKYZIÇCI	<b>ZAMAN SERİLERİ 1</b> Oturum Başkanı - Chair Person: Nevin GÜLER	<b>STATİSTİK METHODS</b> Oturum Başkanı - Chair Person: Şişe ŞAHİN	<b>OLASILIK VE STOKASTİK SÜREÇLER</b> Oturum Başkanı - Chair Person: Fethi Yıldırım
14:00 - 14:15	Abdullah YALÇINKAYA, Mehmet Niyazi ÇANKAYA, Ömer ALTINDAĞ, Yetkin TÜTAÇ Hipotez Testlerinde Normallik Varsayımında Uygulanabilirliği Üzerine Bir Çalışma	Adnan MAZMANOĞLU, Ayşen SAGBAS, Seda MAZMANOĞLU Sağlık Çalışmalarında Karşılaştıkları Sorunları Çözümlenmesi İçin Faktör Analizi Üzerine Bir Çalışma	Ayşe İŞİ, Hakkı POLAT, Fatih CEMREK Bireysel Emeklilik Sisteminde Devlet Katkıların Etkisinin Yapsal Kırılma Analizi ile İncelenmesi	Ayşe FENMEZCİ, Nevin GÜLER Sağlık Bilimsel Ç-Regresyon Model ve Otoregresif Modellerin Öngörüsüne Uygulanması Simülasyon Çalışması	Derya TÜRKAN, Çağdaş Hakan ALADAĞ, Ozgur YENİAY A Genetic Algorithm For Portfolio Optimization	Anah KIZAN, Halil TAMİL Bütüncül Risk Değerleri
14:15 - 14:30	Safin GÜNDÜZ, Ali İhsan GENÇ İki Üçgenli Rastgele Değişkenin Bölümünün Dağılımı	Alper SİNAN, Ayşe BENC Medyana dayalı parametrik olmayan tahmini yöntemi ile polinomial modelin tahmini	Tuğba SÖNÜT AÇAR, M. Revan ÖZALE Otokorelasyon ve Çoklu İlişkili Veri İçerisinde Etkinlik Ölçümleri	Emre KOCACAK, Erol EĞRİOĞLU Yapay Sinir Ağlarına Dayalı Yüksek Dereceden Bulank Arma Zaman Serisi Öngörü Yönteminin Bir Uygulanması	Emre DÜNDER, Mehmet Ali GENİZ Comparison of Discretization Methods For Learning Bayesian Networks	Cisam ÖCAL, Tülay KESEMEN, Zafer KÜÇÜK, Zafer KHANİYEV Genelleştirilmiş Beta Müdahaleli Yarı-Markov Baskıya Yünlüye Süreci İçin Asimptotik Sonuçlar
14:30 - 14:45	Aladdin ŞAMILDOĞU, Çiğdem GİRİFTİNOĞLU, Şevda ÖZDEMİR Maksimum Entropi ve Maximum Entropi Yöntemleri ile Sağlık Veri Analizi	Furkan BAŞER, Ayşen APAYDIN Sınırlı Veri Üzerinde Destek Vektör Makinelerinin Lojistik Regresyon ile Karşılaştırılması	Nurbanur RİHSA, Hüseyin TATLIÖL Türkiye Kredi Temerrüt Takası Piyasasının Uzun Dönemli Bağımlılık Yaptığını İncelenmesi ve Modellenmesi	Nevin GÜLER, Özlem TÜRKŞEN Bulank Kümeleme Dayalı Yerel Regresyon ve Otoregresif Modellerin Öngörüsüne Uygulanması	Nihal PERİN ALAKOÇ, Ayşen APAYDIN Fuzzy Linear Programming: An Application	Halil TAMİL İkinci Rapor Değerleri
14:45 - 15:00	Ayşe TURAN BUĞATEKİN, Sınan ÇALIK Kesikli Dağılım İçerisindeki Sıra İstatistiklerinin Moment Çıkarım Fonksiyonu	Birel TOPÇU, Sınan ÇALIK, Serap ALPASOY, Aydın AKYÜZ, Mehmet YILMAZ, Cemal GAZELİOĞLU Yapısal Etkilik Model Analizi: Prehertansiyel Bireylerde, Antropometrik Biyometrik ve Etilokardiyografik Parametrelerin Üzerine Bir Uygulanması	Sibel AKIN KEMALOĞLU, Emel KUZULU Öner Risk Ölçümüne Dayalı Portfolyo Optimizasyonu: Dinamik Kopulu Modeli	Fatma Esin KURT, Yusuf DEMİR Piyasa Risklerinin Belirlenmesinde Risk Maruz Değer Yöntemi ve Borsa İstanbul'da Bir Uygulanması	Adeyye OLATAN, Ayşe TANSU TUNÇÖLEK An Application of Shock Models to The Inventory	Yedat SAĞLAM, Murat SAĞIR, Erdiç YÜCESOY, Mügehan ZORLU, Tolga ZAMAN Ağır Jackson Modelinde Momentlerin Bulunması
15:00 - 15:15	Mesrin GÜLER Genel Parçalanmış Lineer Model Altında Bazı Kabul Edilebilir Tahmin Ediciler	Tolga ZAMAN, Yedat SAĞLAM, Erdiç YÜCESOY, Murat SAĞIR Fikrî Varlıkların Değerlendirilme Üzerine Bir Çalışma	Ceren Eda CAN, Gül ERGİN, Ahmet YALINIZ Saklı Markov Modeli Kullanılarak Türkiye'de Borçluların Risk Değerlendirilmesi	Özge GÜNDÜZOĞLU, Erol EĞRİOĞLU, Çağdaş Hakan ALADAĞ, Uluk YOLCU Güçsü Aktivasyon Fonksiyonuna Dayalı Çarpımsal Neuron Modeli Üzerine Bir Çalışma	Erkan Abdülgaffar GADİOĞLU The Policy Determinants of the Regional Disparities in the Higher Education in Mathematical Sciences: A Multivariate Case Study of Lessons to be Drawn From the Privatization Efforts of the Turkish Higher Education System With Special Reference to Statistical Sciences	Senol DEMİR, İhsan DÜNER, Zafer KÜÇÜK, Tülay KESEMEN Geçmiş ve Pareto Müdahaleli Çözümlü Yöntem Seçimi, Sınırlı Fonksiyonlarının Momentleri İçin Asimptotik Açılımlar
15:15 - 15:30	Özge KURBAN, M. Revan ÖZALE Lineer Karar Modellerinde Etki Analizinin Uygulanması	Özge PASTIR, Handan ANARALLI, Şengül CANŞİR Dünya Üzerinde Tıbbi Gözetim Araştırma Hastanesinde Tedavi Gören Hastaların Spondiloid Hastalılarının Fesetli Kemiklerin İncelenmesi	Zeynep İLHAN, Nergis YILMAZ Türkiye'deki Sporcuların Sermaye Yapı Seçiminin Mimik Modeli Kullanılarak Analizi	Hakkı POLAT, Fatih CEMREK, Ayşe İŞİ Kopulu Varyans Modellerinin Yıllık Doğal Gaz Fiyatlarına Uygulanması	Sule SAHİN, Banül MILLEK Some Aspects of Simulation with Stochastic Investment Models	Ali HERT, Özgün AYDIN İstatistiksel Dağılım Parametreleri Tahminlerinin Etkin Global Optimizasyon (Ego) ile Ede Edilmesi
15:30 - 15:45	ÇAY/KAHVE ARASI - COFFEE BREAK					
15:45 - 17:15	<b>PANEL - İSTATİSTİK BÖLÜMLERİNİN BÜYÜKLÜĞÜ VE GELECEĞİ</b>					
17:15 - 18:30	<b>AKADEMİK TOPLANTI</b>					

**29 EKİM / OCTOBER 2013 SALI / TUESDAY**

Salon / Hall	SALON 1	SALON 2	SALON 3	SALON 4	SALON 5	SALON 6
<b>BİLDİRİ OTURUMLARI 2 SESSION 2</b>	<b>UYGULAMALI İSTATİSTİK 2</b> Oturum Başkanı - Chair Person: Jale BALİBOYOĞLU	<b>ZAMAN SERİLERİ 2</b> Oturum Başkanı - Chair Person: Gülşah KIROĞLU BAŞARIR	<b>VERİ MADENCİLİĞİ</b> Oturum Başkanı - Chair Person: Türkmen ERBAY DALIKLIÇ	<b>STATISTICS THEORY 1</b> Oturum Başkanı - Chair Person: Çağdaş Hakan ALADAĞ	<b>ÇOK DEĞİŞKENLİ İSTATİSTİK</b> Oturum Başkanı - Chair Person: Sınan ÇALIK	<b>EGİTİM</b>
08:30 - 08:45	Ergül DEMİR Çoklu Seçimli Testlerde Kayıtlı Veri Madenciliğinin İncelenmesi: CBS Örneği	Ebrucan İSLAMOĞLU, Faruk ALPASLAN Anlık Değerli Zaman Serileri Öngörü Yöntemlerinin Karşılaştırılması	Özkan ASLAN, Nevin GÜLER Kümeleme Dayalı İstatistiksel Filtreleme - Karşılaştırmalı Bir Çalışma	Burcu ÜÇER, Tuğba YILDIZ Fisher Information in Progressive Type II Censored Order Statistics and Their Concomitants for Some Copula Models	Muğni OKURSOY, Nurgün YÜKSEL Açıklayıcı Faktör Analizinde Faktör Çıkarma Metodunun Seçimine İlişkin Bir Simülasyon Çalışması	
08:45 - 09:00	Gökçen EFENDİOĞLU, Nuri ÇELİK Medyan Parçalanma Tekniki ve Uygulanması	Gökçen YAPAR, Ramile TAŞLILAN Değerlendirme Üzerinde Düzeltme Yöntemi	Selver Ezi YALINIZ, Hasan GÜZEL Bankacılık Pazarlama Sonuç Tahmini İçin Bir Örnekleme Sınırlama Uygulanması	Selma TÖRKER, Gülşen ÜSTÜNDOĞU ŞIRAY A New Restricted Estimator	B. Barış ALKAN, Cemal ATAKAN Manova Varyanslarının Sınırlanmış Durumunda Uzaklık Analizi Yönteminin Kullanımı	
09:00 - 09:15	Jale BALİBOYOĞLU Varyans Analizi Modellerinde Bağımlılık Varsayımının Etkisi	Hakan BOZDAĞ, Hakan DEMİRİRİ Havaların Yüklü Tahminin Bayes Yöntemi ile Öngörülmesi: Antalya Havalarının Örneği	Volkan UZUN, Emin YILDIZTEPE LSA ve PLSA ile Türkiye Meteoroloji Anomalizasyonunun Ölçülmesi	Emre DÜNDER, Mehmet Ali GENİZ A New Approach for Learning Structure in Hybrid Bayesian Networks	Gülhayat GÖLBAŞI ŞİMŞEK, Fatma NİYAN, Selahattin AKRODUR Kararlı Korelasyon Analizinde Bayut Sayısına Karar Verilmesinde Paralel Analiz Yaklaşımı	
09:15 - 09:30	Nergiz KASIMBEYLİ Bir Bayyut Kesme Ve Ana Matris Seçim Problemi İçin İki Amaçlı Tam Sayılı Programlama Modeli	Selim DÜNMEZ Zaman Serileri Analizinde Yeni Bir Yapay Sinir Ağları Çalışması ve Karşılaştırmalı Analiz	Yasin KAYA EK'deki Normal ve Erken Katkıncı Kesilme Durumlarının Sınırlanması	Özlem TÜRKŞEN Uncertainty Assessment of The Response Surface Model Parameters with Bayesian and Fuzzy Approaches	Siddik ARSLAN, Fikri ÖZTÜRK, Salih ÇELEBİOĞLU Çok Değişkenli Regresyon Kapalılarında Üretici Fonksiyonun Aile Parametresinin Tahmini Edilmesi	SÜZÜLAY HAZAR Saha Uygulanması ve Veri Toplama Teknikleri Ayşen APAYDIN İstatistiksel Modeller ve Araştırmalarda Doğru Bilinen Yanlışlar
09:30 - 09:45	Erkan ARI, Zeki YILDIZ Sıralı Lojistik Regresyonda Orantısal Oran Modeli İçin Paralel Doğrular Varsayımı		Tolga BERBER, Uğur SEVİK, Adil ALPOÇAK, Pınar BALCI Göçül Kitlelerinin Yoğunluklarının Marmara Bölgesindeki İstatistiksel Olarak Elde Edilmesi	Ayşe TANSU TUNÇÖLEK Optimal Replacement Model for Systems with Single Spare and Lead Time	Sınan ÇALIK Çok Değişkenli Sıra İstatistikleri Üzerine	
09:45 - 10:00	ÇAY/KAHVE ARASI - COFFEE BREAK					
Salon / Hall	SALON 1	SALON 2	SALON 3	SALON 4	SALON 5	SALON 6
<b>BİLDİRİ OTURUMLARI 3 SESSION 3</b>	<b>BULANK TEDİRİ</b> Oturum Başkanı - Chair Person: Kamile ŞANLI KÜLA	<b>EKONOMETRİ</b> Oturum Başkanı - Chair Person: Serpil TÜRKÜYLMAZ	<b>STATISTICS THEORY 2</b> Oturum Başkanı - Chair Person: Olcay ARSLAN	<b>APPLIED STATISTICS 1</b> Oturum Başkanı - Chair Person: Bülent GENÇÖLÜ	<b>TIME SERIES</b> Oturum Başkanı - Chair Person: Erol EĞRİOĞLU	<b>EGİTİM</b>
10:00 - 10:15	Ali Zafer DALAR, Erol EĞRİOĞLU, Çağdaş Hakan ALADAĞ, İsmail Burhan TÜRKŞEN Top-1 Bulank Fonksiyon Yaklaşımının Zaman Serilerine Bir Uygulanması	Cengiz ARKAN, Onur YONA Türkiye Ekonomisinin Bazı Göstergelerinin Üzerine Ekonometrik Çözümler ve Öngörü	Fatma Zehra DOĞRU, Olcay ARSLAN Robust Estimators for The Shape Parameters of Burr XII Distribution	Osman Nuri ERDEM, Sevilap KESTEL The Impact of Replacement Migration Policy On Demographic Indicators And Labor Force In Turkey	Eren BAŞ, Vedide Rezan USLU, Uluk YOLCU, Erol EĞRİOĞLU A Fuzzy Time Series Approach Based On Differential Evolution Algorithm Using DeBesset Mutation Strategy	
10:15 - 10:30	EiH Burcu DİLLEN, Burhan TÜRKŞEN Bulank Sistem Modelleme	Serpil TÜRKÜYLMAZ, Mesut BALİBEY Borsa Endeksi Getirilerinde İkili Uzun Hafıza Özelliğinin Analizi İçin Bir Uygulanması	Gülner KEMALBAY On Marginal Distribution of Conditional Largest Order Statistics From Bivariate Sample	Erkan İŞIKLI, Nezir AYDIN, Bekir AYTAZ Inverse Forecasting in Reverse Logistics: The Case Of Electrical and Electronic Equipment	Erol EĞRİOĞLU, Çağdaş Hakan ALADAĞ, Uluk YOLCU A New Adaptive Network Fuzzy Inference System For Time Series Forecasting	Furkan BAŞER Bazı Parametrik Testler



29 EKİM / OCTOBER 2013 SALI / TUESDAY

Salon / Hall	SALON 1	SALON 2	SALON 3	SALON 4	SALON 5	SALON 6
10:30 - 10:45	M. Bahar BAŞKUR, Burhan TÜRKŞEN Kapa Kasası İmlatının Full Tip-2 Bulancık Yaklaşımı Modellemesi	Adnan KARABİBRAHİMOĞLU, Yasin ASAR, Hakan BAŞBOZKURT, Asır GENÇ Türkiye'nin Gayri Safi Yurtiçi Hissesinin (GSY) İstatistiksel Olarak Modellemesinde Regülasyon Parametrelerinin Uygulanması	Deniz TOPIUZ, Ümit UŞAN Negative Cumulative Dependence	Harika ERDOĞDU, Mehmet Ali CENGİZ, Süleyman DÜNDAR Determining The Effect of Socio-Economic Characteristics on the Net Migration Rate Towards the Provinces in Turkey Using Bayesian Model Averaging	Özge ÇAĞAÇ YOLCU, Feriit KAPLAN A Hybrid Fuzzy Time Series Approach For Forecasting Exchange Market Of Istanbul	
10:45 - 11:00	Arçay ÇAY ATALAY, Etiler MUBİN DOĞAN, Ahmet ATILALY Türkiye'de Sosyomatematik ve Sosyal Ekonomik Gelişimi Göre Herin Kümesi	Buse BÜYÜK, Mehmet YILMAZ Türkiye Talebine Ekonomik Bir Yaklaşım - Türkiye Örneği	Sibel ALADAĞ, Hülya CİNGİ Median In Simple Random Sampling	Gülçay EDİMLİ İNAN, Ayşen APAYDIN Robust Quadratic Hedging Problem: In Incomplete Markets	Ufuk YOLCU, Özge ÇAĞAÇ YOLCU, Feriit KAPLAN, Cengiz HAKAN ALADAĞ An Improved High Order Fuzzy Time Series Approach	
11:00 - 11:15	Emre DEMİR, Nevil DEMİR Bulanık Veri Zarfama Analizi Yöntemi ile Çorum İlindeki Liselerin Etimolojik Ölçümü	Feyza GÜNNAY, Abdullah YALÇINKAYA, Mustafa GAZİOĞULLARI Türkiye'de Ekonomik Değişikliklerin Yüceltilmesi Güven Endeksi Üzerindeki Etkisi	Yelken TUĞ, Özgür ARSLAN, Mehmet Niyazi ÇANKAYA Robust Estimators for the Distributed Lag Model Based on 1 and Skew t Distributions	Nispet PEKİN ALADAĞ, Ayşen APAYDIN A Survey on Nonrandom Patterns of Fuzzy Control Charts	Bülent ALPTEKİN, Metin Can ALIYERDOĞDU, Çağrı BAL, Cengiz Hakan ALADAĞ Different Performance Measures For Forecasting Tourism Demand Of Turkey	

11:15 - 11:30 ÇAY/KAHVE ARASI - COFFEE BREAK

11:30 - 12:30	DAVETLİ - INVITED SPEAKER Oturma Başkanı - Chair Person : Cengiz ÇELİKOĞLU Konuşmacı - Speaker: Prof. Dr. Burhan TÜRKŞEN					
---------------	--	--	--	--	--	--

12:30 - 14:00 YEMEK ARASI - LUNCH

Salon / Hall	SALON 1	SALON 2	SALON 3	SALON 4	SALON 5	SALON 6
--------------	---------	---------	---------	---------	---------	---------

BİLDİRİ OTURUMLARI 4 SESSION 4	UYGULAMALI İSTATİSTİK 3 Oturma Başkanı - Chair Person Zehra MULLIK	İSTATİSTİK YÖNTEMLER Oturma Başkanı - Chair Person Serpil AKTAŞ ALTUNAY	BİYOSTATİSTİK Oturma Başkanı - Chair Person Nuri ÇELİK	MODELLEME VE SİMÜLASYON Oturma Başkanı - Chair Person Orhan KESEMEN	KALKINMA BAKANLIĞI İZEL OTURUMU 1 Oturma Başkanı - Chair Person Ayşen APAYDIN	ACTUARIAL SCIENCES Oturma Başkanı - Chair Person Uğur KARABEY
14:00 - 14:15	Erkan ARI, Zeki YILMAZ Sıralı ve Çok Terimli Lojistik Modellerin Uyum İyiliği Göstergeleri Açısından Karşılaştırılması	Elvan ANTÜRK HAYAT, Özgür ALPAY, Aytila Hatice ELHAN, Derya BÜTÜNA Kullandıkları VZ'de Akutik Hastaların Belirli Mesajları	Ayşen Melike AYTIĞUS KOŞAN, Nispettin ÖZ, Aytila Hatice ELHAN, Derya BÜTÜNA Developing An Item Bank For Progress Test And Application Of A Computerized Adaptive Testing By Simulation In Medical Education	Alper SAĞDUĞLU, Yusuf AKTAŞ, İbrahim UŞLAN Toz Metalürjisiyle Üretim Kuvvetinde Yaşamalı Bronz Yataklarda Yağın Yaşam, Gözetim ve Kontrolü	Leyla BİLEN KAZANCIK, Gülce SAKIN Bölgesel Boyutlu Olsun Belirleme Çalışması	Başak BİLUT KARABEYİK, Sıla SAHİN The Translated Gamma Process And Finite Time Run Probability: An Application Of Turkey Compulsory Traffic Insurance
14:15 - 14:30	Şengül ÇANGIR, Handan ANKARALI, Özge PASIN P/C Boyutu Çapraz Tablolarda Çoklu Karşılaştırma Yöntemlerinin İncelenmesi: Bir Uygulama	Özge KARADAĞ, Hülya ÖLMEŞ, Serpil AKTAŞ ALTUNAY Bağlılık Ölçümleri Testlerinde Yöntemlerin Karşılaştırılması	Fatma KIZILAYLA, Cemil ÇOLAK, Oklay KIZILAYLA Sol Aritmi Çapının Temel Bilgilerin Regresyonu, Kısmi En Küçük Kareler Regresyonu Ve Yapı Sınır Ağları İle Tahminlerinin Karşılaştırılması	Murat DOĞAN, Özgür ÖZAYDIN, Veysel YILMAZ Doğrusal Faktör Analizinde Ölçüm Hatası, Tahmin Yöntemleri Ve Normal Uyum Ölçütürlerine Etkisi	Sevgi SAHAÇ, Gonca BUDAK T-71 Bölgesi İçin Sosyo-Ekonomik Gelişim Analizi	Bilgi YILMAZ, Bilal Akper İNKAYA, Yeliz Yolcu ÖZUR Comparison Of Different Methods To Compute The Greeks

14:30 - 14:45	Fatih ÇEMREK, Arçay ISI, Hakkı POLAT Az Gelişim Ülkelerinin Sağlık ve Askeri Harcamalara İlişkin Değişkenlerin Bakımından İncelenmesi	Ufuk HURİYEY, Farnaz GÜLHASEMİ SORHABİ Yol-Bakım Birimlerinde Yol-Bakım Araçları İle Yol Güvenliğini Arttırmak İçin Testlerin Planlanması	Nuri ÇELİK, Gül ÖLÇÜN KARACAN Moleküler Varyans Analizi (Amova) Tekniği	Neslihan GÜNDÜZ, Celal AYDIN Sağlam Ölçek Tahmin Edicilerinin Çarpıklık Bağlı Olarak Davranışlarının Karşılaştırılması	Zeynep ERSAYIN Kalkınma Değerlendirmesi: Çukurova ve İzmir Kalkınma Ajanslarının KOBİ Destekleri	Uğur KARABEY Risk Capital Allocation: Importance Of Copula Selection
---------------	--	--	--	---	---	---

14:45 - 15:00	Gonca ÇİNER, Veysel YILMAZ Yaşamın Kararlarının Anlamak: Bir Yapısal Etki Modellemesi Önerisi	Emel BAŞAR Yaşamın Süzümüne Analizinde Gırtlak Kirliğin Modeli	Ash SÜNER, Özge DİCLE, Selman SÖRMEK, Gökhan KARAKÖLALİ, Can Cengiz ÇELİKOĞLU Dört: Beklenen Kanser Tedavisinde Karar Destek	Rana SEN, Veysel YILMAZ Model Belirlemesi, Ölçüm Hatası Ve Tahmin Yöntemini Yayımlı Etkilik Modelleri Uyum Ölçütürüne Etkisi	Leyla BİLEN KAZANCIK, Hasan SANKALMIŞ, Tuğba DENİZ Sürdürülebilir Bölgesel Kalkınmanın Ölçülmesi: "Sürdürülebilirlik Çemberleri" ve Uygulama	Gizem ÖZAK, Sevgi KEŞTEL, Ahmet GENÇ An Early Warning Model For Insurance Companies In Turkey
---------------	--	---	---	---	---	--

15:00 - 15:15	Adnan KARABİBRAHİMOĞLU, Hakan BAŞBOZKURT, Arçay BAŞBOZKURT, Teşin ASAR, Asır GENÇ Ölüm Nedenlerinin Mekansal Analizi	Erdinc KÖLALY, Hakan DANIŞOĞLU Bölgesel Kestiri Çok Etkeli Tasarımlarda En İyi Bölge Seçimi	Hande KONSUK, Ayten YİĞİTER Bölgesel Durumun Varsayımı Altında Yönelimli Olayların Homojen Pazar Süreci İle Modellemesi	Fatma Gül ANKUL, Birgül ŞENOLU Bunlar Yaşamın Konum Ve Ölçek Parametrelerinin Tahmini: Monte-Carlo Simülasyon Çalışması	Alper DEMİR TRC3 Bölge İçin Rekabet Analizi	Fulya AKAT, Sevgi KEŞTEL Mortgage Agency Approach In Turkish Mortality Estimation
---------------	---	--	--	--	--	--

15:15 - 15:30	Savaş Demir ATALAY, Ali MERT İki Ortak Bozulma Kriteri Sistemlerin Profüt Güvenlilikleri	Adem DOĞANER, Sinan ÇALIK Sağlık Yönetimi Enformasyon Sistemlerinin Erken Teşhis Ve Tanı İçin Stokastik Modeller İle Geliştirilmesi	Orhan KESEMEN, Buğra Kaan TIRYAKI Ki-Kare Uyum İyiliği Testinde Minimum Sınıf Sayısının T-Testi Kullanarak Benzetim Belirlemesi	Fuat KARAGÜNEY, Nurtan ELVAN KÖNSOY, İsmail ABAS, Mehmet Emin YİĞİTBAŞI, Mehmet BÜKRÜ ANILTAÇ Bölgesel Düzeyde Mekansal Kararların Alınması Sürecinde İstatistiksel Saha Araştırılmalı Konya Karaman Örneği	Murat BÜYÜKYAZICI Optimal retention for profit maximizing under VaR levels constraints	
---------------	---	--	--	--	---	--

15:30 - 15:45 ÇAY/KAHVE ARASI - COFFEE BREAK

29 EKİM / OCTOBER 2013 SALI / TUESDAY

Salon / Hall	SALON 1	SALON 2	SALON 3	SALON 4	SALON 5	SALON 6
BİLDİRİ OTURUMLARI 5 SESSION 5		İSTATİSTİK TEORİ 2 Oturma Başkanı - Chair Person Yücel TERZİ	VERİ ANALİZİ VE MODELLEME Oturma Başkanı - Chair Person Aydın KARADAĞCA	AKTÜYERLER BİMLERİ 2 Oturma Başkanı - Chair Person Seray KEŞTEL	APPLIED STATISTICS 2 Oturma Başkanı - Chair Person Ömer Utku ERZENGİN	KALKINMA BAKANLIĞI İZEL OTURUMU 2 Oturma Başkanı - Chair Person Turkan Erbay Dalkılıç
15:45 - 16:00		Esin AVCI, Nural BEYRÖĞLU, Nural YAYI Meme Kanseri Verisinin Bayesci Sağlık Analizi İle İncelenmesi	Ercüment YILMAZ, Temel KAYIÇOĞLU, Sadettin KAYIPRAZ, Ömer Said SEZGİN Konik İyonlu Bilgisayarlı Tomografi Görüntülerindeki Çene Patolojilerinin Rasgele Yürüyüşü Belirlemesi	Ashhan SENTÜRK ACAR, Uğur KARABEY Trafik Sigortasının Risk Ölçümü	Demirberk ÖNLU, Seray KEŞTEL Linkage Between Education, Economic Growth And Co2 Emission: A Multivariate Analysis On Turkey	Fatma AVŞAR, Hale Mahzemeni GÜLBAZ, Candan Ümit ÖZDEN, Cem BAYRAK, Hülya YORULMAZ Doğu Marmara 2014-2022 Bölge Planı Kapsamında Hazırlanan Mikro Bölgeleme Çalışması
16:00 - 16:15		Hatice Yağmur GÜRKAN, Gül ERGÜN Markov Zinciri Monte Carlo Yöntemini Dinamik Doğrusal Modellerde Uygulamaları	Hasan ERTAŞ, Selahattin KAÇIRANLAR Regresyon Modeli İçin Hemen Hemen Yavaş Genelleştirilmiş Lu M Tahmini Edici	Bethül Zehra KARAGÜL, Murat BÜYÜKYAZICI Hasarlar Ve Yatırımlar Arasındaki Bağımlılık İle Hayat Düş Sigorta Sözleşmelerinde Dinamik Finansal Analiz - Bir Uygulama	Ömer Utku ERZENGİN, Sıdıka Gamze ERZENGİN Fractal Dimension of Cement Pastes Microstructure based on Statistical Power of Regression Lines Comparison	Osman TOKŞİN Apare Mal Desteklerinde Sorun Odaklı İzleme Değerlendirme Modeli - Etklilik Analizi Örneği
16:15 - 16:30		Mehmet Ali CENGİZ, Naci MURAT, Tuba SENE OECD Ülkelerinin Enerji Performanslarını Bayesci Stokastik Sınırlı Analizi İle Değerlendirmesi	Esra PAMUKÇU, Hüseyin BÖZÜDOĞAN Aynı Karar Ölçüm Problemlerinde Bilgi Karmaskılı Kriteri (komp) Ve Düzeltme Kuvvetsiz Yayımları İle Boyut İndirgenme Alternatif Bakış Gen Versi Örneği	Esra NEYRUZ, Sema TÜZEL, Enerajli BODU Türkiye Ölümülük Oranlarının Düzeltirilmesi ve Tahmin Edilmesi	Rafail KASIMBEYLİ A New Scalarization Method in Nonlinear and Nonconvex Multi-Objective Optimization	Emine ARSLAN PAHLI, Yasin DALGIÇ, Tufan İlçe Kalkınma Endeksi
16:30 - 16:45		Naci MURAT, Mehmet Ali CENGİZ, Erol TERZİ Bayesci Yayımlı Etkilik Modellerde Farklı Tüpe Ölçütürleri Analizi İçin Eşik Değer Yaklaşımı	Uğur ŞEVİK, Tolpa BERBER, Hüdayet ERDÜL Rcino Gözetimlerinin İstatistiksel Olarak Kalite Değerlendirmesi	Funda KUL, Özgür KARADAĞ, Merai SUCU Ölümülük Gelişim Oranlarının Modellemesi	Asena ÖZDEMİR, Demirberk ÖNLU, Seray KEŞTEL, Kaanpaz YILDIRAK Türkish Electricity Price Behavior Based On Some Economic Indicators: An Econometric Approach	Hasan BİLUT, Yücel ÖNER, Taner TUĞÇE Türkiye'de Kalkınma Ajansları Kapsamındaki İllerin Bazı Sosyoekonomik Göstergeler Bakımından Çok Değişkenli İstatistiksel Analizlere Değerlendirilmesi
16:45 - 17:00		Yüksel TERZİ, Nesrin ALKAN, Mehmet Ali CENGİZ Bayesci Çok Regresyonda Monte Carlo Karşılaştırma Kriterlerinin Değerlendirilmesi	Orhan KESEMEN, Tuncay ULUTURUT Çokgen Alanlarda Ki Değişkenli Ki-Kare Uyum İyiliği Testi	Güven ŞİMŞEK, Durdu KARASOY Yaşam Çözümünde Zayıflık Modeli ve Yapı İçin Regülasyon Etkilikleri Üzerine Bir Uygulama	Bekir ÇETİNTAY, Selma GÜRLER A Comparative Simulation Study For Modified Rank Set Sampling Methods	Erhan DEMİRCAN Çok Değişkenli İstatistiksel Yöntemlerin TRC2 Bölgesi İçin İyileştirme Uygulamaları
17:00 - 17:15		Nimet TÜRKER, Berria GÜLLAY, Selahattin KAÇIRANLAR Geçmişte Dağılım Modellerde Bayesci Lu ve Ridge Tipi Tahmin Edicilerin Karşılaştırılması	Aydın KARADAĞCA, Murat ERDOĞDU, Alper SINAN Parametrik Olmayan Regresyonda Optimal Bart Genleşimi Seçim Yöntemlerinin Karşılaştırılması		Ayşen SAĞBAS, Adnan MAZMANOĞLU An Algorithm For The Determination Of Economic Design Of X-Bar Control Charts	
20:00 - 24:00	GALA YEMEĞİ: CUMHURİYET BALOSU					

HİPOTEZ TESTLERİNDE NORMALLİK VARSAYIMI BOZULDUĞUNDA 1. TİP HATANIN  
DAYANIKLILIĞININ İNCELENMESİ

Abdullah YALÇINKAYA\*, Mehmet Niyazi ÇANKAYA, Ömer ALTINDAĞ, Yetkin TUAÇ  
Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06100, Ankara, TÜRKİYE  
ayalcinkaya@ankara.edu.tr, mncankaya@ankara.edu.tr, altindag@ankara.edu.tr, ytuac@ankara.edu.tr

Giriş

$H_0: \theta = \theta_0$  hipotezini  $H_1: \theta \geq \theta_0$  alternatifine karşı test etmek için kullanılan testlerin birçoğu belirli bir varsayılan model altında oluşturulur. Fakat varsayılan model yanlış olduğunda 1. tip hata ( $\alpha$ ) varsayılan modeldekinden farklı olacaktır. Bu yüzden yanlış model seçiminin  $\alpha$  üzerindeki etkisini incelemek oldukça önemlidir. Bununla ilgili bir teorem aşağıdadır.

**Teorem:**  $T_n$  varsayılan model altında  $\frac{\sqrt{n}[T_n - \mu(\theta_0)]}{\tau(\theta_0)} \rightarrow N(0,1)$  olan test istatistiklerinin bir dizisi olsun. Bu durumda

$H_0: \theta = \theta_0, H_1: \theta \geq \theta_0$  hipotez çifti için  $\frac{\sqrt{n}[T_n - \mu(\theta_0)]}{\tau(\theta_0)} \geq z_\alpha$  testi asimptotik  $\alpha$  düzeye sahiptir. Farz edilsin ki

varsayılan model yanlış olsun ve gerçek model altında  $\frac{\sqrt{n}[T_n - \mu(\theta_0)]}{\tau'(\theta_0)} \rightarrow N(0,1)$  sağlansın.  $\frac{\sqrt{n}[T_n - \mu(\theta_0)]}{\tau(\theta_0)} \geq z_\alpha$

testinin kesin düzeyi  $\alpha'_n$  olmak üzere  $\alpha' = \lim \alpha'_n$  vardır ve  $\alpha' = 1 - \Phi \left[ z_\alpha \frac{\tau(\theta_0)}{\tau'(\theta_0)} \right]$  şeklindedir.

Böylelikle,  $\tau'(\theta_0) \leq \tau(\theta_0)$  iken  $\alpha' \leq \alpha$  dir (Lehmann, 1999).

**Tanım:**  $\mathcal{F}$  varsayılan,  $\mathcal{F}'$  de gerçek model olmak üzere,  $\mathcal{F}$  modeli altında  $\alpha$  asimptotik düzeye sahip olan bir teste  $\mathcal{F}'$  altında

- Eğer  $\theta = \theta_0$  iken  $\alpha' \leq \alpha$  ise, Tutucu (Conservative),
- Eğer  $\theta = \theta_0$  iken  $\alpha' \geq \alpha$  ise, Liberal,
- Eğer  $\theta = \theta_0$  iken  $\alpha' = \alpha$  ise, Dayanıklı (Robust)

denir (Lehmann, 1999).

Normallik varsayımı altında varyans parametresi olan  $\sigma$  için hipotez testi göz önüne alınsın. Eğer  $X_1, X_2, \dots, X_n$  birbirinden bağımsız ve aynı  $N(0, \sigma^2)$  dağılımlı ise  $H_0: \sigma = \sigma_0$  hipotezini  $H_1: \sigma \geq \sigma_0$  a karşı asimptotik  $\alpha$  düzeyinde test etmek için bir red bölgesi

$$\frac{\sqrt{n} \left( \frac{\sum (X_i - \bar{X})^2}{n} - \sigma_0^2 \right)}{\sqrt{2} \sigma_0^2} \geq z_\alpha \quad (1)$$

şeklindedir. Normallik varsayımının sağlanmaması durumunda üstteki red bölgesinin asimptotik düzeyi  $\alpha$  düzeyinden farklı olacaktır.  $X$ 'ler  $Var(X_i) = \sigma^2, Var(X_i^2) = \lambda^2 < \infty$  özelliklerine sahip birbirinden bağımsız ve aynı herhangi  $F$  dağılımlı rasgele değişkenler olsunlar.

Bu durumda;

BİLDİRİ OTURUMLARI I

SESSION I

İstatistik Teorisi 1

$$\frac{\sqrt{n} \left( \frac{\sum (X_i - \bar{X})^2}{n} - \sigma_0^2 \right)}{\lambda} \rightarrow N(0,1) \quad (2)$$

olur ve testin gerçekteki asimptotik seviyesi  $\alpha'$  bu ifade üzerinden yukarıdaki teorem yardımıyla elde edilir (Lehmann, 1999).

## 2. Kontamine Dağılım Altında 1. Tip Hatanın Dayanıklılığı

Kabul edilsin ki  $X_1, X_2, \dots, X_n$  gerçekte  $0 \leq \rho \leq 1$  olmak üzere  $F = \rho F_Y + (1 - \rho) F_Z$  şeklinde verilen

Kontamine dağılıma sahip olsun. Burada  $Z$  rasgele değişkeni Standart Normal dağılımı göstermek üzere  $Y$  rasgele değişkeninin ise sırasıyla aşağıdaki dağılımlara sahip olduğu düşünülün.

- $Y \sim N(0, v^2)$
- $Y \sim SCN(a), f_{SCN}(y) = \frac{1+\alpha y^2}{1+\alpha} \phi(y), -\infty < y < \infty, a \geq 0$
- $Y \sim MGT(p = 1, \mu = 0, \sigma = 1, \lambda = 1, \beta, q)$  (Arslan, 2004)

Bu çalışmada;  $X_1, X_2, \dots, X_n$  rasgele örnekleminin varyans parametresi  $\sigma$  için  $H_0: \sigma = \sigma_0$  hipotezinin  $H_1: \sigma \geq \sigma_0$  e karşı testinde varsayılan model Normal dağılım olduğunda testin asimptotik düzeyinin  $Y$  rasgele değişkeninin ve  $\rho$  nun durumlarına göre nasıl etkilendiği Teorem ve (1), (2) ifadeleri yardımı ile incelenecektir. Ayrıca benzer çalışma konum parametresi için de gerçekleştirilecektir.

## KAYNAKLAR

- [1] Lehmann E.L. (1999), Elements of Large-Sample Theory, New York, Springer-Verlag.
- [2] Gui W., Chen P.H., Wu H. (2013), A Symmetric Component Alpha Normal Slash Distribution, Journal of Statistical Theory and Applications, Vol. 12, No. 1, 55-66.
- [3] Arslan O. (2004), Family of Multivariate Generalized t Distribution, Journal of Multivariate Analysis 89, 329-337.

## ABSTRACT

### INVESTIGATION FOR THE ROBUSTNESS OF SIGNIFICANCE LEVEL WHEN THE NORMALITY ASSUMPTION IN HYPOTHESIS TESTS IS VIOLATED

Many of the tests which used to test  $H: \theta = \theta_0$  are constructed under a postulated model. However, when the postulated model is not correct, true significance level will be different than that of the postulated model. A test which has asymptotic level  $\alpha$  under  $\mathcal{F}$  is said to be conservative under  $\mathcal{F}'$  if  $\alpha'(F) \leq \alpha$  for all  $F \in \mathcal{F}'$ , liberal if  $\alpha'(F) \geq \alpha$  for all  $F \in \mathcal{F}'$ , robust if  $\alpha'(F) = \alpha$  for all  $F \in \mathcal{F}'$  with  $\theta = \theta_0$ . In this study, the true significance level for testing  $H: \sigma = \sigma_0$  against  $\sigma > \sigma_0$  under different contaminated distributions is obtained. Further, the robustness of significance level for testing the location parameter is also investigated.

**Key Words:** Contaminated Distribution, Generalized t Distribution, Hypothesis Tests, Robustness, Significance Level, Symmetric Component Normal Distribution

## İKİ ÜÇGENSEL RASTGELE DEĞİŞKENİN BÖLÜMÜNÜN DAĞILIMI

Selim GÜNDÜZ\*, Ali İhsan GENÇ

Çukurova Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 01130, Adana, Türkiye, [sgunduz@cu.edu.tr](mailto:sgunduz@cu.edu.tr)  
Çukurova Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 01130, Adana, Türkiye, [agenc@cu.edu.tr](mailto:agenc@cu.edu.tr)

### 1. Giriş

Başta biyoistatistik, aktüerya ve sistem mühendisliği olmak üzere uygulamalı bilimlerde, rastgele değişkenlerin çarpımı ve bölümü gibi çeşitli fonksiyonlarının dağılımının belirlenmesi problemleri ile sıkça karşılaşmaktadır. Örneğin, biyoistatistik ve sağlık biliminde, ağırlığın, boy uzunluğunun karesine bölünmesiyle vücut kitle indeksi tanımlanmakta, benzer şekilde gelir, gider, harcama, maliyet v.b ekonomik değişkenlerin oranlarıyla da, ekonomi indekslerde çeşitli tanımlamalar yapılmaktadır.  $X$  ve  $Y$ , bağımsız rastgele değişkenler olmak üzere,  $XY$  ve  $X/(X + Y)$  nin dağılımları, başta Donahue (1964) ve Springer (1979) olmak üzere pek çok bilim insanı tarafından çalışılmıştır. Literatürde,  $X$  ve  $Y$  bağımsız rastgele değişkenler olmak üzere, gamma dağılımı, beta dağılımı, genelleştirilmiş pareto dağılımı, laplace dağılımı gibi çeşitli dağılımlar için  $XY$  ve  $X/(X + Y)$  nin tam dağılımları elde edilmiştir.

Üçgenel dağılım, veri toplamanın zor olması veya rastgele değişkenin dağılımının belirlenememesi durumunda kullanılan uygun bir dağılımdır. Üçgenel dağılımla ilgili detaylı bilgilere, Kotz ve Van Dorp (2004) çalışmasından ulaşılabilir. Bu dağılımla ilgili son yıllarda yapılan çalışmalardan biri olarak Glickman ve Xu (2008), üçgenel dağılıma sahip bağımsız iki rastgele değişkenin çarpımının dağılımını elde etmiştir.

Bu çalışmada,  $X$  ve  $Y$ , üçgenel dağılıma sahip bağımsız iki rastgele değişken olmak üzere,  $X/Y$  bölümünün tam dağılımı bulunacak ve elde edilen sonuçlar kullanılarak,  $Y/(X + Y)$  bölümünün dağılımı elde edilecektir. Ayrıca sistem güvenilirliği üzerine bazı uygulamalar yapıldıktan sonra sonuç ve öneriler ile çalışma sonlandırılacaktır.

### 2. Bölümün Dağılımı

Üçgenel dağılıma sahip  $X$  rastgele değişkeninin olasılık yoğunluk fonksiyonu;

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(m-a)}, & a \leq x \leq m \\ \frac{2(b-x)}{(b-a)(b-m)}, & m \leq x \leq b \end{cases} \quad (1)$$

şeklinde verilir ve  $X \sim \text{Tr}(a, m, b)$  yazılır. Aynı zamanda  $Y \sim \text{Tr}(c, n, d)$ ,  $X$  ten bağımsız bir rastgele bir değişken olsun. Bu çalışmada ilk olarak  $V = X/Y$  bölümünün dağılımı bulunacaktır.  $X$  ve  $Y$  rastgele değişkenlerinin pozitif olduğu varsayımıyla,  $V = X/Y$  nin tam dağılımı,  $V = X(1/Y)$  nin çarpımı olarak yazılarak, Glickman ve Xu (2008) nun yöntemi kullanılacaktır.  $Z = 1/Y$  olmak üzere  $V = XZ$  nin olasılık yoğunluk fonksiyonu;

$$f_V(v) = \int_{x_1}^{x_2} \frac{1}{x} f_X(x) f_Z\left(\frac{v}{x}\right) dz \quad (2)$$

integrali kullanılarak elde edilir. Olasılık yoğunluk fonksiyonu parçalı olduğu için  $X$  ve  $Z$  nin parçalarının tüm kombinasyonları ve bu kombinasyonlara ait tüm alt durumlar düşünülmelidir. (Bknz. Çizelge1.)

MİNİMİNENT VE MAXİMİNENT YÖNTEMLERİ İLE SAĞKALIM VERİ ANALİZİ

Aladdin ŞAMİLOV, Çiğdem GİRİFTİNOĞLU\*, Sevda ÖZDEMİR  
Anadolu Üniversitesi, Yunus Emre Kampüsü, Fen Fakültesi, İstatistik Bölümü, 26470 Eskişehir/TÜRKİYE

e-mail: [asamilov@anadolu.edu.tr](mailto:asamilov@anadolu.edu.tr), [cgiriftinoglu@anadolu.edu.tr](mailto:cgiriftinoglu@anadolu.edu.tr), [sevdaozdmr@gmail.com](mailto:sevdaozdmr@gmail.com)

Sağkalm Analizi için MinMinxEnt ve MaxMinxEnt Yöntemleri

Bu çalışmada, 1935 – 1944 yılları arasında Connecticut şehrindeki bağırsak kanserine yakalanmış erkek hastaların verileri göz önünde bulundurulmuş ve veriler Çizelge 1’de gösterilmiştir. Burada sansüre uğrayan 52 hasta çalışmaya dahil edilmeyerek gözlenen ölüm olasılıkları hesaplanmış ve bu olasılıkların toplamı 0.9821 olarak bulunmuştur. Bu değer 1’den küçük olması sebebiyle  $1 - 0.9821 = 0.0179$  farkı, her bir sansürlü gözleme eşit olarak paylaştırılmış ve sansürlü gözlemlerin ait oldukları aralıktaki verilerin olasılıklarına eklenerek yeni olasılıklar elde edilmiştir. Elde edilen bu yeni olasılıklara ise düzeltilmiş olasılıklar adı verilmiş ve sonuçlar Çizelge 1’de gösterilmiştir.

Çizelge1. 1935 – 1944 yılları arasında Connecticut şehrindeki erkek hastaların bağırsak kanseri verileri, Gözlenen Olasılıklar, Düzeltilmiş Olasılıklar

t	Aralık başlangıcında hayatta olan hasta sayısı $n_i$	$n_i'$	Aralık boyunca ölen hasta sayısı $d_i$	Aralık boyunca sansüre uğrayan hasta sayısı $c_i$	Gözlenen Olasılıklar $q_i$	Düzeltilmiş Olasılıklar $P_i^*$
0 – 1	388	336	167	2	0.4970	0.5030
1 – 2	219	167	45	1	0.1339	0.1369
2 – 3	173	121	45	1	0.1339	0.1369
3 – 4	127	75	19	0	0.0565	0.0565
4 – 5	108	56	17	0	0.0506	0.0506
5 – 6	91	39	11	1	0.0327	0.0357
6 – 7	79	27	8	0	0.0238	0.0238
7 – 8	71	19	5	0	0.0149	0.0149
8 – 9	66	14	6	1	0.0179	0.0208
9 – 10	59	7	7	0	0.0208	0.0208

Çizelge1. İki üçgensel rastgele değişkenin bölümünün olasılık yoğunluk fonksiyonunun belirlenmesindeki durumlar

Durumlar	Alt durumlar
$a \leq x \leq m, \frac{1}{d} \leq z \leq \frac{1}{n}$	$\frac{a}{n} < \frac{m}{d}, \frac{a}{n} = \frac{m}{d}, \frac{a}{n} > \frac{m}{d}$
$a \leq x \leq m, \frac{1}{n} \leq z \leq \frac{1}{c}$	$\frac{a}{c} < \frac{m}{n}, \frac{a}{c} = \frac{m}{n}, \frac{a}{c} > \frac{m}{n}$
$m \leq x \leq b, \frac{1}{d} \leq z \leq \frac{1}{n}$	$\frac{m}{n} < \frac{b}{d}, \frac{m}{n} = \frac{b}{d}, \frac{m}{n} > \frac{b}{d}$
$m \leq x \leq b, \frac{1}{n} \leq z \leq \frac{1}{c}$	$\frac{m}{c} < \frac{b}{n}, \frac{m}{c} = \frac{b}{n}, \frac{m}{c} > \frac{b}{n}$

Bu durumlar ve alt durumlar tek tek göz önünde bulundularak (2) no lu integral hesaplanırsa aşağıda verilen (3) no lu fonksiyon yardımı ile,  $V = X/Y$  nin tam dağılımı (4) no lu eşitlikteki gibi elde edilir.

$$U(x) = \begin{cases} 1, & x < 0 \\ 2, & x = 0 \\ 3, & x > 0 \end{cases} \quad (3)$$

$$f_V(v) = f_{1U\left(\frac{a-m}{n-d}\right)}(v) + f_{2U\left(\frac{a-m}{c-n}\right)}(v) + f_{3U\left(\frac{m-b}{n-d}\right)}(v) + f_{4U\left(\frac{m-b}{c-n}\right)}(v) \quad (4)$$

Benzer şekilde,  $W = X/Y$  bölümünün dağılımından elde edilen sonuçlar kullanılarak, bölümünün dağılımı bulunur.

$$W = \frac{Y}{X+Y} = \frac{1}{1+V}$$

KAYNAKLAR

- [1] Donahue, J.D., 1964. Products and Quotients of Random Variables and Their Applications. ARL 64-115. Aerospace Research Laboratories, Ohio.
- [2] Glickman, T. S., Xu, F. 2008. The distribution of the product of two triangular random variables, Statist. and Probab. Letters, 78, 2821-2826.
- [3] Gündüz S. and Genç A.İ. (2013). The distribution of the quotient of two triangular random variables. (Submitted).
- [4] Kotz, S., van Dorp, J.R., 2004. Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications. World Scientific, Singapore
- [5] Springer, M.D., (1979). The Algebra Of Random Variables, Wiley, New York.

ABSTRACT  
THE DISTRIBUTION OF THE QUOTIENT OF TWO TRIANGULAR RANDOM VARIABLES

The exact distributions of the quotients  $X/Y$  ve  $Y/(X+Y)$  when  $X$  and  $Y$  are independent and triangularly distributed random variables are obtained. These quotients are useful especially in operations research and reliability engineering, and some reliability applications of the results are also given.

**Keywords:** triangular distributions, distribution of quotient, functions of random variable

Not: Burada  $n_1$ , aralık başlangıcında hasta olduğu planlanan ama çalışma sırasında sansüre uğrayan 52 gözlemin çalışmaya dahil edilmemesi ile aralık başlangıcındaki sağkalan hastaların sayısı göstermektedir.

Bu çalışmada, Kullback – Leibler ölçümü açısından, olasılık toplamları 1' den küçük olan dağılıma en yakın olan dağılımın bulunması, belli anlamda "sansürlü" veriyi dikkate alarak problemi çözmeye imkanı sağlamaktadır. Bu durum göz önünde bulundurularak, MinMinxEnt Yöntemi yukarıda verilen veriye uygulanmış ve  $(\text{MinMinxEnt})_m$ ,  $m = 1, 2, \dots, 4$  dağılımları elde edilmiştir. Bu sonuçlar ve elde edilmiş dağılımların veriyeye uyumunun  $\chi^2$  ve RMSE Kriteri sonuçları Çizelge 2'de gösterilmiştir.

Çizelge2.

$(\text{MinMinxEnt})_m$ Distribution	D(p : q)	$\chi^2$ Hesap Değeri	$\chi^2$ Tablo Değeri	RMSE
$(\text{MinMinxEnt})_1$	0.0260387930	0.0118	$\chi_{8, \alpha}^2 = 15.51$	0.0338
$(\text{MinMinxEnt})_2$	0.0260457605	0.0116	$\chi_{7, \alpha}^2 = 14.07$	0.0333
$(\text{MinMinxEnt})_3$	0.0260561615	0.0112	$\chi_{6, \alpha}^2 = 12.59$	0.0329
$(\text{MinMinxEnt})_4$	0.0260786357	0.0103	$\chi_{5, \alpha}^2 = 11.07$	0.0317

Çizelge 2'ye bakıldığında  $\chi^2$  istatistiksel kriterine göre Bağırsak kanseri hastalarının yaşam sürelerinin modellemeye tüm MinMinxEnt Dağılımlarının veri setini açıkladığı görülmektedir. RMSE kriteri değerlerine bakıldığında ise MinMinxEnt Dağılımlarından  $(\text{MinMinxEnt})_4$  dağılımının veri setini en iyi açıklayan MinMinxEnt Dağılımı olduğu sonucuna varılmıştır.

Benzer şekilde  $(\text{MaxMinxEnt})_m$ ,  $m = 1, 2, \dots, 4$  dağılımları da bulunmuş ve gereken araştırmalar yapılmıştır.

#### KAYNAKLAR

- [1] Deshpande J. V., Purohit S.G. (2005), Life Time Data: Statistical Models and Methods, India, Series on Quality, Reliability and Engineering Statistics.
- [2] Lee E. T., Wang J.W. (2003), Statistical Methods for Survival Data Analysis, Oklahoma, Wiley-Interscience.
- [3] Shamilov A. (2006), A Development of Entropy Optimization Methods, Wseas Transactions on Mathematics, Vol. 5, Issue 5 568-575
- [4] Shamilov A. (2007), Generalized Entropy Optimization Problems and The Existence of Their Solutions, Physica A : Statistical Mechanics and Its Applications, 382(2) 465-472
- [5] Shamilov A. (2010), Generalized Entropy Optimization Problems with Finite Moment Functions Sets, Journal of Statistics and Management Systems, Vol. 13, Issue 3 595-603

**Key Words:** Censored Observation, Generalized Entropy Optimization Methods, MinxEnt, MinMinxEnt, MaxMinxEnt distribution

#### KESİKLİ DÜZGÜN DAĞILIMDAKİ SIRA İSTATİSTİKLERİN MOMENT ÇIKAN FONKSİYONLARI

Ayşe TURAN BUĞATEKİN, Sinan ÇALIK  
Fırat Üniversitesi Fen Fakültesi İstatistik Bölümü 23119 Elazığ/ TÜRKİYE  
[aturan@firat.edu.tr](mailto:aturan@firat.edu.tr) [scalik@firat.edu.tr](mailto:scalik@firat.edu.tr)

#### 1. Giriş

Moment çıkaran fonksiyonlar matematik ve istatistiğin çeşitli alanlarında, özellikle kombinator analiz ve fark denklemleri teorisinde önemli bir araç olarak kullanılmaktadır. Bu araç olasılık teorisinde ve stokastik süreçler teorisinin Markov zincirleri, dallanan süreçler ve tesadüfi yürüyüş gibi alanlarında vazgeçilmezdir. Moment çıkaran fonksiyonlar bekleme hatları ve inventar modellerinin analizinde de sık sık yer almaktadır.

Kesikli dağılımlardaki sıra istatistiklerinin ilk iki momentini Khatri (1962), alternatif bir metotla Arnold vd. (1992) ve m. inci momentini Çalık ve Güngör (2010) elde etmiştir. Balakrishnan (1986) keyfi sürekli dağılımlardaki n boyutlu bir örnekteki sıra istatistiklerinin tekli ve çarpım momentleri için mevcut olan indirgeme ilişkisi ve eşitsizliği kesikli durumlar için vermiştir. Nagaraja (1992) kesikli sıra istatistiklerindeki tüm gelişmeleri açık bir şekilde tartışmıştır.

Kesikli düzgün dağılımdaki sıra istatistiklerinin ilk iki momentini Arnold vd. (1992) ve değişik bir yolla Ahsanullah ve Nevzorov (2001) vermişlerdir. Ayrıca Turan (2008) kesikli düzgün dağılımdaki sıra istatistiklerinin örnek ekstremlerinin beklenen değer ve varyansları için, n=10 örnek boyutuna kadar, cebirsel ifadeler ve sayısal sonuçlar elde etmiştir.

Bu çalışmada ise kesikli düzgün dağılımdaki sıra istatistiklerinin örnek ekstremlerinin moment çıkaran fonksiyonları ve momentlerin bulunmasını kolaylaştıracak olan s inci türevleri elde edilmiştir.

#### 2. Sıra İstatistikleri ve Dağılımları

$X_1, X_2, \dots, X_n$  örnekleme artan sırada  $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$  biçiminde sıralansın. Bu sıralanmış  $X_{1:n}, X_{2:n}, \dots, X_{n:n}$  tesadüfi değişkenlerine sıra istatistikleri denir.

$X_{r:n}$ ' nin birikimli dağılım fonksiyonu (cdf),

$$F_{r:n}(x) = \int_0^{F(x)} \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r} dt \quad (1)$$

$$= I_{F(x)}(r, n-r+1), -\infty < x < \infty$$

olarak yazılabilir [Arnold vd., 1992].  $F_{r:n}(x)$ ' in ifadesi kullanılarak  $X_{r:n}$ ' nin pmf' si,

$$f_{r:n}(x) = \int_{F(x^-)}^{F(x)} \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r} dt \quad (2)$$

şeklinde elde edilir [Arnold vd., 1992].

$X_1, X_2, \dots, X_n$  pmf' si  $f(x) = 1/k$  ve cdf' si  $F(x) = x/k$   $x = 1, 2, \dots, k$  olan n tane bağımsız ve aynı dağılımlı kesikli anakütledeki düzgün tesadüfi değişkenler olsun. (1)' den  $X_{r:n}$ ' nin pmf' si



$$f_{r:n}(x) = \frac{n!}{(r-1)!(n-r)!} \int_{(x-1)/k}^{x/k} u^{r-1} (1-u)^{n-r} du \quad (3)$$

şeklinde yazılabilir.

### 3. Kesikli Düzgün Dağılımdaki Sıra İstatistiklerin Moment Çıkaran Fonksiyonları

**Teorem 3.1.**

$X_1, X_2, \dots, X_n$  kesikli düzgün dağılıma sahip tesadüfi değişkenler ve  $X_{1:n}$  bu tesadüfi değişkenlere karşılık gelen 1 inci sıra istatistiği olsun.  $X_{1:n}$ ' nin moment çıkararı fonksiyonu

$$M_{X_{1:n}}(t) = \sum_{i=1}^k [e^{it} - e^{(i-1)t}] \left( \frac{k+1-i}{k} \right)^n$$

ve  $M_{1:n}^{(s)}(t)$  bu moment çıkararı fonksiyonun s inci türevini göstermek üzere,

$$M_{1:n}^{(s)}(t) = \sum_{i=1}^k e^{it} [i^s e^{it} - (i-1)^s e^{(i-1)t}] \left( \frac{k+1-i}{k} \right)^n$$

şeklinde yazılabilir.

**Teorem 3.2.**

$X_1, X_2, \dots, X_n$  kesikli düzgün dağılıma sahip tesadüfi değişkenler ve  $X_{n:n}$  bu tesadüfi değişkenlere karşılık gelen n inci sıra istatistiği olsun.  $X_{n:n}$ ' nin moment çıkararı fonksiyonu

$$M_{X_{n:n}}(t) = \sum_{i=1}^{k-1} [e^{it} - e^{(i+1)t}] \left( \frac{i}{k} \right)^n + e^{kt}$$

ve  $M_{n:n}^{(s)}(t)$  bu moment çıkararı fonksiyonun s inci türevini göstermek üzere,

$$M_{n:n}^{(s)}(t) = \sum_{i=1}^{k-1} [i^s e^{it} - (i+1)^s e^{(i+1)t}] \left( \frac{i}{k} \right)^n + k^s e^{kt}$$

**KAYNAKLAR**

- [1] Ahsanullah, M. and Nevzorov, V. B., 2001. Ordered Random Variables, Nova Science Publishers, Inc., New York.  
[2] Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N., 1992. A First Course in Order Statistics, John Wiley and Sons, New York.  
[3] Çalik, S., Güngör, M. And Colak, C., 2010. On the Moments of Order Statistics from Discrete Distribution. Pak. J. Statist., **26(2)**, 417-426.  
[4] Khatri, C. G., 1962. Distribution of Order Statistics for Discrete Case, Ann. Inst. Statist. Math., **14**, 167-171.

**Key Words:** Order statistics, discrete uniform distributions, moment generating function.

### GENEL PARÇALANMIŞ LİNEER MODEL ALTINDA BAZI KABUL EDİLEBİLİR TAHMİN EDİCİLER

Nesrin GÜLER

Sakarya Üniversitesi, FEF, İstatistik Bölümü, 54187, Sakarya, Türkiye, nesring@sakarya.edu.tr

#### Giriş ve Ön Bilgiler

$E(y) = X\beta$  ve  $\text{cov}(y) = V$  olmak üzere, tam model olarak bilinen

$$M = \{y, X\beta, V\} = \{y, X_1\beta_1 + X_2\beta_2, V\} \quad (1)$$

parçalanmış lineer modeli ele alınsın.  $K\beta$ ,  $M$  modeli altında tahmin edilebilir parametrik fonksiyon vektörü olmak üzere,  $K\beta$  vektörünün tüm lineer tahmin edicilerinin kümesi

$$LE(K\beta) = \{Fy : F \in \mathbf{R}^{k \times p}\} \quad (2)$$

olarak gösterilsin.  $Gy \in L(K\beta)$  olmak üzere,  $E(Gy) - K\beta$ ,  $Gy$  lineer tahmin edicisinin yanlışlığı olarak tanımlanır ve bu fark sıfır olduğunda,  $Gy$  yansız bir tahmin edici olur. Yanlılık bir tahmin edicinin niteliğini ölçmek için kullanılan bir ölçüttür. Bunun yanı sıra bir ölçüt olarak kayıp fonksiyonu da kullanılır. Kayıp fonksiyonun beklenen değeri ortalama hata kareler (MSE) matrisi olarak tanımlanır ve bu matrisin izi ise risk fonksiyonu olarak bilinir [3]. Tahmin edicileri risk fonksiyonuna göre karşılaştırırken, bazı tahmin edicilerin uygun olmadığı görülebilir. Örneğin, tüm parametre değerleri üzerinden eğer bir tahmin edicinin riski, diğer bir tahmin edicinin riskinden fazlaysa, bu tahmin edici diğerine göre kabul edilemez olarak adlandırılır. Diğer bir deyişle tüm  $\beta \in \mathbf{R}^{p \times 1}$  için  $R(K\beta; Gy) \leq R(K\beta; Ay)$  olacak şekilde  $Gy \in L(K\beta)$  bulunamıyorsa  $Ay \in L(K\beta)$  lineer tahmin edicisine,  $M$  modeli altında kabul edilebilir tahmin edici denir [5]. Burada

$$R(K\beta; Gy) = iz[E(Gy - K\beta)(Gy - K\beta)'] = iz[\text{cov}(Gy) + (E(Gy) - K\beta)(E(Gy) - K\beta)'] \quad (3)$$

$Gy$  yansız tahmin edicisi diğer tüm yansız tahmin ediciler arasında Löwner sıralamasına göre en küçük kovaryansa sahipse, bu durumda  $Gy$  yansız tahmin edicisi,  $X\beta$  için en iyi lineer yansız tahmin edici (BLUE) olarak bilinir. Bir  $Gy$  yansız tahmin edicisinin  $M$  modeli altında  $X\beta$  için BLUE olmasının gerek ve yeter koşulu  $G$  matrisi

$$\begin{pmatrix} V & X \\ X' & 0 \end{pmatrix} \begin{pmatrix} G' \\ L \end{pmatrix} = \begin{pmatrix} 0 \\ X' \end{pmatrix} \quad (4)$$

denkleminin bir çözümü olacak şekilde bir  $L$  matrisinin mevcut olmasıdır. Bu denklem Pandora's Box denklemi olarak bilinir.

$$C_M = \begin{pmatrix} C_{1M} & C_{2M} \\ C_{3M} & -C_{4M} \end{pmatrix} = \begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}^{-1} \quad (5)$$

olmak üzere,  $M$  modeli altında  $X\beta$  nin BLUE'su  $BLUE(X\beta|M) = XC_{2M}'y$  ile gösterilir [4].

#### Kabul Edilebilir Tahmin Edici

Çalışmada  $M$  modeli ile birlikte

$$M_1 = \{y, X_1\beta_1, V\}, M_r = \{M_1, y, M_1X_2\beta_2, M_1VM_1\} \text{ ve } M_o = \{y, M_1X_2\beta_2, V\} \quad (6)$$

indirgenmiş modelleri de ele alınmaktadır. Bu modeller altında  $M_1X_2\beta_2$  parametre vektörünün BLUE'ları Pandora's Box denkleminde elde edilen simetrik parçalanmış matrislerin genelleştirilmiş tersleri yardımıyla elde edilmektedir. Genelleştirilmiş Frisch-Waugh-Lovell Teoremi'ne göre  $BLUE(M_1X_2\beta_2|M) = BLUE(M_1X_2\beta_2|M_r)$  dir [2]. Bu eşitlik kullanılarak çalışmada öncelikle  $M_1$  ve  $M_o$  modelleri altında  $M_1X_2\beta_2$  parametre vektörünün BLUE'larının  $M$  modeli altında BLUE kalması ile ilgili sonuçlar verilmektedir. Daha sonra  $M_1$  ve  $M_o$  modelleri altında  $M_1X_2\beta_2$  için BLUE'larının  $M$  modeli altında BLUE kalmadığı zaman bu tahmin edicilerin  $M$  modeli altında kabul edilebilir tahmin ediciler olmaları için gerek ve yeter koşullar [1] tarafından verilen genel Gauss-Markov modeli için verilmiş olan bir karakterizasyon yardımıyla elde edilmektedir.

#### KAYNAKLAR

- [1] Baksalary, J. K. & Markiewicz, A. (1988), Admissible linear estimators in the general Gauss-Markov model. Statist. Plan. Inf., 19, 349-359.
- [2] Gross, J. & Puntanen, S. (2000), Estimation under a general partitioned linear model. Linear Algebra Appl., 321, 131-144.
- [3] Puntanen, S., Styan, G. P. H. & Isotalo, J. (2011), Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty. Springer, Heidelberg.
- [4] Rao, C. R. (1971), Unified theory of linear estimation. Sankhyā, Ser.A, 33, 371-394. [Corrigendum (1972), 34, p.194 & p.477.]
- [5] Sengupta, D. & Jammalamadaka, S. R. (2003), Linear Models: An Integrated Approach. World Scientific, River Edge, NJ.

#### ABSTRACT

##### SOME ADMISSIBLE ESTIMATORS UNDER GENERAL PARTITIONED LINEAR MODEL

The general partitioned linear model, denoted by  $M = \{y, X_1\beta_1 + X_2\beta_2, V\}$  and known as full model, with its reduced models is considered. Some results are given related to BLUEs of  $M_1X_2\beta_2$  under reduced models to remain under full model using the Pandora's Box equation introduced by [4]. Furthermore, the necessary and sufficient conditions for the BLUEs of  $M_1X_2\beta_2$  under reduced models are obtained when they considered as admissible estimators under full model.

**Key Words:** Admissible estimator, BLUE, Frisch-Waugh-Lovell theorem, partitioned linear model, reduced linear model.

#### LİNEER KARMA MODELLERDE ETKİ ANALİZİNİN UYGULAMASI

Özge KURAN\*, M. Revan ÖZKALE

Çukurova Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü, 01330, Adana, TÜRKİYE

ozge-kuran@hotmail.com, mrevan@cu.edu.tr

#### 1. Model

İstatistiksel modeller, verinin tanımlanması ve veri yapısını en iyi temsil eden matematiksel formun belirlenmesi amacıyla kullanılan modellerdir. En yaygın kullanılan istatistiksel model sıradan lineer regresyon modeli olmasına rağmen bazen sabit etkilere ilave olarak rastgele etkileri de içeren modellere uyan veri yapıları ile de karşılaşmak mümkündür. Bu durumda, lineer regresyon modellerinin genişletilmiş bir şekli olan lineer karma modeller ortaya çıkmaktadır. Genel olarak lineer karma bir model,

$$\begin{aligned} y &= X\beta + Zu \\ &+ \varepsilon \end{aligned} \quad (1)$$

şeklinde ifade edilir. Burada,  $y$ ,  $n \times 1$  tipinde gözlemlenebilir yanıt vektörü,  $\beta$ ,  $p \times 1$  tipinde sabit etkilere ilişkin parametreler vektörü,  $X$  ve  $Z$  sırasıyla  $n \times p$  ve  $n \times q$  tipinde sabit ve rastgele etkilere ilişkin tasarım matrisleri,  $u$ ,  $q \times 1$  tipinde rastgele etkilere ilişkin parametreler vektörü ve  $\varepsilon$ ,  $n \times 1$  tipinde rastgele hataların bir vektörüdür.  $u \sim N(0, \sigma_u^2 D)$  ve  $\varepsilon \sim N(0, \sigma_\varepsilon^2 I_n)$  şeklinde olup, karşılıklı olarak bağımsızdırlar. (1) ile verilen modelde,

$$u = \begin{bmatrix} u_1 \\ \vdots \\ u_r \end{bmatrix}, Z = [Z_1 \cdots Z_r] \text{ biçimindedir ve bu model, lineer karma model tiplerinden varyans bileşenleri modeli}$$

adını alır. Bu modelde,  $i = 1, \dots, r$  için  $q = q_1 + q_2 + \dots + q_r$  olmak üzere,  $u_i$  ve  $Z_i$  sırasıyla  $q_i \times 1$  ve  $n \times q_i$  tipindedir.  $u_i \sim N(0, \sigma_u^2 D_i)$  dağılımından alınan bağımsız rastgele değişkenlerin vektörü ve  $D$ ,  $i$ -inci bloğu  $\gamma_i I_{q_i}$  olan blok-köşegen bir matristir. Bu çalışmada  $\gamma_i = \frac{\sigma_u^2}{\sigma_\varepsilon^2}$  şeklinde olup, varyans bileşen oranları adını alır (Zewotir ve Galphin, 2005).

#### 2. Lineer Karma Modellerde Etki Analizi

(1) ile verilen varyans bileşenleri modeli altında

$$E(y) = E(X\beta + Zu + \varepsilon) = X\beta + ZE(u) + E(\varepsilon) = X\beta$$

$$\begin{aligned} \text{Var}(y) &= \text{Var}(X\beta + Zu + \varepsilon) = Z\text{Var}(u)Z^T + \text{Var}(\varepsilon) = \sigma_u^2 ZDZ^T + \sigma_\varepsilon^2 I_n \\ &= \sigma_\varepsilon^2 \underbrace{(ZDZ^T + I_n)}_H = \sigma_\varepsilon^2 \underbrace{\left( \sum_{i=1}^r \gamma_i Z_i Z_i^T + I_n \right)}_H = \sigma_\varepsilon^2 H \end{aligned}$$

olmak üzere,  $y \sim N(X\beta, \sigma_e^2 H)$  şeklindedir ve  $\gamma_i$  ler biliniyorken  $\beta$ ,  $\sigma_e^2$  ve  $u$  nun maksimum likelihood (ML) tahminleri  $\hat{\beta} = W(X, X)^{-1} W(X, y)$ ,  $\hat{\sigma}_e^2 = W(d, d)/n$  ve  $\hat{u} = DW(Z, d)$  ile verilir (Zewotir ve Galphin, 2005). Burada,  $d = y - X\hat{\beta}$  ve  $W(A, B) = A^T H^{-1} B$  şeklindedir (Hemmerle ve Hartley, 1973).  $\hat{\beta}$  ya,  $\beta$  nun en iyi lineer yansız tahmin edicisi (BLUE) ve  $\hat{u}$  ya  $u$  nun en iyi lineer yansız ön tahmin edicisi (BLUP) adı verilir (Henderson ve ark., 1959).

Bir analizden elde edilen sonuçlar, bir veya daha fazla gözlemden büyük ölçüde etkilenebilir, yani, gözlemlerin tümü istatistiksel modellerde eşit etkiye sahip değildir. Gözlem silme tanılamaları, bireysel gözlemlerin istatistiksel model üzerindeki etkisini ölçmek için kullanılırlar. Christensen ve ark. (1992) ile Zewotir ve Galphin (2005), gözlem silme tanılamalarından yararlanarak lineer karma modellerdeki etki analizi üzerinde durmuşlardır. Bu çalışmada,  $i$ -inci silinmiş gözlemin sabit ve rastgele etkilere ilişkin parametre tahminleri üzerindeki etkisini ortaya çıkarmak amacıyla Zewotir ve Galphin (2005) tarafından elde edilen  $\hat{\beta}_{(i)}$ ,  $\hat{\sigma}_{e(i)}^2$ ,  $\hat{u}_{(i)}$  formüllerinin kullanılmasıyla, klasik regresyon etki tanılamalarının lineer karma modele genişletilmiş olan biçimleri incelenmiştir. Sabit etkili parametre tahminleri üzerindeki etki analizi için Cook uzaklığının benzeri ( $CD(\beta)$ ), varyans oranının benzeri ( $|VR - 1|$ ), Cook-Weisberg istatistiğinin benzeri ( $|CWI|$ ) ve Andrews-Pregibon istatistiğinin benzeri ( $AP$ ) ölçüleri üzerinde durulmuştur. Rastgele etkili parametre tahminleri üzerindeki etki analizi için ise Cook uzaklığının benzeri ( $CD(u)$ ) ölçüsü incelenmiştir.

### 3. Uygulama

Beckman ve ark. (1987), New Mexico'da kurulu ABD Enerji Bakanlığı'na bağlı ulusal bir laboratuvar olan Los Alamos tarafından yüksek etkili partikül hava (HEPA-High Efficiency Particular Air) filtreleri üzerinde yapılan çalışma sonucunda elde edilen aerosol penetrasyon verisinin bir kısmını kullanarak, lineer karma modellerde yerel etki tanılamalarını geliştirmişlerdir. Bu çalışmada, Beckman ve ark. (1987) tarafından kullanılan aerosol penetrasyon verisi, lineer karma modellerdeki etki tanılamalarının incelenmesi amacıyla Matlab yardımıyla analiz edilmiştir.

### KAYNAKLAR

- [1] Beckman, R. J., Nachtsheim, C. J., ve Cook, R. D. (1987), Diagnostics for Mixed Model Analysis of Variance, *Technometrics*, 29: 413-426.
- [2] Christensen, R., Johnson, W., ve Pearson, L. M. (1992), Case-deletion Diagnostics for Mixed Models, *Technometrics*, 34: 38-45.
- [3] Hemmerle, W. J., ve Hartley, H. O. (1973), Computing MLE for the Mixed A. O. V. Model Using the W Transformation, *Technometrics*, 15(4): 819-831.
- [4] Henderson, C. R., Kempthorne, O., Searle, S. R., ve von Krosigk, C. M. (1959), The Estimation of Environmental and Genetic Trends from Records Subject to Culling, *Biometrics*, 15(2): 192-218.
- [5] Zewotir, T., ve Galphin, J. S. (2005), Influence Diagnostics for Linear Mixed Models, *Journal of Data Science*, 3: 153-177.

### AN APPLICATION OF THE INFLUENCE ANALYSIS IN LINEAR MIXED MODELS

In this study, linear mixed model influence diagnostics were examined in order to reveal the effect of  $i$ -th deleted observation on parameter estimates about fixed and random effects. To demonstrate the importance of these influence diagnostics in linear mixed model, aerosol penetration data set (Beckman ve ark., 1987) was analyzed with the help of Matlab.

**Key Words:** Case Deletion, Influence Observations, Fixed Effects, Random Effects

## BİLDİRİ OTURUMLARI 1

### SESSION 1

#### Uygulamalı İstatistik 1

SAĞLIK ÇALIŞANLARININ KARŞILAŞTIKLARI ŞİDDETİ OLUŞTURAN EN ETKİN FAKTÖRÜN  
LİNEER MODELLERDEN

2- FAKTÖRLÜ ÇAPRAZ SINIFLANDIRMA ÇÖZÜMLEME YÖNTEMİYLE ARAŞTIRILMASI VE BİR  
UYGULAMA

Adnan MAZMANOĞLU<sup>1</sup>, Aysun SAGBAS<sup>2</sup>, Seda MAZMANOĞLU<sup>3</sup>

<sup>1</sup>Department of Statistics, University of Istanbul Aydın, Florya Campus, Istanbul/Turkey  
[adnanmazmanoglu@aydin.edu.tr](mailto:adnanmazmanoglu@aydin.edu.tr)

<sup>2</sup>Department of Industrial Engineering, University of Namik Kemal, Corlu, Tekirdag/ Turkey [asagbas@nku.edu.tr](mailto:asagbas@nku.edu.tr)

<sup>3</sup>University of Kocaeli, Faculty of Medicine, doctor of medicine, Izmit /Turkey [sedamazmanoglu@gmail.com](mailto:sedamazmanoglu@gmail.com)

1. GENEL AÇIKLAMA

Sosyal bilimlerin en önemli alanlarından biri olan Sosyal psikoloji, toplumda hangi tür etkiler altında kalıp davranışlarımızı şekillendirdiğimize öğretir. Bireye ait bir özellik olan tutumu etkileyen faktörleri tanıma, tutumlarımızı hangi durumlarda değiştirdiğimiz çok önemlidir. Toplumsal olayların dalgalanmalarının rastlantısal oluşları, gelecek hakkında karar vermemizi güçleştirmektedir. Çünkü her olayda ilk bakışta gözlenemeyen bir düzen, ayrıca bu olayı doğuran bir veya birkaç etken(faktör) vardır.

İşte sağlık kurumlarında bilhassa hekimlerimize yapılan şiddet; hasta, hasta yakınları ya da diğer başka bir bireyden gelen, sağlık çalışanı için risk oluşturan sözel ya da davranışsal tehdit, fiziksel saldırı veya cinsel saldırıları içermektedir. Son yıllarda dünyanın her yerinden

yaş, cinsiyet, ırk, din, dil, eğitim düzeyi ayırt etmeksizin toplumdaki bütün bireyleri etkileyen şiddet, giderek günlük yaşamımızın bir parçası haline gelmekte, herkes ve her sektör için önemli bir sağlık sorunu olarak karşımıza çıkmaktadır. Sağlık alanında hizmet verenler diğer iş alanlarında çalışanlara göre 16 kez daha fazla saldırıya uğramaktadır(1).

2. MODEL

Özellikle nitel değişkenlerin söz konusu olduğu lineer modeller doğada karşımıza çok fazla çıkmaktadır. Nitel değişkenli lineer modellerin incelenmesinde(analizinde) kullanılan yöntem "Varyans Analizi(V.A.)" dendiğini biliyoruz. Klasik anlamda 1-faktörlü sınıflandırmadan n-faktörlü sınıflandırmaya genişletebileceğimiz durumlar olduğu gibi faktörlerin çapraz, iç-içe, çapraz + iç-içe olduğu durumlar ve verinin dengeli olup olmaması durumları ve en sonunda faktörlerin etkileşimli(çapraz modellerde) olup olmamaları durumlarında V. A. tekniği uygulanmaktadır. Önemli sorun nitel değişkenli modeller için oluşturulan

$$Y = X\beta + \varepsilon$$

Şeklinde gösterilen lineer modelden elde edilen

$$(X'X)b = X'Y \quad \text{veya} \quad (X'X)\hat{\beta} = X'Y$$

normal denklemlerinde  $X'X$  matrisinin rankı tam olmayıp(not full rank) klasik anlamda tersinin hesaplanamaması. Bu ve benzeri matrislerin son zamanlarda ortaya atılmış olan "genelleştirilmiş ters matrislerle çözümleme" yöntemiyle kısıtlamalara gidilmeden bir çözüm mümkün olabilmektedir. Çalışmamızda veri tablosu için

düşündüğümüz model 2 – faktörlü çapraz sınıflandırma modelidir. Uygulama için birinci faktör olarak hekimler(doktorlar)(kamu ve özel kurumlarda çalışan) ikinci faktörün hastaların sosyolojik durumlarıdır. Kamuda çalışan, araştırma hastaneleri, üniversite hastaneleri, devlet hastaneleri ve sağlık ocaklarında çalışan hekimler ile yabancı sermayeli hastanelerde, Türk sermayesiyle faaliyet gösteren hastanelerde, özel vakıf hastanelerinde vb. kurumlarda çalışan başta hekimlerin olmak üzere karşılaştıkları şiddetin düzeylerini, uzun bekleme süreleri, hasta ve yakınlarının aşırı istekte bulunması, eğitim düzeylerinin düşük olması, stresli hasta yakınları, uzun çalışma koşulları, yanlış anlama gibi iletişim problemleri, personel yetersizliği ve yorgunluğu, yetersiz güvenlik ve polis desteği, kriz yönetmede yetersizlik vb. ayrıca, fiziksel şiddet; silahla tecavüz, ölümlü şiddet, kesici aletle yaralama, sözlü hakaret, kalabalık ve gürültülü ortamlar nedeniyle şiddet uygulamalarına rastlanmaktadır. Hastaların ve/veya hasta yakınlarının hangi sosyolojik nedenlerle bu şiddete başvurduklarını araştırmayı düşündük.

ABSTRACT

**Keywords:** Variance analysis, The 2-Way Crossed Classification, without interaction, factorial design, workplace violence, g-inverse.

Social psychology, which is one of the most important branches of the social sciences, teaches that under which effects we are changing our behaviours. The identification of the factors which affects behavior which is a characteristic of the individual and when we do change our behaviours is very important. The fact that the floating of the social events is random makes it difficult for us to decide about future. Because, in every event there is an order which can't be observed at first sight and also there are one or more factors which leads that event.

The violence which was done in health institutions especially to our doctors includes verbal or behavioural threats, physical attacks or sexual attacks which comes from patients, patient relatives or another individual and which forms a risk for the health worker. In last years, violence which affects all individuals in the society regardless of age, sex, race, religion, language and education levels is gradually becoming a part of our life and faces us as an important health problem for everybody and every sector. The health workers are attacked 16 times more than the individuals which work at the other areas (1).

Kaynaklar(References)

- (1) Behçet Al, Suat Zengin, Yahya Deryal, Cem Gökçen, Demet Arı Yılmaz, Cuma Yıldırım(2012), Increased Violence Towards Health Care Staff(Sağlık Çalışanlarına Yönelik Artan Şiddet), the journal of academic emergency medicine.
- [2] Gordon Lee Gillespie, PhD RN PHCNS-BC • Donna M. Gates, EdD RN FAAN • Margaret Miller, EdD CNS RN • Patricia Kunz Howard, PhD RN CEN FAEN(2010)  
) ; "Workplace Violence in Healthcare Settings: Risk Factors and Protective Strategies", Rehabilitation Nursing • Vol. 35, No. 5 • September/October.
- [3] S. R. Searl(1971), Linear models, John Wiley&Sons

## MEDYANA DAYALI PARAMETRİK OLMAYAN TAHMİN YÖNTEMİ İLE POLİNOMİAL MODELİN TAHMİNİ

Alper SİNAN<sup>1</sup>, Aşır GENÇ<sup>2</sup>

<sup>1</sup>Sinop Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Sinop

<sup>2</sup>Selçuk Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Konya

E-mail :alpsin@sinop.edu.tr, Email:agenc@selcuk.edu.tr

### 1. Giriş

Parametrik olmayan yaklaşım regresyon fonksiyonunun bilinmeyen formda olması veya, parametrik modelde varsayımların sağlanmaması durumunda da kullanılan tekniklere verilen ortak bir isimdir. Parametrik olmayan tahmin yöntemlerinden biri olan medyana dayalı tahmin yöntemi bağımlı değişken ile bağımsız değişken arasındaki bağlantıyı doğrusal bir model yardımıyla açıklamayı amaçlar. Literatürde basit doğrusal model için kullanılan Brown-Mood ve Theil yöntemleri mevcuttur. Bu çalışmada; bazı veri setleri arasındaki bağlantının polinom biçiminde bir eğri ile açıklanmasının daha uygun olacağı düşünülerek yeni bir medyan tahmini yöntemi geliştirilmiştir.

### 2. Brown- Mood ve Theil Yöntemi

Bu yöntemler basit doğrusal regresyon modelini ele alan, temelde birbirine yakın yaklaşımlardır. Basit doğrusal regresyon modeli

$$Y_i = \mu + \beta x_i + \varepsilon_i, \quad i = 1, 2, \dots, N \quad (1)$$

biçiminde olsun.  $Y_i$ ; bağımlı değişkenler,  $x_i$ ; bağımsız değişken,  $\mu$  ve  $\beta$  parametrelerdir.  $N$  adet gözlem çiftlerinden oluşan örneklem  $(x_1, y_1), \dots, (x_N, y_N)$  çiftleri ile gösterilsin. Brown-Mood yönteminde amaç grafik üstünde bu gözlem çiftlerini göstererek elde edilen noktalara en yakın doğruyu geçirmektir. Bu yöntemde veriler gruplanarak meydanları yardımıyla önce doğrunun geçtiği noktalar tespit edilir ardından noktalar yardımıyla doğru denklemini ve buna bağlı olarak  $\mu$  ve  $\beta$  parametre değerleri bulunur. Theil yönteminde ise gözlem değerlerinin medyanı yerine her bir gözlem çifti ile oluşturulacak modelden elde edilen tahmin değerlerinin medyanı kullanılır (Okur 2009).

### 3. Doğrusal ilişkiye sahip olmayan veriler için medyan tahmini

Brown-Mood ve Theil yöntemleri değişkenler arasındaki bağlantıyı bir doğru yardımıyla açıklamaktadır. Bazı veri setleri arasındaki bağlantının polinom biçiminde bir eğri ile açıklanmasının daha uygun olacağı düşünülerek çalışmamızda yeni bir medyan tahmini yöntemi geliştirdik. Veri setindeki gözlemlerin serpilme diyagramı çizildiğinde bağımlı değişken  $y_i$ ' ler ile bağımsız değişken  $x_i$ ' ler arasındaki ilişkinin doğrusal olmayan bir biçimde, ikinci dereceden bir polinom ile ifade edilebilecek biçimde ise regresyon model eşitlik 2' deki gibi kurulur.

$$Y_i = \mu + \beta_0 x_i + \beta_1 x_i^2 + \varepsilon_i, \quad i = 1, 2, \dots, N \quad (2)$$

Bu model üzerinden parametre tahminini yapmak için önerilen yöntem Brown-Mood yöntemindeki gibi verilerin bölgelere ayrılması ile ilgilendiği gibi Theil yöntemindeki gibi olası tüm parametrelerin hesaplanarak medyanının alınması fikrini de kullanmaktadır (Sinan, 2010).

### 4. Uygulama

Medyana dayalı tahmin yöntemlerinden en çok kullanılan yöntem olan Brown-Mood yöntemi ile yeni önerilen medyan yöntemini karşılaştırmak amacı ile karşılaştırılacak farklı durumlar düşünülerek 25'er adet gözlem çiftinden oluşan 6 farklı yapay veri seti oluşturulmuştur. Elde edilen 6 farklı durumun her biri için Brown-Mood ve önerilen yöntemin tahmin sonuçları karşılaştırılmıştır.

### 5. Sonuç

Değişkenler arasındaki ilişkinin doğrusal olması durumunda Brown-Mood yöntemi ile önerilen medyan yöntemi birbirlerine yakın sonuçlar vermektedir ancak bağımlı değişken ile bağımsız değişken arasındaki ilişkinin doğrusal olmadığı durumlarda, önerilen medyan yönteminin klasik yöntemlere göre daha iyi sonuç verdiği grafikler ve hata kareler ortalamaları yardımıyla gösterilmiştir.

### KAYNAKLAR

- [1] Brown, G., W., and Mood, A. M. (1951), "On Median Tests for Linear Hypotheses," Proceedings Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 159-166.
- [2] Miller J.C. and Miller, J.N., (1993), "Statistics for Analytical Chemistry", Ellis Horwood PTR Prentice Hall, Analytical Chemistry Series, 3rd ed. pp 159-161.
- [3] Okur, S., (2009), "Parametrik Ve Parametrik Olmayan Basit Doğrusal Regresyon Analiz Yöntemlerinin Karşılaştırmalı Olarak İncelenmesi" Yüksek Lisans Tezi, Çukurova Üniv. Fen Bilimleri Enstitüsü.
- [4] Sinan, A., (2010), "Panel Verili Yarıparametrik Regresyon Modelleri" Doktora Tezi, Selçuk Üniv. Fen Bilimleri Enstitüsü.

### ABSTRACT

#### THE ESTIMATION OF POLYNOMIAL MODEL WITH NONPARAMETRIC ESTIMATION METHOD BASED ON MEDIAN

The nonparametric estimation method based on median is a useful method for estimating a simple linear regression model when the assumptions of parametric model is not provided. Also it is easy to apply and robust for outliers in the data set. In this study a new nonparametric estimation method based on median suggested for the estimation of polynomial regression models. The results of two methods are compared on six different data sets and it is shown that new method is more useful for the estimation of the polynomial regression models.

**Key Words:** nonparametric regression, median, polimial regression model



## SINIFLANDIRMA AMAÇLI DESTEK VEKTÖR MAKİNELERİNİN LOJİSTİK REGRESYON İLE KARŞILAŞTIRILMASI

Furkan BAŞER<sup>1</sup>, Ayşen APAYDIN<sup>2</sup>

<sup>1</sup>Gazi Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Uluslararası Ticaret Bölümü 06830-Gölpazarı, Ankara, <sup>2</sup>  
Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü 06100-Tandoğan, Ankara

E-mail: furkan.baser@gmail.com<sup>1</sup> E-mail: apaydin@science.ankara.edu.tr

### 1. Giriş

Uygulamalı istatistikte, gözlemlerin ayrık iki kategoride sınıflandırılması biçimindeki problemler ile sıklıkla karşılaşmaktadır. Varsayımların doğrulanması durumunda sınıflandırma yöntemleri, yeni gözlemleri ait olduğu gruplara, eğitim örnekleri yardımıyla bir ayırıcı fonksiyona bağlı olarak atanmasını sağlar. Lojistik Regresyon (LR) çözümlemesi nesnelerin sınıflandırılması problemlerinde günümüzde yaygın olarak kullanılan bir yöntemdir.

Destek Vektör Makineleri (DVM), teorik çözümlemeler ile sayısal algoritmaları yapısında birleştiren bir sınıflandırma ve regresyon yöntemidir. İstatistiksel öğrenme teorisinde bu teknik, deneysel risk minimizasyonundan (Empirical Risk Minimization – DRM) ziyade yapısal risk minimizasyonuna (Structural Risk Minimization – YRM) dayandırılan bir öğrenme algoritması olarak geliştirilmiştir. YRM tümevarım prensibi, sonlu örneklem için Vapnik–Chervonenkis (VC) boyutuna bağlı olarak optimum model karmaşıklığını belirlemek üzere biçimsel bir mekanizma sağlar. Klasik sinir ağları ile karşılaştırıldığında DVM, bir tek global optimum çözüm elde edebilir ve boyut sorunu ile karşılaşmaz. Bu ilgi çekici özellikleri DVM'yi sıklıkla tercih edilir bir teknik haline getirmektedir. DVM ilk olarak örüntü tanıma problemlerini çözmek üzere tasarlanmıştır [1, 2]. Vapnik'in  $\epsilon$ -duyarsız kayıp fonksiyonunun ortaya atılması ile birlikte DVM, fonksiyon yakınsama ve regresyon model tahmin problemlerine de genişletilmiştir [3, 4].

Bu çalışmada, sınıflandırma problemlerinde DVM yöntemi ile LR çözümlemesinden elde edilen bulgular hatalı sınıflandırma sayılarına göre karşılaştırılacaktır. Bu amaçla, çeşitli doğum karakteristiklerine bağlı olarak düşük doğum ağırlıklı bebeklerin kestirim problemlerine ilişkin bir veri seti kullanılacaktır.

### 2. Sınıflandırma Amaçlı Destek Vektör Makineleri

$\mathcal{X}$  girdi örüntülerinin uzayını (örneğin  $\mathfrak{R}^d$ ) göstermek üzere; eğitim verisi  $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\} \subset \mathcal{X} \times \mathfrak{R}$  biçiminde ele alınsın. Eğitim verisini,

$$D(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \quad (1)$$

hiperdüzlem karar fonksiyonu tarafından uygun  $\mathbf{w}$  ve  $b$  katsayıları ile sınıflandırıldığı varsayılınsın. DVM yaklaşımına göre sınıflandırma problemlerinde  $L_{\Delta}(y, f(\mathbf{x}, \omega)) = \max(1 - yf(\mathbf{x}, \omega), 0)$  kayıp fonksiyonuyla deneysel risk,

$$R_{\text{emp}}(\omega, Z_n) = \frac{1}{n} \sum_{i=1}^n L_{\Delta}(y_i, f(\mathbf{x}_i, \omega)) \quad (2)$$

ile belirtilir. Diğer benzer sınıflandırma yöntemlerinde olduğu gibi DVM'de, sayısal optimizasyona uygun bir kayıp fonksiyonu kullanarak sınıflandırma hatasını minimum yapmaya çalışır.  $C$  yeterince büyük bir sabit olmak üzere programlama problemi,

Amaç fonksiyonu:

$$\min_{\mathbf{w} \in \mathcal{X}, b \in \mathfrak{R}} \left[ \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{n} \sum_{i=1}^n \xi_i \right]$$

Kısıtlar:

(3)

$$y_i [(\mathbf{w} \cdot \mathbf{x}_i) + b] \geq 1 - \xi_i, \quad i = 1, \dots, n$$

biçiminde oluşturulur [5]. Bu yapıda kullanıcı tarafından belirlenmesi gereken  $C$  katsayısı, karmaşıklık ile ayrılabilir olmayan örneklerin miktarı arasındaki değişimi kontrol eder.

### KAYNAKLAR

- [1] Chen, R. C. and Hsieh, C. H. (2006), Web page classification based on a support vector machine using a weighted vote schema, *Expert Systems with Applications*, 31(2), 427–435.
- [2] Tsujinishi, D. and Abe, S. (2003), Fuzzy least squares support vector machines for multiclass problems, *Neural Networks*, 16(5–6), 785–792.
- [3] Wu, Q. (2009), The forecasting model based on wavelet m-support vector machine, *Expert Systems with Applications*, 36(4), 7604–7610.
- [4] Dong, H., Yang, S., Wu, D. (2007), Intelligent prediction method for small-batch producing quality based on fuzzy least square SVM, *Systems Engineering-Theory and Practice*, 27(3), 98–104.
- [5] Cherkassky, V. and Mulier, F. (2007), *Learning From Data: Concepts, Theory, and Methods*, New Jersey, John Wiley & Sons.

### ABSTRACT

#### COMPARISON BETWEEN LOGISTIC REGRESSION AND SUPPORT VECTOR MACHINES FOR CLASSIFICATION PURPOSES

The classification of individuals is a common problem in applied statistics. For binary classification problems, logistic regression (LR) is one of the most widely used classification methods. More recently, Support Vector Machines (SVM) has been receiving considerable attention in pattern recognition and regression function estimation problems. In this paper, the fundamentals of LR and SVM are described, and the question of which one is better to discriminate is investigated. An application with real data associated with giving birth to a low birth weight baby is presented as an illustration.

**YAPISAL EŞİTLİK MODEL ANALİZİ: PREHİPERTANSİF BİREYLERDE, ANTROPOMETRİK BİYOKİMYASAL VE EKOKARDİYOGRAFI PARAMETRELER ÜZERİNE BİR UYGULAMA**

Birol TOPÇU<sup>1</sup>, Sinan SARAÇLI<sup>2</sup>, Şeref ALPSOY<sup>3</sup>, Aydın AKYÜZ<sup>3</sup>, Mehmet YILMAZ<sup>4</sup>, Cengiz GAZELOĞLU<sup>5</sup>

<sup>1</sup>Yrd.Doç.Dr., Namık Kemal Üniversitesi, Tıp Fakültesi Biyoistatistik Anabilim Dalı, Tekirdağ,

<sup>2</sup>Yrd.Doç.Dr., Afyon Kocatepe Üniversitesi, Fen-Ed. Fakültesi İstatistik Bölümü, Afyonkarahisar,<sup>3</sup>

Yrd.Doç.Dr., Namık Kemal Üniversitesi, Tıp Fakültesi Kardiyoloji Bölümü, Tekirdağ, <sup>4</sup> Doç.Dr., Ankara

Üniversitesi, Fen Fakültesi İstatistik Bölümü, Ankara, <sup>5</sup>MSc. Namık Kemal Üniversitesi Tıp Fakültesi

Biyoistatistik Anabilim Dalı, Tekirdağ

E-mail: [topcubirol@gmail.com](mailto:topcubirol@gmail.com)

**Giriş**

Yapısal Eşitlik Modeli (YEM), son yıllarda birçok bilim alanında kullanılmaya başlanmış bir model olmakla beraber tıp alanında çok fazla kullanılmamaktadır. Ancak bazı çalışmalarda literatürde mevcuttur. Örnek olarak; insülin tedavisi altındaki hastalarla yapılan bir çalışmada, hipoglisemiyi önlemede en önemli faktörün farkında olmak olduğu bulunmuştur [2]. Başka bir çalışmada ateroskleroz risk faktörlerinin, yaşam tarzı değişkenleri (sigara, alkol ve egzersiz) ile doğrudan veya dolaylı olarak ilişkili olduğu ve bu risk faktörlerinden etkilendiği gösterilmiştir [3].

YEM; modelde yer alan tüm gözlenen ve gözlenmeyen değişkenler arasındaki ilişkileri aynı anda test edebilen, faktör analizi ve çoklu regresyon analizinin birleşimi niteliğinde kapsamlı bir istatistiksel yöntemdir [1].

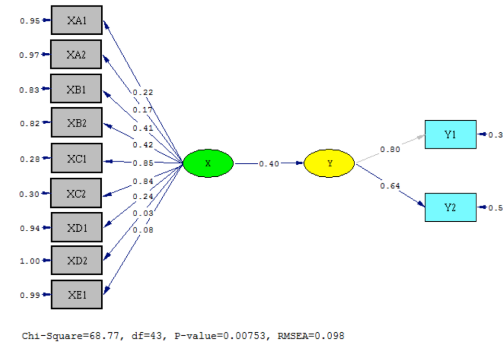
Bu çalışmada bizde yapısal eşitlik modellemesi ile prehipertansif bireylerde, antropometrik, biyokimyasal ve ekokardiyo grafi parametrelerinin arasındaki etkileri belirlemeyi amaçladık.

**Uygulama**

Çalışmaya, Mart 2010 ile Kasım 2011 zaman aralığında Namık Kemal Üniversitesi Araştırma ve Uygulama Hastanesi Kardiyoloji kliniğine gelen ardışık 3000 hasta arasından prehipertansif olduğu tespit edilen 60 hasta alınmış ve bu bireylerden elde edilen verilere ilk olarak Açıklayıcı Faktör Analizi uygulanmıştır.

Bazal özellikler ve biyokimyasal parametreler ölçüldükten sonra yapısal eşitlik modeli için 5 faktör oluşturulmuştur. Faktör 1 lipid faktörü olarak adlandırıldı ve total kolesterol, trigliserid ve düşük dansiteli lipoprotein (LDL) kolesterolü içermektedir. Benzer biçimde, faktör 2 şeker faktörü hemoglobin A1c (HgbA1c) ve açlık kan şekeri; faktör 3 antropometrik faktör bel çevresi ve kalça çevresini; faktör 4 eko faktörü sol ventrikül diastolik çap ve sol ventrikül ejeksiyon fraksiyonunu; faktör 5 böbrek faktörü serum kreatinin, düzeylerini ifade etmektedir.

5 faktörden oluşan kan değerleri ile tansiyon (diastolik ve sistolik tansiyon) arasındaki ilişkileri belirlemek için kurulan YEM, Şekil 1'de verilmiştir.



**Şekil 1. Hastaların kan ve tansiyon ölçümlerinin yapısal eşitlik modeli**

Kurulan Yapısal Eşitlik Modelinin uygunluğu incelenecek olursa, Ki-kare/sd = 68.77/ 43 = 1.599 değeri kritik değer olarak kabul edilen 3 değerinden küçük olduğundan ve RMSEA (Root Mean Square Error Approximation) değeri =0,098, Kabul edilebilir değerler olan  $0.05 \leq RMSEA \leq 0,10$  aralığında yer aldığından Kurulan modelin uygun olduğu söylenebilir. [4]

EFA analizinde yer alan XA3 (total kolesterol) değişkeni YEM de anlamlı çıkmadığından analizden çıkartılmış ve ilgili sonuçlar elde edilmiştir.

XD2 (EF) ve XE1 (Creatin) değerlerinin ise kan değerleri üzerinde, sırası ile 0.03 ve 0.08 katsayısı ile en az etkiye sahip değişkenler olduğu söylenebilir. EF ekokardiyo grafi bir parametredir ve prehipertansiflerde metabolik sendrom ve obezitenin fazla olduğu düşünülürse EF nin bunlarda yüksek olması vücudun sempatik yük altında olduğunun bir göstergesidir. Creatinin değerlerinin yüksek olması da ilerde gelişecek hipertansiyon ve buna bağlı renal hasar gelişeceğinin bir göstergesi olabilir.

Kan değerleri tüm alt boyutlarıyla birlikte tansiyon üzerinde pozitif yönlü bir etkiye sahip olarak kan değerlerindeki bir birimlik artışın tansiyon değerlerinde 0.40 birimlik bir artışa sebep olabileceği söylenebilir. AKŞ, HbA1C, total ve LDL kolesterol ve trigliserit ve kreatinin değerlerinde yükselme prehipertansif dönemde başlamaktadır. Bu da bize bu dönemde yaşam şekli değişikliği ile önlem almamız gerektiğini düşündürmektedir.Y (Tansiyon) değişkeni üzerinde ise Sistolik tansiyon değeri 0.80 katsayısı ile diastolik tansiyon değerinden daha etkili olarak belirlenmiştir. Prehipertansiflerde daha çok sistolik tansiyon yüksekliği ön plandadır.

**KAYNAKLAR**

- [1] Hair J., Anderson R., Black W., (1998). Multivariate data analysis with readings, 5<sup>th</sup> ed., Prentice-Hall.
- [2] Hepburn D.A., Deary I.J., MacLeod K.M., Frier B.M., (1994). Structural equation modeling of symptoms, awareness and fear of hypoglycemia, and personality in patients with insulin-treated diabetes, Diabetes Care. 17 (11):1273-80.
- [3] Oh H.S., Seo W.S., (2001). Development of a structural equation model for causal relationships among arteriosclerosis risk factors, Public Health Nurs. 18 (6): 409-17.
- [4] Schermelleh-Engel, K. and Moosbrugger, H., (2003), Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures, Methods of Psychological Research Online, 8 (2), 23-74.

### TEK YARDIMCI DEĞİŞKENLİ ORAN TAHMİN EDİCİLERİNİN TAYLOR SERİSİ YAKLAŞIMIYLA KARŞILAŞTIRILMASI VE BİR UYGULAMA

Tolga ZAMAN<sup>1\*</sup>, Vedat SAĞLAM<sup>1</sup>, Erdiç YÜCESOY<sup>1</sup>, Murat SAĞIR<sup>1</sup>

<sup>1</sup>Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Ana Bilim Dalı, 55139 ,Samsun, TÜRKİYE

<sup>2</sup>Amasya Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Ana Bilim Dalı, Amasya, TÜRKİYE

E-mail: zamanaman61@hotmail.com

Bu çalışmada, basit rasgele örnekleme yönteminde geliştirilen bazı tek yardımcı değişkenli oran tahmin edicilerinin hata kareler ortalamaları Taylor serisi yaklaşımı kullanılarak incelenmiştir. Taylor serisi yaklaşımı doğrusal olmayan bir tahmin ediciyi gözlemlerin doğrusal bir fonksiyonuna yaklaştırıp, bu doğrusal yaklaşıma uygun bir varyans eşitliği yazar. Bu doğrusallaştırma işlemi Taylor serisi yaklaşımıyla elde edilerek tahmin edicinin varyansı hesaplanır. Daha sonra bu yöntem kullanılarak elde edilen bu tahmin edicilere ait hata kareler ortalamaları dikkate alınarak karşılaştırmalar yapılmış ve belirli koşullar elde edilmiştir. Bu koşullar altında hangi tahmin edicinin hata kareler ortalamasının daha küçük, dolayısıyla daha etkin olduğuna karar verilmiştir.

Uygulamada, Trabzon'daki 18 ilçede bulunan 111 tane ortaöğretim okuluna ait veriler kullanılmıştır. Bu veriler Trabzon Valiliği İl Millî Eğitim Müdürlüğünden alınan gerçek verilerdir. Burada yardımcı değişken öğrenci sayısı ve ilgilenilen değişken öğretmen sayısı olarak alınmıştır ve öğrenci sayısı yardımıyla ortalama öğretmen sayısı oran tahmin edicileri ile tahmin edilmiştir. Burada kitleden örnekleme seçme işlemi rastgele sayılar tablosu kullanılarak yapılmış, dolayısıyla örnekleme alınan birimlere eşit şans verilerek kitledeki değişkenli örneklemede korunması sağlanmıştır.

**Anahtar Kelime:** Basit Rasgele Örnekleme; Taylor Serisi Yaklaşımı; Ortalama; Oransal Tahmin Edici; Oran; Çarpımsal Tahmin Edici; Yardımcı Bilgi; Hata Kareler Ortalaması.

### COMPARISON OF RATIO ESTIMATORS WITH SINGLE AUXILIARY VARIABLE BY TAYLOR SERIES APPROACH AND AN APPLICATION

**Abstract:** In this study mean square errors of ratio estimators with a auxiliary single variable are found with Taylor series approach. Taylor series approach provides a suitable variance equation by approaching a non-linear estimator to a linear function of linear observations. This method of linearization is provided by Taylor series approach and the variance of the estimator is calculated. After having mean square errors of the estimators, some comparisons are done and some conditions are given considering these mean square errors. Under these conditions it is decided which of the mean square error is smaller and yet which one is more effective.

In the application, datas of 111 junior schools in the counties of Trabzon is used. These actual datas are provided by the Trabzon National Education Management of Trabzon Governorship. In these datas the auxiliary is the number of students and the considered variable is the number of teachers. The number of teachers is estimated via the number of students by ratio estimators method. In the application random sampling choosing is done by the table of random numbers.

**Keywords :** Simple Random Sampling; Taylor Series Approach; Mean; Ratio Estimator; Proportion; Product Estimator; Auxiliary Information; Mean Square Error.

### KAYNAKLAR:

- [1] Kadılar, C., Çıngı, H., 2003. A study on the chain ratio-type estimator, Hacettepe Journal of Mathematics and Statistics, 32, 105-108.
- [2] Kadılar, C., Çıngı, H., 2006. New Ratio Estimators Using Correlation Coefficient, İnterstat 4, pp, 1-11.
- [3] Naik, 6., V. D., Gupta, P. C., 1996. A note on estimation of mean with known population proportion of an auxiliary character. Jour. Ind. Soc. Agr. Stat., 48(2), 151-158.
- [4] Ortaöğretim Okulları, Öğretmen ve Öğrenci Sayıları, 2012. Trabzon Valiliği İl Millî Eğitim Müdürlüğü.

### DÜZCE ÜNİVERSİTESİ ARAŞTIRMA HASTANESİNDE TEDAVİ GÖREN ANKİLOZAN SPONDİLİTLİ HASTALARIN YERSEL KÜMELENMELERİNİN İNCELENMESİ

Özge PASIN<sup>1</sup>, Handan ANKARALI<sup>2</sup>, Şengül CANGÜR<sup>3</sup>

Düzce Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD, 81620, Düzce, TÜRKİYE,  
<sup>1</sup>ozgepasin90@yahoo.com.tr, <sup>2</sup>hankarali@yahoo.com, <sup>3</sup>sengulcangur@duzce.edu.tr

Konuma dayalı gözlemlerle elde edilen grafik ve grafik-olmayan bilgilerin toplanması, saklanması, işlenmesi ve kullanıcıya sunulması işlemlerini bütünlük içerisinde gerçekleştiren coğrafi bilgi sistemi, son yıllarda sağlık alanı araştırmalarında ve özellikle hastalık dağılım haritalandırılmasında, hastalıkların görülme sıklığının belirlenmesinde, hastaların takibinde, sağlık servislerine coğrafi ulaşımı öğrenmede, personel yönetiminde, bölgesel hastalık analizlerinde, sağlık hizmetlerinin dağılımının belirlenmesi gibi konularda kullanılmaktadır. Ancak bu konuda ülkemizde sınırlı bölgesel çalışmalar mevcuttur. Bu çalışmada, Düzce Üniversitesi Tıp Fakültesi Fizik Tedavi ve Rehabilitasyon polikliniğine tanı veya tedavi amacıyla çeşitli tarihlerde başvuran 52 ankilozan spondilitli hastadan, 2013 yılında toplanan veriler yardımıyla hastaların Düzce ilinde belirli bir alanda yersel kümeleme gösterip göstermediklerinin araştırılması amaçlanmıştır. Bu amaçla elde edilen veriler ARCGIS 10 programına aktarılmıştır. Hastaların yersel kümelemeleri, programda yer alan Moran I, Getis ORD (General G), En yakın komşu yöntemi (Average nearest neighbour) ve Ripley'in K fonksiyonu yöntemleri ile incelenmiştir. Bu yöntemler yere göre global kümeleme yöntemleridir. Moran I yöntemi, olgular arasındaki mekânsal otokorelasyonu ölçen bir sistemdir. Getis ORD ise, mekânsal alanda yüksek değerlerin (olguların) mi yoksa düşük değerlerinin mi kümelendiğinin tespitinde kullanılan bir yöntemdir. En yakın komşu yöntemi de noktaların en yakın komşu uzaklıklarından yararlanarak kümelemelerini araştırmaktadır. Ripley'in K fonksiyonu ise, coğrafi bölgeye ait nokta veriler ile mekânsal yapıyı analiz etmek için kullanılmaktadır. Bu yöntem belirli bir mesafe ile diğer olgular arasında ne kadar olgu/vaka bulunduğunu ölçer ve sadece en yakın komşu mesafesi ile değil tüm yapılar arasındaki mesafeleri olarak mekânsal bağımlılıklarla ilgili bir özet sunar. Çalışmada yersel kümeleme için yukarıda kısaca tanımlanan dört yöntem kullanılarak, elde edilen sonuçlar karşılaştırmalı olarak yorumlanmıştır. Bulgular değerlendirildiğinde, Düzce il merkezinde anlamlı düzeyde bir kümelemenin olduğu belirlenmiştir. Ancak, gerçekleşen kümeleme, ankilozan spondilit hastalığının nadir görülmesi sebebiyle toplanan veri sayısının az olmasından ve sadece bir sağlık merkezine ait verilerin kullanılmasından dolayı gerçeği tam olarak yansıtmayabilir. Buna karşın verilerin toplandığı sağlık merkezinin, Düzce ilinde en büyük ve kapsamlı sağlık merkezi olması nedeniyle sonuçların, bir ön çalışma niteliğinde önem taşıdığı da göz ardı edilmemelidir. Sonuç olarak planlama, korunma ve tanı-tedavi işlemlerinde oldukça önemli rol oynayan hastalık, personel, sağlık merkezi gibi konularıyla ilgili kümeleme yapılarının incelenmesi önerilir. Bu kümelerin incelenmesi için gerekli veri, çeşitli sağlık kuruluşlarından temin edilebilir. Ancak sağlık kuruluşlarının, adres bilgileriyle birlikte düzenli kayıtlar tutması önemli bir konudur.

### KAYNAKLAR

- [1] ARCGIS 10 uygulama dokümanı (ESRI)
- [2] <http://www.biomedware.com/?module=Page&SID=clusterseer-help-and-tutorials>
- [3] Erdem Karabulut, Reha Alpar, Enis Özyayar, (2006), Hastalıkların Yere Göre Kümelemesinde Kullanılan Yöntemler, İnönü Üniversitesi Tıp Fakültesi Dergisi 13(1) 37-43
- [4] G. Özkan, H.C. Güngör, 2007, Cbs'nin Sağlık Alanında Kullanımı ve Örnekleri, TMMOB Harita ve Kadastro Mühendisleri Odası Ulusal Coğrafi Bilgi Sistemleri Kongresi, KTÜ, Trabzon

### ABSTRACT

### INVESTIGATION OF SPATIAL CLUSTER OF ANKYLOSING SPONDYLITIS PATIENTS IN DUZCE UNIVERSITY RESEARCH HOSPITAL.

With the help of the data collected in Duzce University Physical Therapy and Rehabilitation Clinic, in 2013, 52 ankylosing spondylitis patients were investigated for spatial clusters. Spatial clustering was examined with Moran I, Getis ORD, Average nearest neighbor and Ripley's K function in ARCGIS 10 program. It was determined that a significant level of visualisation cluster in the centre of Duzce. However, due to the rarity of the disease ankylosing spondylitis, and used only one health center, clustering may not reflect exactly the data. But, because of the health center is largest institution, results can be evaluated a preliminary study.

**Keywords:** Geographic information system, spatial clustering methods, ARCGIS10, ankylosing spondylitis, disease cluster

## **BİLDİRİ OTURUMLARI 1**

### *SESSION 1*

*Finans, Sigortacılık, Risk Yönetimi*

## BİREYSEL EMEKLİLİK SİSTEMİNDE DEVLET KATKISININ ETKİSİNİN YAPISAL KIRILMA ANALİZİ İLE İNCELENMESİ

Ayşe İŞİ, Hakkı POLAT, Fatih ÇEMREK

\* Gazi Üniversitesi Gazi Meslek Yüksekokulu Çubuk Yerleşkesi Çubuk/Ankara, [isiayse@gmail.com](mailto:isiayse@gmail.com)

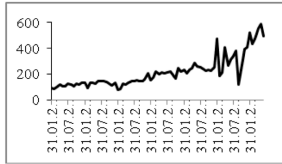
\*\*Eskişehir Osmangazi Üniversitesi Meşelik Kampüsü İstatistik Bölümü /Eskişehir, [hakkipl@gmail.com](mailto:hakkipl@gmail.com),  
[femrek@gmail.com](mailto:femrek@gmail.com)

### 1. Giriş

Bireysel Emeklilik Sistemi (BES), mevcut kamu sosyal güvenlik sisteminin tamamlayıcısı olarak kurulmuştur [3]. Bireysel Emeklilik Tasarruf ve Yatırım Sistemi Kanunu, sosyal güvenlik reformunun bir parçası olarak Ekim 2001 tarihinde yürürlüğe girmiştir. 29 Haziran 2012 tarih ve 28338 sayılı Resmî Gazete'de yayımlanan Bireysel Emeklilik Tasarruf ve Yatırım Sistemi Kanunu ile Bazı Kanun ve Kanun Hükmünde Kararnamelerde Değişiklik Yapılmasına Dair Kanun ile devlet katkısı sistemine geçilmiş ve böylece BES'te yeni bir dönem başlamıştır.

Bu çalışmanın amaçlarından biri, kanun değişikliği ile getirilen devlet katkısının BES üzerindeki etkisini incelemek, bir diğeri de yapısal kırılma içeren bir serinin durağanlık analizinde yapısal kırılmanın dikkate alınması gerekliliğini vurgulamaktır. Zira literatürde yapılan çalışmaların çoğunda zaman serilerinin durağanlığı incelenirken yapısal kırılmalar dikkate alınmamış ve aslında durağan olan serilere gereksiz fark alma işlemleri uygulanarak yanıltıcı sonuçlar elde edilmiştir. Bu amaçla 2007:01-2013:06 dönemine ilişkin aylık katkı payı tutarı (KP) (milyon TL) veri seti üzerinde çalışılmıştır.

2004:1-2013:6 dönemine ilişkin aylık katkı payı tutarları (milyon TL) serisinin zaman grafiği Şekil 1'de verilmiştir.



Şekil 1. Katkı Payı Tutarı (milyon TL) Serisinin Grafiği

Şekil 1 incelendiğinde BES katkı payı tutarı serisinin ortalamada ve varyansa durağan olmayan bir yapıya sahip olduğu, ayrıca serinin genel eğiliminde 2011:12 ve 2012:09 anlarında yapısal bir değişiklik (kırılma) olduğu görülmektedir. Özellikle 2012:09 noktasında görülen serideki sıçrama ve sonraki dönemde görülen artan trend, serinin 2012:09 anından sonra hem düzeyinde hem de eğiminde yapısal bir değişim olduğunu göstermektedir.

### 2. Bulgular

Serinin durağanlığı, Augmented Dickey Fuller (ADF) birim kök testi ile analiz edilmiş ve serinin tüm modeller için %1, %5 ve %10 anlamlılık seviyelerinde durağan olmadığı belirlenmiştir.

Zaman serilerinde durağan-dışılığın bir nedeni de serinin ana kütle regresyon denklemi boyunca farklı örneklemeler açısından değişiklikler (yapısal kırılmalar) göstermesidir. Genelde ekonomide yapısal kırılmaların bir nedeni olarak ekonomi politikalarındaki değişimler ile önemli olayların neden olduğu değişiklikler sayılabilir [2]. BES'teki devlet katkısı gelişmesi de bu değişimlerden biri olarak değerlendirilebilir.

BES'teki bu gelişmenin yapısal kırılmaya yol açıp açmadığının incelenmesi için serinin CUSUM ve CUSUM of squares grafikleri incelenmiş ve seriyeye Chow yapısal kırılma testi uygulanmıştır. Analizler sonucunda seride 2011:12 ve 2012:09 dönemlerinde yapısal kırılma belirlenmiştir. KP serisinde 2012:09 dönemindeki yapısal kırılmanın

serideki durağandışılığın bir nedeni olabileceği düşünülerek Perron (1989)'ün önerdiği Model C ile yapısal kırılma modelleri dikkate alınarak birim kök testi uygulanmış ve sonuçlar Tablo 1'de verilmiştir.

Tablo 1. Perron(1989) Modellerine İlişkin ADF Testi Sonuçları

Model	t- ist.	Olasılık (p) değeri	AIC
Toplamsal Sapmalı Model (AO)	-8,7416	0,000	10,603
Kademeli Sapmalı Model (IO)	-8,3633	0,000	10,677

Bu durumda seri gerçekte durağan iken serideki yapısal kırılmadan kaynaklanan bir sahte durağan-dışılığın olduğu değerlendirilmiştir.

### 3. Sonuçlar

Katkı payı tutarı serisine uygulanan analizlerin sonuçları değerlendirildiğinde KP serisinde 2012:09 döneminde yapısal bir kırılma tespit edilmiş ve bu kırılmanın seride sahte durağan-dışılığa neden olduğu belirlenmiştir. Haziran 2012 tarihinde Resmî Gazetede yayımlanarak yürürlüğe giren BES'te devlet katkısı içeren yasa değişikliği BES'i olumlu yönde etkilemiş ve 2012:09 döneminden itibaren katkı payı tutarlarında hızlı bir artışa yol açmıştır.

### KAYNAKLAR

[1] Perron, P.,(1989), The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, *Econometrica*, 57 (6): 1361-14013.

[2] Sevüktekin, M ve Nargeleçekenler, M. (2010), *Ekonometrik Zaman Serileri Analizi Eviews Uygulamalı*, Ankara, Nobel Yayınları.

[3] <http://www.egm.org.tr/>

### ABSTRACT

#### INVESTIGATING OF STATE CONTRIBUTION EFFECT OF PRIVATE PENSION SYSTEM VIA ANALYSIS OF STRUCTURAL BREAK

One of the purposes of this paper, to examine the impact on the state's contribution PPS sector and the other purpose is to emphasize the necessity of taking into account the structural break on the stationarity analysis of a series with structural breaks For this purposes, amount of monthly contribution for the period 2007:01-2013:06 (TRY million) was studied on the data set. As a result of analysis, series of the contribution amount has been identified structural break in 2012:09 and there have been a non-stationarity structure due to the structural break. The change of the new law has increased amount of contribution since the 2012:09.

**Key Words:** Private Pension System (PPS), structural breaking analysis, Chow test



## OTOKORELASYON VE ÇOKLU İÇ İLİŞKİ VARLIĞINDA ETKİNLİK ÖLÇÜMLERİ

Tuğba SÖKÜT AÇAR<sup>1</sup>, M. Revan ÖZKALE<sup>2</sup>

<sup>1</sup>Çanakkale Onsekiz Mart Üniversitesi, Fen Bilimleri Enstitüsü, 17020, Çanakkale, Türkiye, t.sokut@comu.edu.tr

<sup>2</sup>Çukurova Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, 01330, Adana, Türkiye, mrevan@cu.edu.tr

### 1. Giriş

$$y = \mathbf{1}\beta_0 + X\beta_1 + \varepsilon \quad (1)$$

ile verilen çoklu lineer regresyon modeli ele alınsın. Burada,  $\mathbf{y}$ :  $n \times 1$  tipinde rastgele değişkenlerin gözlenen yanıt vektörü,  $\mathbf{1}$ :  $n \times 1$  tipinde 1'lerden oluşan vektör,  $\beta_0$  ve  $p \times 1$  boyutlu  $\beta_1$  bilinmeyen parametreleri içeren vektör,  $X = (x_1, x_2, \dots, x_r)$ :  $n \times r$  tipinde merkezileştirilmiş ve standartlaştırılmış ( $\mathbf{1}'x_i = 0$ ,  $x_i'x_i = 1$ ,  $i = 1, 2, \dots, r$ ) bilinen açıklayıcı değişkenler matrisi ve  $\varepsilon$ :  $n \times 1$  tipinde  $E(\varepsilon) = 0$ ,  $E(\varepsilon\varepsilon') = \sigma^2 I_n$  varsayımları ile hataların gözlenmeyen vektörüdür.  $\beta' = (\beta_0 \beta_1)$  ve  $Z = [\mathbf{1} X]$ :  $n \times p$  ( $p = r + 1$ ) tipinde matris olmak üzere (1) modeli

$$y = Z\beta + \varepsilon \quad (2)$$

olarak yazılır. (1) modeline en küçük kareler (EKK) yöntemini uygulayabilmek için  $X$  matrisinin tam kolon ranklı olması gerekmektedir. Ancak, bir veya daha çok açıklayıcı değişkenin diğerlerinin doğrusal bir kombinasyonu olması durumunda ortaya çıkan çoklu iç ilişki varlığında  $X$  matrisi tam kolon ranklı değildir. Bu durumda  $X'X$  matrisi singüler olacağından dolayı EKK tahmin edici uygulanamaz. Sabit varyanslı ve ilişkisiz hataya sahip lineer regresyon modellerinde çoklu iç ilişki olması durumunda Hoerl ve Kennard (1970) EKK'ye alternatif olarak ridge tahmin ediciyi (1) modeli altında:  $\hat{b}_k = (X'X + kI)^{-1}X'y$ ,  $k \geq 0$  olarak önermiştir.

Regresyon analizinde önemli bir varsayım seçilen modelin verideki tüm gözlemler için uygun olmasıdır. Fakat uygulamada bir veya birden çok gözlem, verinin çoğunluğunun oluşturduğu modelden farklılık gösterebilir (Weisberg, 1985). Bu farklılıklar regresyon sonuçları üzerinde önemli bir etki yaratacağından dolayı tanılama kavramı lineer regresyon modellerinde önemli bir yer almaktadır. Kaldıraç (Leverage) ve Etkinlik (Influence) kavramları ilişkisiz hataya sahip lineer regresyon modelleri için sıklıkla ele alınan tanılama metodlarıdır; Belsley ve ark. (1980), Cook (1977). Etkinliğin ölçümünde kullanılan istatistikler; Cook uzaklığı (Cook, 1977), DFFITS ve DFBETAS (Belsley ve ark., 1980) iken kaldıraç nokta Şapka matrisi üzerinden belirlenir. Etkinlik ve kaldıraç nokta kavramını sabit varyanslı ve ilişkisiz hataya ve çoklu ilişkiye sahip modeller altında bazı yazarlar tarafından incelenmiştir: Walker ve Birch (1988), Özkale (2013), Steece (1986).

Bu çalışmada (2) ile verilen modelden farklı olarak hataların  $u_i = \rho u_{i-1} + v_i$ ,  $v_i \sim WN(0, \sigma_v^2)$  ile modellendiği

$$y = Z\beta + u, \quad E(u) = 0, \quad E(uu') = \Phi = \sigma^2\Psi \quad (3)$$

modeli ele alınmıştır. (3) modeli altında çoklu iç ilişki olması durumunda Trenkler (1984)'in önerdiği ridge tahmin edici  $\hat{\beta}(k) = (Z'\Psi^{-1}Z + kI_p)^{-1}Z'\Psi^{-1}y$ ,  $I_p = \text{diag}(0, 1, \dots, 1)$  ele alınmış ve  $\hat{\beta}(k)$  tahmin edici için COOK uzaklığı, DFFITS ve DFBETAS ölçümleri incelenmiştir. Elde edilen teorik sonuçlar simülasyon ile üretilen örnek üzerinden irdelenmiştir.

### KAYNAKLAR

- [1] Belsley D. A., Kuh E. and Welsch R. E., 1980. Regression Diagnostics, John Wiley, New York.
- [2] Cook R. D., 1977. Detection of Influential Observation in Linear Regression. Technometrics, Vol.19, No.1, 15-18.
- [3] Özkale M. R., 2013. Influence Measures In Affine Combination Type Regression, Journal of Applied Statistics, DOI: 10.1080/02664763.2013.809568.
- [4] Steece B. M., 1986. Regressor Space Outliers in Ridge Regression, Commun. Statist.-Theory Meth., 15(12), 3599-360
- [5] Trenkler G., 1984. On the Performance of Biased Estimators in the Linear Regression Model with Correlated or Heteroscedastic Errors, Journal of Econometrics 25, 191-203.
- [6] Walker E. and Birch J. B., 1988. Influence Measures in Ridge Regression, Technometrics, 30(2), 221-227.
- [7] Weisberg, S., 1985. Applied Linear Regression, 2nd ed., Wiley, New York.

### THE INFLUENCE MEASURES IN THE PRESENCE OF AUTOCORRELATION AND MULTICOLLINEARITY

Influence concepts have an important place in linear regression models. Case deletion is a useful method for assessing the influence of single case. Some of the influence measures (DFFITS and COOK's D) are discussed by Walker and Birch (1988), when the ridge regression used in case of multicollinearity. However, the errors structure in the linear regression models dealt with by Walker and Birch (1988) are uncorrelated. In this paper, the influence measures for ridge regression estimator with autocorrelated errors are discussed. To determine the behavior of the influential observations depending on ridge parameter and autocorrelation coefficient, a Monte Carlo simulation were applied.

**Key Words:** Autocorrelation, Influence, Leverage Point, Multicollinearity, Ridge Regression

## TÜRKİYE KREDİ TEMERRÜT TAKASI PRİMİNİN UZUN DÖNEM BAĞIMLILIK YAPISININ İNCELENMESİ VE MODELLENMESİ

Nurbanu BURSA\* ,Hüseyin TATLIDİL

Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Beytepe, Ankara, TÜRKİYE

E-mail: [nurbanubursa@hacettepe.edu.tr](mailto:nurbanubursa@hacettepe.edu.tr)

E-mail: [tatlidil@hacettepe.edu.tr](mailto:tatlidil@hacettepe.edu.tr)

### Giriş

Kredi türlerini adı altında, kredi riskinin çeşitli risk aktarım enstrümanları vasıtasıyla bu riskleri almak isteyenlere aktarılmasına olanak sağlayan yeni yöntemler geliştirilmektedir. Kredi türlerinin finansal piyasalarda en yaygın olarak işlem gören enstrümanı kredi temerrüt takası (credit default swap) yani CDS'tir. CDS'ler, bir alacaklımın, üçüncü bir kişiye belli bir ücret ödeyerek, alacağını garantilemesini sağlayan sözleşmelerdir. Bu riski, CDS satıcısı belirli bir ücret karşılığında üstlenir. Bu ücrette CDS primi denir. Primlerin miktarını borçlu olan ülkenin ya da şirketin iflas etme olasılığı belirler. İflas etme olasılığı arttıkça ödenecek primler de doğru orantılı olarak artar.

Gün geçtikçe hacim olarak önemli bir duruma gelmeye başlayan CDS'lerin bu yükselişinde, devletlerin ülke riskinin bir göstergesi olarak kullanılmalarının ve diğer risk göstergelerine göre daha doğru sonuçlar sunmalarının payı büyüktür. Ülke CDS'lerinin primi, eurobond ihraç eden ülkenin temerrüde düşmesi durumunda, söz konusu varlığı elinde tutanların önceden belirlenmiş bir miktarda tazminlerini sağlamak amacıyla, CDS satıcılarına ödedikleri, yükümlülük tutarının belirli bir yüzdesi olarak hesaplanan yıllık primlerdir. Bu haliyle CDS primleri, ülkeler için temerrüt riskinin doğrudan bir ölçütü olmaktadır. Bir anlamda bu primler, uluslararası piyasalarda o ülkenin borçlarını ödeyebilme kabiliyetine olan güveni göstermektedir.

Çalışmada 2851 gözlemden oluşan 01.11.2000 - 30.11.2012 tarihleri arasındaki Türkiye' nin 5 yıl vadeli CDS primi tek değişkenli zaman serisindeki uzun dönem bağımlılık yapısı trendden arındırılmış dalgalanma analizi (detrended fluctuation analysis-DFA) ile ortaya çıkarılmış, bu analiz yardımıyla belirlenen kesirli fark parametresiyle de ARFIMA modellemesi gerçekleştirilmiştir.

### Uzun Dönem Bağımlılık Yapısının İncelenmesi ve Modellemesi

Uzun dönem bağımlılık, bir sürecin otokorelasyon fonksiyonunda gözlenen hiperbolik biçimdeki yavaş düşüş olarak ifade edilmektedir. Bir zaman serisinin uzun dönem bağımlılık özelliğine sahip olması bu zaman serisinin uzak gözlemleriyle bile arasında bir ilişkinin olduğunu göstermektedir.

Uzun dönem bağımlı zaman serilerinin çözümlenmesinde doğrusal bir model ve ardışık bağımlı bütünleşik hareketli ortalamalar (ARIMA) modellerinin genişletilmiş bir hali olan ardışık bağımlı kesirli bütünleşik hareketli ortalamalar (ARFIMA) modelleri kullanılmaktadır.

ARIMA ( $p, d, q$ ) sürecinde  $d$  değeri için yalnız tamsayılar dikkate alınmaktadır. Ancak teorik olarak  $d$ 'nin değeri tamsayı ile sınırlı değildir. Doğrusal zaman serileri modeli olan ARFIMA ( $p, d, q$ ) modeli ile tamsayı olmayan  $d$  değerlerinin de kullanılabilmesi sağlanmıştır.

$X_t$  zaman serisi,

$$\Phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t \quad -0.5 < d < 0.5 \quad (1)$$

model denklemi ile ifade edilirse  $X_t$ ' ye ARFIMA süreci denir.

Bir ARFIMA sürecinde fark parametresi  $d$ ,  $[-0.5, 0]$  aralığında ise süreç kısa dönem bağımlılığa,  $[0, 0.5]$  aralığında ise de uzun dönem bağımlılığa sahiptir. Bir serinin uzun dönem bağımlılık özelliğinin incelenmesi için  $d$  kesirli fark parametresinin belirlenmesinde yinelenmiş ölçekli düzeltilmiş aralık istatistiği (R/S), Whittle yaklaşık en çok olabilirlik yöntemi ve dalgacık yöntemi gibi pek çok yöntem kullanılmaktadır. Son zamanlarda öne çıkan yöntemlerden biri de ölçekleme teorisi çerçevesinde geliştirilen DFA'dır. Bu analiz, zaman serilerindeki uzun dönem hafızanın varlığını ya da yokluğunu araştırmak için ölçekleme üsteli olarak bilinen sayısal bir parametre ( $\alpha$ ) sağlamaktadır. Bu yöntemle, durağan olmamaktan kaynaklı, sahte uzun dönem bağımlılıkların belirlenmesinden kaçınılmaktadır.

Aynı zamanda, DFA ile elde edilen ölçekleme üsteli  $\alpha$  ile  $d$  kesirli fark parametresi arasında aşağıdaki gibi bir ilişki bulunmaktadır.

$$d = \alpha - 0.5 \quad (2)$$

### KAYNAKLAR

- [1] Tatlıdil, H. ve Bursa, N. (2011). Kredi Temerrüt Takası ve Risk Yönetimi, İktisat ve Toplum Dergisi, 12, 58-65.
- [2] Leite, A., Rocha, A. P., Silva, M. E., Gouveia, S., Carvalho, J., Costa, O., (2007). Long-Range Dependence in Heart Rate Variability Data: ARFIMA Modelling vs Detrended Fluctuation Analysis, Computers in Cardiology, 34, 21-24
- [3] Günay, S., Eğrioglu, E., Aladağ, Ç. H. (2007). Tek Değişkenli Zaman Serileri Analizine Giriş, Hacettepe Üniversitesi Yayınları, Ankara.
- [4] Wang, Y., Wei, Y., Wu, C., (2011). Detrended Fluctuation Analysis On Spot and Futures Markets of West Texas Intermediate Crude Oil, Physica A, 390, 864-875.
- [5] Bektaş, C., (2007). Dünya Borsaları Getiri ve Mutlak Getiri Serilerinde Uzun Dönemli Bağımlılık Analizi, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.

### ABSTRACT

#### EXAMINING AND MODELLING THE STRUCTURE OF LONG-TERM DEPENDENCY OF TURKEY'S CREDIT DEFAULT SWAP SPREADS

Credit default swap (CDS) spreads which are known as indicator of country risk and modelling these spreads have great importance in terms of financial world. In this study, firstly, information is given about long-term dependency, ARFIMA models and detrended fluctuation analysis which can be called (DFA). Secondly, the structure of long-term dependency is revealed with DFA on Turkey's CDS spread time series from 2000 to 2012. In addition to this, mentioned univariate time series is modelled by ARFIMA.

**Key Words:** Credit Default Swaps, Detrended Fluctuation Analysis, ARFIMA.

### CVAR RİSK ÖLÇÜSÜNE DAYALI PORTFÖY OPTİMİZASYONU: DİNAMİK KOPULA MODELİ

Sibel AÇIK KEMALOĞLU\*

Emel KIZILOK KARA

Ankara Üniversitesi Fen Fakültesi  
Bölümü

Kırıkkale Üniversitesi Fen Edebiyat Fakültesi İstatistik  
İstatistik Bölümü

Tandoğan/ANKARA

Yahşihan/KIRIKKALE

[acik@science.ankara.edu.tr](mailto:acik@science.ankara.edu.tr)

[emel.kizilok@gmail.com](mailto:emel.kizilok@gmail.com)

#### Özet

Bu çalışma, çok değişkenli finansal verilerin bağımlılık yapısının kopula kullanılarak istatistiksel olarak modellenmesi ile ilgilidir. Finansal veriler, ekonomik faktörlerden çok etkilendiğinden çoğunlukla zamana göre değişkenlik gösterir. Bu nedenle, çalışmada zamana göre değişkenliği göz önüne alan dinamik kopula modeli kullanılmıştır. Ayrıca Monte Carlo simülasyonu ile Mean-CVaR modeline dayalı portföy optimizasyonu yapılmıştır. Uygulama olarak Türk finans piyasasındaki dört farklı endeks ile bir portföy oluşturulmuştur. Bu portföydeki varlıkların marjinal dağılımları tahmin edilerek farklı kopula modelleri için parametre tahminleri verilmiştir. Belirlenen kopula modelinden oluşturulan portföy için CVaR risk ölçüsüne dayalı portföy optimizasyonu yapılmıştır.

#### Giriş

Modern Portföy Teorisi, 1950'li yıllardan sonra ortaya çıkan ve varlıklar arasındaki ilişkileri göz önüne alarak portföy oluşturmaya dayalı bir yaklaşımdır. Markowitz, ortalama varyans modelini ortaya koymuş ve modern portföy teorisinin temellerini atmıştır[1]. Ortalama varyans modelinde portföyün riski varyans ile ölçülür. Getiri verilerinin normal dağılmadığı durumda riski varyans ile ölçmek yerine diğer risk ölçülerini kullanmak daha uygundur. Riske Maruz Değer (VaR -Value at Risk) ve Koşullu Riske Maruz Değer (CVaR -Conditional Value at Risk) finansal alanda en çok bilinen risk ölçüleridir. Finansal alanda yapılan ilk çalışmalarda risk ölçüsü olarak VaR kullanılmıştır. Ancak VaR, bir risk ölçüsünde istenen alt toplamsallık ve konvekslik özelliğini sağlamamaktadır. Özellikle finans ve portföy optimizasyonunda bu özelliklerin sağlanması önemlidir. CVaR bu özellikleri sağladığı için VaR'a alternatif olarak önerilmektedir[2]. Bu çalışmada risk ölçüsü olarak CVaR alınmış ve portföy optimizasyonu için Ortalama- CVaR (Mean-CVaR) modeli kullanılmıştır.

Portföy CVaR'ını daha doğru tahmin edebilmek için varlık getirilerinin kuyrukları arasındaki lineer olmayan bağımlılığı ortaya çıkarmak gerekmektedir. Bu çalışmada, getiri serileri arasında bağımlılık yapısını modellemek için ilk defa Sklar [3] da bahsedilen kopula modeli kullanılmıştır. Ancak kullanılan finans verileri zamana göre değişkenlik gösterdiğinden verilere bilinen kopula modeli yerine dinamik kopula modeli uygulanmıştır[4]. CVaR'a dayalı portföy optimizasyonu yapabilmek için kopula parametresinin tahmin edilmesi gerekmektedir. Bu amaçla iki aşamadan oluşan Marjinallere İlişkin Çıkarsama Fonksiyonları (IFM) yöntemi kullanılmıştır. Bu yöntemde, ilk aşamada marjinal dağılım parametre tahminleri, ikinci aşamada ise copula parametresinin tahmini yapılmaktadır[5]. Marjinal ve kopula uyum iyiliklerinin testi için AIC ve BIC kriterlerinden yararlanılmıştır. Böylelikle, portföyün bağımlılığının GARCH(1,1)-t dinamik kopula modeline uyduğu görülmüştür. Son olarak Monte Carlo simülasyon yöntemi ile bu modelden üretilen veriler için CVaR risk ölçüsüne dayalı portföy optimizasyonu yapılmış ve portföydeki varlıkların ağırlıkları belirlenmiştir.

#### KAYNAKLAR

[1] Markowitz H.(1952), Portfolio Selection, Journal of Finance, 25, 77-91.

[2] Rockafellar R. T. and Uryasev S. (2002), Optimization of conditional value at risk, Journal of Risk, 2, 21-40.

[3] Sklar A. (1959), Functions De Repartition An Dimensions At Leurs Marges, Publications De L'institut De Statistique De L'université De Paris, 8,229-231

[4] He H. and Li P. (2011), Dynamic Asset Allocation Based on Copula and CVaR, IEEE.

[5] Cherubini U., Luciano E. and Vecchiato W. (2004), Copula Methods in Finance. John Wiley and Sons, New York.

#### ABSTRACT

#### PORTFOLIO OPTIMIZATION BASED ON CVAR: DYNAMIC COPULA MODEL

This paper is concerned with the statistical modeling of the dependence structure of multivariate financial data using copula. Since financial data is greatly affected by the economic factors, it often varies according to the time. Therefore, dynamic copula model is used that takes into account the time-varying. In addition, portfolio optimization based on Mean-CVaR model is applied with Monte Carlo simulation. As an application, a portfolio with four different Indexes is constructed from the Turkish financial markets. The marginal distributions of assets in the portfolio are estimated and parameter estimates are given for the different copula models. The portfolio optimization based on CVaR is made for the portfolio created from the specified copula model.

**Key Words:** Dynamic copula, CVaR, portfolio optimization

SAKLI MARKOV MODELİ KULLANILARAK TÜRKİYE DIŞ BORÇLARININ RİSK  
DEĞERLENDİRİLMESİ

\*Ceren Eda CAN Gül ERGÜN

Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Beytepe, Ankara, TÜRKİYE

E-mail: [cerencan@hacettepe.edu.tr](mailto:cerencan@hacettepe.edu.tr)

E-mail: [gul@hacettepe.edu.tr](mailto:gul@hacettepe.edu.tr)

Ahmet YALNIZ

Cankaya Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Bankacılık ve Finans Bölümü, Ankara, TÜRKİYE

E-mail: [yalniz@cankaya.edu.tr](mailto:yalniz@cankaya.edu.tr)

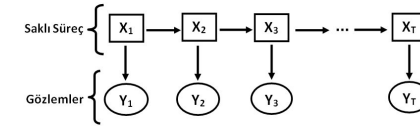
Giriş

Dış borçlanma, gelişmekte olan ülkeler için önemli bir finans kaynağıdır ve ülkeye yabancı kaynak girişi sağladığı için ülke ekonomisine başlangıçta olumlu etki yaratmaktadır. Buna karşılık, faiz ve anapara ödemeleri esnasında iç kaynakların ülke dışına transferini gerektirdiği için ülke ekonomisi üzerinde olumsuz etkiler oluşturabilmektedir. Dış borçların sürdürülebilirliği, ulusal gelirin mevcut ve gelecekteki borç yükümlülüklerini ve diğer giderlerini karşılama durumuna bağlıdır. Ülkedeki döviz akışı ve likidite durumunun dış borç servisi için yeterli düzeyde olmaması, var olan borçların yeniden finanse edilmesini gerektirmektedir. Bu durum, ülkede ciddi bir borç krizine neden olabilir. Yabancı kaynak girişinin, ülkede ekonomik dengeyi sağlaması ve koruyucu bir etkide bulunması, etkili bir dış borç yönetimini gerektirmektedir. Türkiye’de yaşanan ekonomik krizlerin çoğunu dış borçlar tetiklemektedir. Olası borç krizlerinin önüne geçilmesi amacıyla, Türkiye’nin dış borç stoklarının modellenmesi büyük önem taşımaktadır. Ülkenin borçlanma limiti belirlenerek gerekli ekonomik politikalar düzenlenmelidir. Finansal zaman serilerinin dinamik bir yapıya sahip olmasından dolayı belli dönemlerde irrasyonel değişimler meydana gelebilir. Bu irrasyonel değişimler, ülkenin o dönemdeki ekonomik durumunu temsil eden bir rejime ait olabilir. SMM sayesinde, seride uzun ve kısa vadede bir yöne doğru gerçekleşen rejim değişiklikleri tanımlanarak, bu rejimler arasında ilişki yapısı göz ardı edilmeden serideki hareketliliğin modellenmesi sağlanmaktadır. Bu çalışmada, Saklı Markov Modeli (SMM) ile Türkiye toplam dış borç stokundaki (milyon ABD \$) değişimlerin modellenmesi ve böylece Türkiye’nin dış borçluluk yapısının belirlenmesi amaçlanmıştır. Modelde, farklı rejimlerin etkilerinin yansıtılmasında Normal dağılım kullanılmıştır. Böylece, Türkiye’deki ekonomik istikrarı sağlayacak gerekli politikaların alınması amacıyla, Normal-SMM ile Türkiye toplam dış borç stokunda gelecek dönemlerde gerçekleşmesi beklenen değişim oranları tahmin edilmektedir.

Saklı Markov Modeli

Markov modeli, ele alınan sistemin durumlarının ve bu durumlara ilişkin olasılıkların bilindiği stokastik bir modeldir ve sistemin herhangi bir t zamanındaki çıktısı gözlem olarak adlandırılmaktadır. Ancak, gerçek uygulamalarda ilgilenilen sistemin gözlemlenebilmesi çoğu zaman mümkün değildir. Bu durum, Markov modelinin uygulama alanını daraltan önemli bir sorundur. SMM’de, sistemin herhangi bir anda içinde bulunduğu durumun bilinmediği varsayılmaktadır.

Kesikli dizin kümesi  $T = \{1, 2, \dots, T\}$  olmak üzere, SMM kesikli zamanlı ikili bir stokastik süreç  $(X, Y) = (X_t, Y_t)_{t \in T}$  olarak tanımlanır.  $X = (X_t)_{t \in T}$  süreci, kesikli zamanlı, homojen, sonlu-durumlu Markov zinciridir ve saklı süreç olarak tanımlanır.  $X_t, t$  anında sistemin içinde bulunduğu durumdur.  $Y = (Y_t)_{t \in T}$  süreci,  $X$  saklı sürecinin etkisi altında ortaya çıkan gözlemlenebilir süreç olarak tanımlanmaktadır.  $Y_t, t$  anında meydana gelen gözlemdir. Gözlemlenebilir bu süreç aracılığıyla, sistemin saklı durumuna ilişkin problemler incelenmekte ve tüm istatistiksel çıkarımlar gözlemlenebilir bu süreç üzerinden yapılmaktadır. SMM’de,  $X$  ve  $Y$  süreçlerinin arasındaki ilişki bir dinamik Bayes ağı oluşturmaktadır. Bu yapı, Şekil 1’de gösterilmektedir



Şekil 1. SMM'nin Yapısı

Bu çalışmada, belirtilen arıksık zaman aralıklarında Türkiye toplam dış borç stokunda meydana gelen değişim oranının Normal dağılım ailesinden türetildiği ve bu ailenin parametrelerinin homojen bir Markov zinciri olduğu Normal-SMM (NSMM) kullanılmıştır.

KAYNAKLAR

- [1] Cappe, O., Moulines, E. and Ryden, T. (2005), *Inference in Hidden Markov Models*, Springer Science+Business Media, New York, USA.
- [2] Rabiner, L.R. (1989), *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, Proceedings of the IEEE, 77:2, 257-285.
- [3] Sarı, M. (2004), *Dış Borç Yönetimi ve Türkiye Uygulamaları*, Türkiye Cumhuriyeti Merkez Bankası Dış İlişkiler Genel Müdürlüğü, Uzmanlık Yeterlilik Tezi, Ankara, Türkiye.
- [4] Zucchini, W. and MacDonald, I.L. (2009), *Hidden Markov Models for Time Series: An Introduction Using R*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Boca Raton, USA.

RISK ASSESSMENT OF TURKISH EXTERNAL DEBT BY USING HIDDEN MARKOV MODEL

*External borrowing is an important source of finance for developing countries. In the first stage, due to the inflow of foreign resources, external debt provides positive impact on the economy of the country. However, at the time of the interest and principal payments, due to outflow of internal resources, the external debt gives rise to negative impact on the economy. If the liquidity and cash flow in the country is not sufficient to service external debt, it may be required to refinance the debt, which can result in serious further debt crisis. In particular, economic crisis in Turkey is mostly triggered by external debt. Therefore, the external debt is one of the most important problems in Turkey. The aim of this study is to model the change of total Turkish external debt stock by Hidden Markov Model (HMM). Here, Normal-HMM is used to forecast the future change level of Turkish external debt and determine its structure in order to adopt necessary policies to ensure economic stability.*

**Key Words:** Hidden Markov Model, Turkish External Debt, External Debt Management

**TÜRKİYE'DEKİ SİGORTA ŞİRKETLERİNİN SERMAYE YAPISI SEÇİMİNİN MIMIC MODEL  
KULLANILARAK ANALİZİ**

Zeynep İLHAN, Veysel YILMAZ

Eskişehir Osmangazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, 26480, Eskişehir, TÜRKİYE

zeynepilhan@ogu.edu.tr

vyilmaz@ogu.edu.tr

**1. Mimic Model ve Sermaye Yapısı Seçimi**

Sermaye yapısı seçimi konusunda literatür oldukça geniş olmakla birlikte alanda önemli çalışmalardan biri olan "The cost of capital, corporation finance, and the theory of investment" Modigliani ve Miller tarafından 1958'de yayınlanmıştır. Titman ve Wessels ise 1988'de şirketlerin sermaye yapısı seçiminde belirleyicilerin modellemesini regresyon analizi kullanarak yapmıştır. Günümüzde ise Fama ve French sermaye yapısı ile ilgili önemli çalışmalara imza atmaktadır. 2006-2011 yılları arasında çalışmasını yürüttükleri son yayınları 2012 yılında "Capital Structure Choices" yayınlanmıştır.

Literatürde, sermaye yapısı incelemelerinde regresyon analizi sıklıkla tercih edilmiş olsa da Chang ve diğerleri, 2008 yılında sermaye yapısını MIMIC model ile açıklayan bir çalışma yayınlamışlardır. Çoğunlukla kayıtdışı ekonomi ve yolsuzluk çalışmalarında kullanılan Mimic Model'in bu kullanımı literatüre farklı bir bakış açısı kazanmıştır.

Mimic Model'in alternatiflerine göre üstünlükleri mevcuttur. Gizil değişkene etki eden birden fazla neden değişkenini kullanmaya ve bunların nisbi anlamlılıklarını belirlemeye olanak sağlarken, aynı anda gizil değişkenin etkinliğinin birkaç farklı gösterge değişkenini de hesaba katabilir. Modelin teorik yapısı da bu durumu net bir şekilde göstermektedir.

MIMIC model eşitlikleri,

$$Y = \lambda\eta + \varepsilon \quad (1)$$

$$\eta = \gamma X + \zeta \quad (2)$$

şeklindedir.  $y' = (y_1, y_2, \dots, y_n)$  değişkenleri, " $\eta$ " gizil değişkeninin göstergeleri (çoğunlukla kayıtdışı ekonomi ya da rüşvet olarak alınır);  $x' = (x_1, x_2, \dots, x_q)$  değişkenleri ise, " $\eta$ " gizil değişkeninin neden değişkenleridir. LISREL bakış açısıyla, Eşitlik (1)  $\eta$  için ölçüm modeli olarak, Eşitlik (2) ise  $\eta$  için yapısal model olarak kabul edilebilir (Jöreskog, Sörbom, 1993).

Yapısal modelden gizil değişkene ilişkin herhangi bir tahmin yapmak söz konusu olamamaktadır. Bu nedenle ölçüm modeli ve yapısal model aşağıdaki biçimde indirgenir belirlir (Baldemir, Özkoç ve İşçi, 2009).

$$Y = \lambda y'X + \lambda \zeta + \varepsilon, \\ = \Pi X + z \quad (3)$$

İndirgenmiş modelde, Eşitlik (3)'de, katsayı matrisi  $\Pi = \lambda y'$  ve uzaklık vektörü  $z = \lambda \zeta + \varepsilon$  biçimindedir (Savaşan, 2003). Ayrıca,  $\text{cov}(z) = \lambda \lambda' \psi + \Theta \varepsilon$  ve  $\psi = \text{var}(\zeta)$ 'dir.  $\Theta \varepsilon$  ise  $\varepsilon$ 'nin diagonal kovaryans matrisidir.

Bu çalışmada, Türkiye'de faaliyet gösteren sigorta şirketlerinin sermaye yapısı incelenirken Mimic modelin alternatif analizlere göre üstünlüklerinden faydalanmak amaçlanmıştır. 2002-2011 yılları arasında Türkiye'de faaliyet gösteren sigorta şirketlerinin Hazine Müsteşarlığı'ndan alınan gelir tablosu ve bilançoları kullanılmıştır.

Titman ve Wessels ile Chang, Lee ve Lee'nin çalışmalarından yola çıkarak, neden değişkenleri olarak büyüme, teklik, borç dışı vergi kalkını, teminat değeri, kârlılık, oynaklık ve endüstri değişkenleri alınmıştır. Gösterge değişkenleri olarak ise uzun vadeli borç, kısa vadeli borç ve dönüştürülebilir borç değişkenleri kullanılmıştır.

**KAYNAKLAR:**

[1] Savaşan, F. (2003), Modelling The Underground Economy in Turkey : Randomized Response and MIMIC Models, Journal of Economics, v.XXIX, No.1, pp.47-76.

[2] Giles, D. (1998), Modelling The Hidden Economy and The Tax-gap n New Zealand, Econometrics Working Paper EWP 9810.

[3] Chang, C., Lee, A. and Lee, C. (2009), Determinants of capital structure choice: A structural equation modeling approach, The Quarterly Review of Economics and Finance 49, 197-213.

[4] Baldemir, E., Özkoç, H. ve İşçi, Ö. (2009), MIMIC Model ve Yolsuzluk Üzerine Türkiye Uygulaması, Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt: 24, Sayı:2, 49-63.

[5] Jöreskog, K., Sörbom, D. (1993), LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language, SSI Scientific Software International, Chicago.

**ABSTRACT**

**AN ANALYSIS OF CAPITAL STRUCTURE CHOICE OF INSURANCE COMPANIES IN TURKEY BY USING MIMIC MODEL**

Modigliani and Miller published an important paper about capital structure choice in 1958 and then Titman and Wessels published a paper about this issue using regression analysis in 1988. Chang and others used Mimic Model and published a paper in 2008. Mimic Model has some advantages in comparison to other analysis. In this paper, we analysed capital structure choice of insurance companies in Turkey between 2002 and 2011. We used growth, uniqueness, non-debt tax shields, assets collateral value, earnings volatility, profitability, industry classification as cause variables and used long-term debt, short-term debt and collateral value as indicator variables.

**Key Words:** Mimic Model, capital structure choice, insurance

SAĞLAM BULANIK C-REGRESYON MODEL VE OTOREGRESİF MODELLERİN ÖNGÖRÜSÜNE  
UYGULANMASI - SİMÜLASYON ÇALIŞMASI

Aytaç PEKMEZCİ, Nevin GÜLER  
Muğla Sıtkı Koçman Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Kötekli, Muğla

aytac0803@mu.edu.tr, nguler@mu.edu.tr

**Otoregressif Modeller**

Otoregressif modeller, bir zaman serisinin geçmişte gözlenen değerlerinden yararlanarak gelecek değerlerini tahmin etmek için kullanılmaktadır.  $y_1, y_2, \dots, y_n$  şeklinde bir zaman serisi düşünüldüğünde p. gecikmeli otoregressif model aşağıdaki gibi tanımlanmaktadır:

$$y_i = \beta_0 + \beta_1 y_{i-1} + \beta_2 y_{i-2} + \dots + \beta_p y_{i-p} + \varepsilon_i \quad (1)$$

Burada  $y_i$ , i. zamanda bağımlı değişkenin değerini,  $\{\beta_0, \beta_1, \dots, \beta_p\}$  tahmin edilecek parametre vektörünü,  $\varepsilon_i$  ise 0 ortalama ve  $\sigma^2$  varyansla normal dağılan hata vektörünü göstermektedir.

Otoregressif modeller zaman serisi tahmininde kullanılan en basit modelleme aracı olmasına rağmen, zaman serisinin doğrusal olmaması, veri sayısının az olması, olasılık varsayımlarının sağlanmaması gibi durumlarda kullanılamamaktadır. Bu gibi durumlarda bulanık mantık, yapay sinir ağları gibi esnek hesaplama teknikleri son zamanlarda yaygın olarak kullanılmaktadır.

**Sağlam Bulanık C-Regresyon Model**

Hathaway and Bezdek (1993) [1], prototipi girdi uzayının ( $x_k$ ) bir fonksiyonuna  $f(x_k)$  karşılık gelen kümeleri teşhis edebilmek amacıyla Bulanık C-Regresyon Modeli (BCRM) geliştirmiştir. Burada  $f(x_k)$ 'i (1) eşitliğinde verildiği gibi tanımlamak mümkündür. Dolayısıyla, BCRM'in ( $Y_{-1}, Y_{-2}, \dots, Y_{-p}$ ) ile  $Y$  arasındaki ilişkiyi tahmin etmek için kullanılabilirliği açıktır. Nitekim şu ana kadar yapılan çalışmalar BCRM'in veri setinin özellikle birbirinden farklı dağılımlara sahip olması ve (veya) bağımsız değişkenler ile bağımlı değişken arasındaki ilişkinin doğrusal olmayan bir yapıda olması durumunda oldukça başarılı sonuçlar verdiğini göstermiştir [2][3]. Ancak BCRM veri setinde yer alan aykırı ve gürültülü noktalardan oldukça etkilenmektedir. Sağlam Bulanık C-Regresyon Model (SBCRM) [4] BCRM'in avantajlarını da koruyarak, hem veri setinde yer alan aykırı değerleri teşhis etmek hem de bu değerlerin modellere etkisini azaltmak için geliştirilmiştir. SBCRM aşağıdaki gibi verilen amaç fonksiyonunun minimize edilmesine dayanır:

$$J(U, \beta_i) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \frac{1}{w_k^q} d_{ik}(x_k, \beta_i) \quad (2)$$

Burada  $n$  veri noktası sayısını,  $\sum_{k=1}^n w_k = w$  veri setindeki noktaların toplam ağırlığını göstermektedir. (2)

eşitliğinde verilen amaç fonksiyonu minimize edildiğinde  $\beta_i$  parametre vektörü,  $U$  üyelik değerleri matrisi ve  $W$  ağırlık matrisi için elde edilen eşitlikler aşağıdaki gibidir:

$$\beta_i = [X_i W_i X_i]^{-1} X_i^T W_i Y \quad (3)$$

**BİLDİRİ OTURUMLARI 1**

*SESSION 1*

*Zaman Serileri 1*

Burada  $W_i$  köşegen elemanları  $u_{ik}/w_{ik}^q$  karşılık gelen diyagonal bir matristir,  $X_1$  ise gecikmeli değerlerden oluşan matrise "1" sütununun eklenmesiyle oluşturulmaktadır.

$$u_{ij} = 1 / \left( \sum_{k=1}^c (d_{ij}/d_{ik})^{m-1} \right) \quad (4)$$

$$w_{ik} = \left( \frac{\sum_{i=1}^c u_{ik}^m d_{ik}(x_k; \theta_i)^{1/(q+1)}}{\sum_{k=1}^n \left( \sum_{i=1}^c u_{ik}^m d_{ik}(x_k; \theta_i)^{1/(q+1)} \right)^q} \right) w \quad (5)$$

SBCRM iteratif bir algoritmadır. Buna göre amaç fonksiyonu minimum değerine yakınsayana kadar (3), (4) ve (5) eşitliklerinin tekrarlı olarak hesaplanması gerekmektedir.

#### Uygulama

Bu çalışmada otoregressif modeller, BCRM ve SBCRM tekniklerinin geleceği tahmin etmedeki performanslarını karşılaştırmak amacıyla MATLAB programı kullanılarak her birinden 10.000 adet olmak üzere AR(1), AR(2) ve AR(3) serileri üretilmektedir. Üretilen serilerin her birine büyük hata model [5] kullanılarak gürültülü ve aykırı noktalar eklenecektir. Karşılaştırma kriteri olarak Ortalama Hata Kareler ve Ortalama Mutlak Yüzde Hata kriterleri kullanılacaktır.

#### KAYNAKLAR

- [1] Hathaway, R.J., Bezdek, J.C., 1993. Switching Regression models and fuzzy clustering, IEEE Transactions on Fuzzy Systems, 1, (3):195-204.
- [2] Güler, N., Bulanık Kümeleme Analizi ve Bulanık Modellemeye Uygulamaları, Y.Lisans Tezi, Muğla Üniversitesi, Fen Bilimleri Enstitüsü, Muğla, 2006.
- [3] Güler, N., "Olabilirlik ve Bulanık Kümelemeye Dayanan Modelleme Teknikleri ve Yazılım Güvenilirliğinin Tahminine Uygulanması", Muğla Sıtkı Koçman Üniversitesi, Fen Bilimleri Enstitüsü, 2012.
- [4] Shen H., Yang, J. Wang, S., 2004. Outlier Detecting in Fuzzy Switching Regression Models, Lecture Notes in Computer Science, Springer-Verlag GmbH, 3192:208-215.
- [5] Frigui, H., Krishnapuram, R., 1999, A Robust Competitive Clustering Algorithm with Application in Computer Vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21: 450-465.

#### ABSTRACT

#### ROBUST FUZZY C-REGRESSION MODEL AND APPLICATION TO FORECASTING OF AUTOREGRESSIVE MODEL – A SIMULATION STUDY

In this study, Fuzzy C-Regression Model, Switching Fuzzy C-Regression Model and Autoregressive models are applied to forecasting of autoregressive models. For this aim, 30.000 number of data set are generated by using MATLAB program. The forecasting performances of these methods are compared according to Mean Absolute Percentage Error (MAPE) and Mean Square Error (MSE).

**Key Words:** Fuzzy C-Regression Model, Robust Fuzzy C-Regression Model, Autoregressive models

#### YAPAY SINİR AĞLARINA DAYALI YÜKSEK DERECEDEDEN BULANIK ARMA ZAMAN SERİSİ ÖNGÖRÜ YÖNTEMİNİN BİR UYGULAMASI

Cem KOÇAK\*, Erol EĞRİOĞLU

\*Hitit Üniversitesi, Sağlık Yüksekokulu, 19000, Çorum, TÜRKİYE, cemkocak@hotmail.com  
Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 5540, Samsun, TÜRKİYE, erole@omu.tr

#### Giriş

Uzunca bir süre literatürde önerilen bulanık zaman serisi öngörü yöntemlerinin büyük çoğunluğu bulanık gecikmeli (Otoregressif(AR)) değişkenlere dayandırılmakta, hatanın gecikmeli (hareketli ortalama (MA)) değişkenleri kullanılmamaktadır. Bu nedenle son yıllarda AR değişkenleri ile birlikte MA değişkenlerini de kullanan ARMA tipi çalışmalara rastlamak mümkündür. Bu çalışmada da, Kocak [1]'in yapay sinir ağlarına dayalı yüksek dereceden bulanık ARMA (p,q) modelinin literatürdeki bulanık AR modellerine göre üstün yönlerini göstermeyi amaçlayan bir uygulama yapılmıştır. Kocak [1]'in bulanık ARMA(p,q) modelinin avantajları aşağıdaki gibi sıralanabilir.

- Çözümlemede, AR değişkenleri ile beraber MA değişkenlerinin de kullanılması ile bulanık AR yöntemlerinde yapılan model belirleme hatası ortadan kaldırılmaktadır.
- Hem AR değişkeni hem de MA değişkeni yüksek dereceden olduğundan, literatürde geliştirilen birinci ve yüksek dereceden bulanık AR modellerine göre daha çok bilgi ile çözümleme yapılması söz konusudur. Bu durum, öngörü performansını artırmayı sağlamaktadır.
- Bulanık ilişki belirlemede yapay sinir ağlarının kullanımı Chen [2,3] yöntemlerindeki bulanık mantık grup ilişkisi tablolarının oluşturulmasındaki karmaşıklığı ortadan kaldırmış ve öngörü performansını artırmada etkili olmuştur.

Bu çalışmada, Kocak [1]'in yöntemini, literatürdeki diğer bulanık zaman serisi yöntemleri ile karşılaştırmak amacıyla Türkiye Cumhuriyeti Merkez Bankasının resmi Web sitesinden alınan 2009 yılına ait altın fiyatları zaman serisi kullanılmıştır. Altın fiyatlarının son 15 gözlemi ve son 30 gözlemi test kümesi olarak alınarak 2 farklı uygulama yapılmıştır. Bu uygulamalardan test kümesi sayısı 30 için çalışmada kullanılan yöntemlerden en iyi öngörülere sahip 5 adet yöntem için elde edilen Performans değerleri Çizelge 1'de verilmiştir.

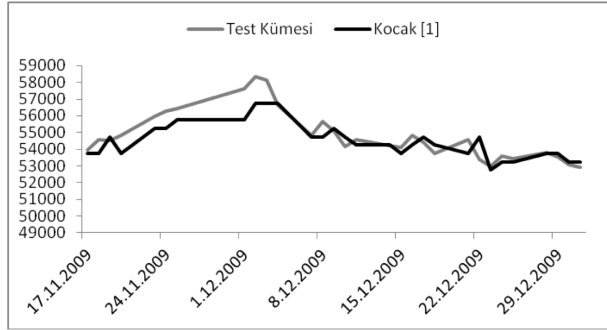
#### Çizelge 1. Yöntemlerin en iyi sonuçları için altın fiyatlarının 30 gözlemlilik test kümesi için elde edilen performansları

Performans	Chen [2]	Chen [3]	Huang [4]	Aladağ vd. [5]	Kocak [1]*
Hata Kareler Ortalamasının Karekökü (HKOK)	1031,12	857,34	1045,23	1003,50	752,48
Ortalama Mutlak Yüzdellik Hata (OMYH)	0,01512	0,01245	0,01530	0,01382	0,01050
Yön Doğruluğu (YD)	0,55172	0,51724	0,55172	0,48276	0,55172

\*En İyi Sonuç

Çizelge 1. incelendiğinde, Kocak [1] yönteminin 752.48 minimum HKOK, % 1.05 minimum OMYH ve % 55.17 maksimum yön doğruluğu değerleri ile diğer yöntemlerin öngörü performanslarından oldukça uygun değerler

sahip olduğu saptanmıştır. Şekil 1' de ise 30 gözlemlilik test kümesi ile Kocak [1] yöntemin öngörülerinin birbirine oldukça yakın sonuçlar olduğu görülmektedir.



Şekil 1. 2009 yılı altın fiyatlarının 30 gözlemlilik test kümesi ve Kocak [1]'in öngörülerinin grafiği fuzzy optimization and decision making

#### KAYNAKLAR

- [1] Kocak C. (2013), A new high order fuzzy ARMA time series forecasting method by using neural networks to define fuzzy relations, Fuzzy optimization and Decision (**Submitted Article**).
- [2] Chen S.M. (1996), Forecasting enrollments based on fuzzy time-series, Fuzzy Sets and Systems, vol.81, pp.311-319.
- [3] Chen S.M. (2002), Forecasting enrollments based on high order fuzzy time series, Cybernetics and Systems, vol.33, pp.1-16.
- [4] Huarng K. (2001), Effective length of intervals to improve forecasting in fuzzy time series, Fuzzy Sets and Systems, vol.123, pp.387-394.
- [5] Aladag C.H., Basaran M.A., Egrioglu E., Yolcu U, and Uslu V.R. (2009), Forecasting in high order fuzzy time series by using neural networks to define fuzzy relations, Expert Systems with Applications, vol.36, pp.4228-4231.

#### ABSTRACT

#### AN APPLICATION OF THE HIGH ORDER FUZZY ARMA TIME SERIES FORECASTING METHOD BASED ON NEURAL NETWORKS

Forecasting performance in determining fuzzy relations has been substantially increased with the use of artificial neural networks and artificial intelligence algorithms. However, the fuzzy time series forecasting methods in literature are frequently based on fuzzy lagged (Autoregressive (AR)) variables. Inclusion of only AR variables in the fuzzy time series forecasting models proposed in literature can cause model specification error. For this reason, Kocak [1] has developed high order fuzzy ARMA(p,q) method by using neural networks to define fuzzy relations for eliminating model specification error. In this study, Kocak [1]'s solution algorithm has been aimed showing the advantages by comparison with fuzzy AR models in Literature.

**Key Words:** Fuzzy Time Series, Autoregressive Moving Average, ARMA Type, High Order

#### BULANIK KÜMELEMeye DAYALI YEREL REGRESYON VE OTOREGRESİF MODELLERİN ÖNGÖRÜSÜNE UYGULANMASI

<sup>1</sup>Nevin GÜLER, <sup>2</sup>Özlem TÜRKŞEN

<sup>1</sup>Muğla Sıtkı Koçman Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Kötekli, MUĞLA  
<sup>2</sup>Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Tandoğan, ANKARA

[nguler@mu.edu.tr](mailto:nguler@mu.edu.tr), [turksen@ankara.edu.tr](mailto:turksen@ankara.edu.tr)

#### Yerel Regresyon

Zamana bağlı olarak elde edilen verilerin analizinde, hata değişkenleri arasında korelasyon (otokorelasyon) söz konusu olmaktadır. Bu durumda, zamanla ilgili değerlendirmeler yapmak için parametrik regresyon analizinin gerekli varsayımları sağlanamayacağından, bilinmeyen yanıt fonksiyonunun tahmininde parametrik olmayan regresyon yöntemlerinden yararlanır. Yerel regresyon (YR), verilere uygun model belirlenmesi amacı ile geliştirilmiş bir parametrik olmayan regresyon yöntemidir.  $\mathbf{X} = [X_1 X_2 \dots X_n]$  bağımsız değişken vektörü ve  $Y$  bağımlı değişken ölçümleri arasındaki ilişki

$$Y_i = f(X_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

ile tanımlansın. Burada,  $f(X_i)$  bilinmeyen bir düzgün fonksiyon,  $\varepsilon_i$ , 0 ortalamalı ve  $\sigma^2$  sabit varyanslı hata terimleridir. Taylor teoremi gereği, herhangi bir sürekli fonksiyona polinomsal yaklaşımda bulunulabileceği göz önünde bulundurularak, YR analizine göre, bilinmeyen  $f$  fonksiyonuna basit parametrik fonksiyon ile yerel olarak yaklaşımda bulunulabileceği varsayılır.  $f$  fonksiyonuna,  $x$  sorgu noktasında yaklaşımda bulunulacak  $p$ . dereceden genel bir polinom yaklaşım olarak

$$f(X_i) \approx \sum_{k=0}^p \beta_k \frac{(X_i - x)^k}{k!} \quad (2)$$

biçiminde tanımlanır. (2) ile tanımlı fonksiyonun  $x$  sorgu noktasındaki YR değerinin tahmin edilmesi için  $\beta_i$ ,  $i = 0, 1, \dots, k$  ile gösterilen katsayılarının bulunması gerekir. Bu parametrelerin belirlenmesi

$$\hat{\beta} = \arg \min_{\beta \in R^p} \sum_{i=1}^n w_i(x) \left[ Y_i - \left( \beta_0 + \beta_1 (X_i - x) + \frac{\beta_2}{2} (X_i - x)^2 + \dots + \frac{\beta_k}{k!} (X_i - x)^k \right) \right]^2 \quad (3)$$

ile tanımlı problemin çözümü ile mümkündür. Burada  $w_i(x)$ ,  $(X_i - x)$  uzunluğuna bağlı olarak atanan  $i$ . gözlemin ağırlığıdır. (3) eşitliği ile tanımlı problemin optimizasyonu sonucu

$$\hat{\beta} = \left[ \hat{\beta}_0 \hat{\beta}_1 \dots \hat{\beta}_k \right]' = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\mathbf{Y} \quad (4)$$



parametre tahminleri elde edilir. Burada,  $\mathbf{X}$  tasarım matrisi,  $\mathbf{W}$  köşegen elemanları  $w_i(x)$ ,  $i = 1, 2, \dots, n$  olan köşegen ağırlık matrisidir. Böylece,  $\hat{\beta}$  tahmin değerlerinin belirlenmesi ile YR tahmini model denklemi  $\hat{f}(x)$  elde edilmiş olur.

#### Bulanık Kümelemeye Dayalı Yerel Regresyon

Bulanık kümeleme, veri noktalarının farklı üyelik dereceleri ile eş zamanlı olarak birden fazla kümeye dahil olmasına izin veren kümeleme tekniklerini içerir. En basit ve yaygın olarak kullanılan bulanık kümeleme tekniği Bulanık C-Ortalamalar algoritmasıdır. Bulanık kümeleme algoritmalarının bir çoğu amaç fonksiyonunun minimize edilmesine dayanır. BCO algoritmasına ilişkin amaç fonksiyonu aşağıdaki şekilde tanımlanmaktadır:

$$J(U, V; X) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (x_k - v_i)^2 \quad (5)$$

Burada  $u_{ik}$ ,  $k$ . veri noktasının  $i$ . kümeye bulanık üyeliğini,  $m$  bulanıklık indeksini,  $v_i$  ise  $i$ . kümenin küme merkezini göstermektedir. Bulanık kümelemeye dayalı yerel regresyonun (YR\_BK) adımları şu şekilde verilebilir:

**Adım 1:** BCO algoritması kullanarak  $v_i$  küme merkezleri hesaplanır.

**Adım 2:** Her bir küme merkezi YR için sorgu noktası ( $x$ ) kabul edilerek (4) eşitliğine göre küme sayısı kadar YR modeli elde edilir.

**Adım 3:** Tüm veri seti için küme sayısı ( $c$ ) kadar tahmin değeri ( $\hat{y}_{ki}$ ,  $k = 1, 2, \dots, n$ ,  $i = 1, 2, \dots, c$ ) hesaplanır.

**Adım 4:** Aşağıdaki eşitlik kullanılarak birleştirilmiş  $y$  tahmin değeri ( $\hat{y}_k$ ) hesaplanır.

$$\hat{y}_k = \frac{\sum_{i=1}^c u_{ki} \hat{y}_{ki}}{\sum_{i=1}^c u_{ki}} \quad (6)$$

Burada,  $u_{ki}$  BCO algoritmasından elde edilmektedir.

#### KAYNAKLAR

[1] Loader, C. R. (1999), Local Regression and Likelihood, New York: Springer.

[2] Bezdek, J.C. (1981), Pattern Recognition With Fuzzy Objective Function, Plenum Press, New York.

#### LOCAL REGRESSION BASED ON FUZZY CLUSTERING AND APPLICATION TO FORECASTING OF AUTOREGRESSIVE MODELS

In this study, local regression based on fuzzy clustering (LR\_FC) method is suggested for forecasting of autoregressive models. The proposed method is applied on a data set which is related to population migration rate, foreign population migration, and consumer price indexes of OECD countries and the obtained results are compared with classical local regression and AR(1) models. It is seen from the analysis results that LR\_FC gives at least as good as results of classical methods with a smaller number of models.

**Key Words:** Local regression, local regression based on fuzzy clustering, autoregressive models.

#### PİYASA RİSKLERİNİN BELİRLENMESİNDE RİSKE MARUZ DEĞER YÖNTEMİ ve BORSA İSTANBUL'DA BİR UYGULAMA

Yusuf DEMİR\*  
Fatma Esin KURT\*\*

\*Süleyman Demirel Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Turizm İşletmeciliği Bölümü Doğu Kampüsü  
32260 Çünür / ISPARTA, [yusufdemir@sdu.edu.tr](mailto:yusufdemir@sdu.edu.tr) Tel: 246 2113050; Faks: +90 0246 - 237 09 20

\*\*Mehmet Akif Ersoy Üniversitesi, Fen –Edebiyat Fakültesi İstatistik Bölümü, İstiklal Yerleşkesi 15030 BURDUR,  
[esinkurt@mehmetakif.edu.tr](mailto:esinkurt@mehmetakif.edu.tr) Tel : +90 248 213308; Faks: +90 248 213 30 99

#### Riske Maruz Değer (RMD) Kavramı ve Hesaplama Yöntemleri

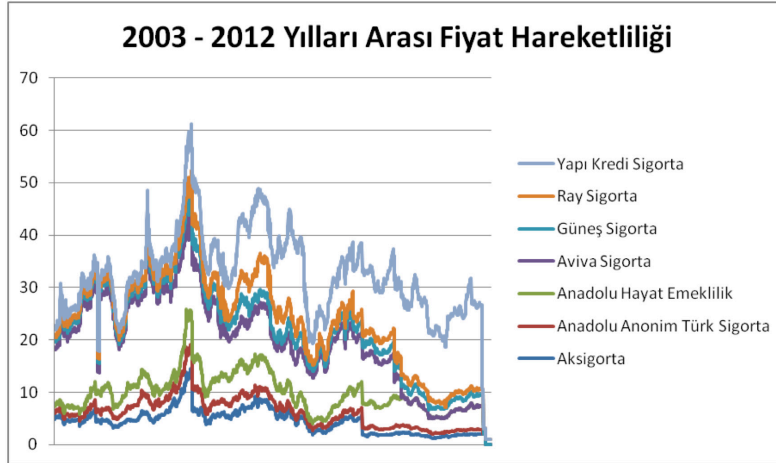
Riske Maruz Değer (RMD) yöntemi son dönemde risk yönetimi alanında yaygın olarak kullanılmaktadır. 1994'te JP Morgan tarafından geliştirilen bir risk yönetim tekniği olan bu yöntem Risk Metrics kullanımı ile gelişmiştir. İşletmeler ulusal ve uluslar arası finansal piyasalara açıldıkça riskler artmakta ve etkin bir risk yönetim sisteminin gerekli olmaktadır.

RMD riskli bir varlık yada portföyün, belirlenen bir güven aralığında potansiyel kayıplarının hesaplanması esasına dayanan bir yöntemdir. RMD hesaplamalarında dört genel yöntem kullanılmaktadır. Bu yöntemler, Varyans-Kovaryans Yöntemi, Tarihi Yöntem, Monte Carlo Simülasyonu Yöntemi ve Delta-Gamma Yöntemleridir.

Bu çalışmanın amacı uygun volatilité tahmin yönteminin belirlenmesinden sonra piyasa risklerinin en uygun Riske Maruz Değer (RMD) tahminci yöntemiyle hesaplanmasıdır. Bu hesaplama süreci, risk faktörlerinin volatilitesi, güven düzeyi ve elde tutma süresi gibi değişkenlerin belirlenmesini içermektedir.

Finansal kurumlar portföylerini çeşitli enstrümanlardan oluşturmaktadırlar. Bu çalışmada oluşturulan portföyde hisse senetleri üzerine yoğunlaşmıştır. Hisse senetleri riskli enstrümanlar grubunda yer almakla birlikte hesaplamalardaki karmaşıklık da fazla olmaktadır.

RMD analizi yardımı ile Borsa İstanbul'da işlem gören sigorta şirketlerinin hisse senetlerinden oluşturulmuş portföyün (1 Ocak – 31 Aralık) on yıllık süreçte riskleri ölçülmüştür. Araştırmanın sonuçlarına göre; portföyün risk yapısı dönemlere göre farklılık göstermektedir.



Şekil 1. Borsa İstanbul'a Kote Hayat, Hayat-Dışı ve Bireysel Emeklilik Şirketlerinin 2003 – 2012 Yılları Arası Hisse Senetleri Günlük Fiyat Değişim Grafiği

#### KAYNAKLAR

- [1] Demireli E. ve Taner B. (2009), Risk Yönetiminde Riske Maruz Değer Yöntemleri ve Bir Uygulama, Süleyman Demirel Üniversitesi İİBF Dergisi
- [2] Ibragimov R. and Walden J. (2011), Value at risk and efficiency under dependence and heavy-tailedness: models with common shocks, Ann Finance, Springer-Verlag.
- [3] Jeffrey T. Tsai, Jennifer L. Wang, Larry Y. Tzeng. (2010), On the optimal product mix in life insurance companies using conditional value at risk, Mathematics and Economics, Elsevier B.V.
- [4] Tapiero, Charles S. (2010), Risk finance and asset pricing value, measurements and markets, New York: Wiley
- [5] Tunç G. (2010), Volatility Modelling and Forecasting Value-at-Risk: Evidence from New and Candidate European Union Countries, Doktora Tezi, İzmir University of Economics

**Anahtar Kelimeler:** Riske Maruz Değer, Sigortacılık, Piyasa Risk Ölçümü

#### GAUSS AKTİVASYON FONKSİYONUNA DAYALI ÇARPIMSAL NÖRON MODEL YAPAY SİNİR AĞI

- Özge GÜNDOĞDU<sup>1\*</sup>, Erol EĞRİOĞLU<sup>2</sup>, Çağdaş Hakan ALADAĞ<sup>3</sup>, Ufuk YOLCU<sup>4</sup>  
<sup>1\*</sup> Cumhuriyet Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, 58140, Sivas, TÜRKİYE,  
E-mail: ozge5gundogdu@hotmail.com  
<sup>2</sup> Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 55139, Samsun, TÜRKİYE,  
E-mail: erole@omu.edu.tr  
<sup>3</sup> Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800, Ankara, TÜRKİYE,  
E-mail: aladag@hacettepe.edu.tr  
<sup>4</sup> Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06100, Ankara, TÜRKİYE,  
E-mail: varyansx@hotmail.com

#### Giriş

Zaman serisi öngörü problemi için yapay sinir ağları yaygın olarak kullanılmaktadır. Zaman serilerinin çözülmesinde farklı yapay sinir ağı (YSA) türleri mevcuttur. Yadav vd. [1]'de önerilen ve başarılı öngörü sonuçları üreten bir YSA türü, çarpımsal nöron modele dayalı yapay sinir ağıdır (ÇNM-YSA). Bu tür yapay sinir ağları ara katman içermediğinden, ara katman eleman sayısı belirleme gibi problemlere de sahip değildir. Aladağ vd. [2], Yolcu vd. [3] çalışmalarında çarpımsal nöron modele dayalı farklı iki YSA ortaya koyulmuştur. Tüm bu ÇNM-YSA' larda aktivasyon fonksiyonu sigmoid olarak alınmıştır. Radyal tabanlı aktivasyon fonksiyonu kullanan yapay sinir ağlarının daha başarılı öngörü sonuçları ürettiği Aslanargun vd. [4] ve literatürdeki diğer bazı çalışmalarda gösterilmiştir. Radyal tabanlı aktivasyon fonksiyonları doğrusal olmayan problemlerin çözümünde başarılı sonuçlar üretmesine rağmen, ÇNM-YSA' larda henüz kullanılmamıştır. Bu çalışmada ÇNM-YSA' da sigmoid aktivasyon fonksiyonu yerine radyal tabanlı olan Gaussian aktivasyon fonksiyonu kullanılmıştır. YSA' nın ağırlıkları ve aktivasyon fonksiyonunun parametreleri garanti yakınsamalı parçacık sürü optimizasyonu (GY-PSO) ile optimize edilmektedir. Bu çalışmanın en önemli iki katkısı; çarpımsal nöron modelde Gaussian aktivasyon fonksiyonunun ilk kez kullanılması ve aktivasyon fonksiyonunun merkez ve yayılım parametrelerinin YSA' nın ağırlıkları ile birlikte GY-PSO ile tek bir optimizasyon süreci içerisinde tahmin edilmesidir. Önerilen Gaussian aktivasyon fonksiyonuna dayalı çarpımsal nöron model yapay sinir ağının (RT-ÇNM-YSA) üstün öngörü performansı bir gerçek hayat zaman serisine uygulanarak kanıtlanmıştır.

#### 2. Uygulama

RT-ÇNM-YSA' nın performansının değerlendirilmesi amacıyla literatürde sıklıkla kullanılan bir gerçek hayat zaman serisi kullanılmıştır. Zaman serisi 1956-1994 yılları arasında çeyreklik gözlemlenen Avusturya bira tüketimi zaman serisidir ve mega litre cinsinden değerler olarak verilmiştir (Janacek [5], Sayfa 84). Toplam 148 gözlem içeren zaman serisinin son 16 gözlemi test verisi olarak, diğer gözlemler ise eğitim verisi olarak kullanılmıştır. Önerilen yöntemin performansının karşılaştırılması amacıyla, önerilen yöntemin yanı sıra literatürde sıklıkla kullanılan bazı yöntemler (SARIMA-Mevsimsel Otoregresif Bütünleşik Hareketli Ortalama Modeli, WMES- Winters'ın Çarpımsal Üstel Düzleştirme Yöntemi, İB-YSA- İleri beslemeli YSA, RT-YSA –Radyal Tabanlı YSA, ÇM-YSA- Çarpımsal Mevsimsel YSA ve D&DO-YSA- Doğrusal ve Doğrusal Olmayan YSA) ile zaman serisi çözümleme sonuçları kullanılmıştır. Zaman serisinin önerilen yöntem dışındaki yöntemler için elde edilen hata ölçütleri (Hata Kareler Ortalaması Karekökü-HKOK, Ortalama Mutlak Yüzdellik Hata-OMYH, Mutlak Yüzdellik Hata Medyanı -MdMYH) Yolcu vd. (2013)'den alınmış ve Çizelge 1'de verilmiştir. RT-ÇNM-YSA' nın uygulanmasında uygun girdi sayısı 1-16 arasında değiştirilerek deneme yanılma yöntemi ile belirlenmiştir. RT-ÇNM-YSA' nın eğitiminde kullanılan GY-PSO' nun parametreleri  $vm_{aps} = 1$ ,  $c_1 = 2$ ,  $c_2 = 2$ ,  $f_c = 5$ ,  $s_c = 5$ ,  $\epsilon = 10^{-6}$ , parçacık sayısı 30 ve maksimum yineleme sayısı 1000 olarak alınmıştır. RT-ÇNM-YSA' nın en iyi sonucu, girdi olarak ilk 12 gecikmeli değişkenin kullanıldığı mimariden elde edilmiştir.

#### Çizelge 1. Tüm Yöntemlerden Elde edilen Öngörüler ve Performans Ölçüleri

Test Verisi	SARIMA	WMES	İB-YSA	RT-YSA	ÇM-YSA	D&DO-YSA	RT-ÇNM-YSA
HKOK	47.0367	53.3295	24.1052	41.7000	22.1700	18.7888	15.8378
OMYH	0.0949	0.1072	0.0476	0.0686	0.0394	0.0357	0.0300
MdMYH	0.0980	0.1032	0.0459	0.0481	0.0328	0.0390	0.0235

#### KAYNAKLAR

- [1] Yadav R.N., Kalra P.K. and John J. (2007). Time series prediction with single multiplicative neuron model, Applied Soft Computing, 7, 1157-1163.
- [2] Aladag C.H., Yolcu U. and Egrioglu E. (2013). A new multiplicative seasonal neural network model based on particle swarm optimization, Neural Processing Letters, 37(3), 251-262.
- [3] Yolcu U., Aladag C.H. and Egrioglu E. (2013). A New Linear & Nonlinear Artificial Neural Network Model for Time Series Forecasting, Decision Support System Journals, 54, 1340-1347.
- [4] Aslanargun A., Mammadov M., Yazici B. and Yolacan S. (2007). Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting, Journal of Statistical Computation and Simulation, 77:1, 29-53
- [5] Janacek G. (2001). Practical time series, Oxford Univ. Press Inc., New York, pp., 156.

#### ABSTRACT

##### MULTIPLICATIVE NEURON MODEL ARTIFICIAL NEURAL NETWORK BASED ON GAUSSIAN AKTIVATION FUNCTION

Artificial neural networks (ANN) are commonly used to forecast time series. There are many kinds of ANN for forecasting purpose. Multiplicative neuron model artificial neural network (MNM-ANN) has been used for forecasting aim recent years. Because this neural network has only one layer and one neuron, users do not have to determine number of neurons in the hidden layer. Sigmoid activation function has been used in the MNM-ANN, so far. In the literature, it is shown that radial basis activation functions produce more accurate forecasts in the ANN. In this study, new MNM-ANN which has Gaussian activation function is proposed. Weights of new MNM-ANN and parameters of activation function are optimized by using guaranteed convergence particle swarm optimization. The performance of the proposed method is analyzed by applying a real world time series.

**Key Words:** Artificial Neural Network, Multiplicative Neuron Model, Gaussian Activation Function, Forecasting, Particle Swarm Optimization.

#### KOŞULLU VARYANS MODELLERİNİN YILLIK DOĞAL GAZ FİYATLARINA UYGULANMASI

Hakkı POLAT\*, Fatih ÇEMREK\*, Ayşe İŞİ\*\*

\*Eskişehir Osmangazi Üniversitesi Meşelik Kampüsü İstatistik Bölümü /Eskişehir

\*\*Gazi Üniversitesi Gazi Meslek Yüksekokulu Çubuk Yerleşkesi Çubuk/ANKARA

#### 1. Giriş

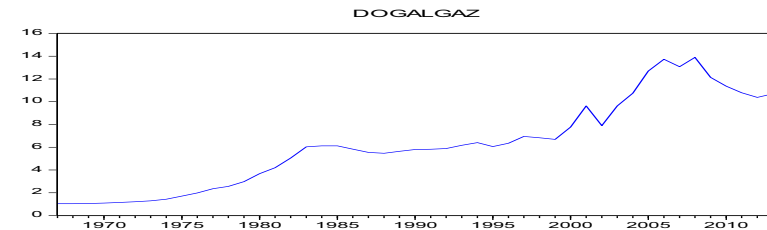
Pazar payının büyümesi, talebin ve kullanımının artması, diğer bu statüdeki tüm enerji kaynakları gibi doğalgaz fiyatlarının da kırılmalı ve diğer dışsal değişkenlerden edilgen bir yapıya gelmesine sebep olmuştur. Özellikle fiyatlarında meydana gelen değişimlerin incelenmesi, enerji kaynağı olarak doğalgazı kullanan kurum ve kuruluşlar için hayati bir önem taşımaktadır.

Bu çalışmada da 1967-2013 yılları arasındaki konut doğalgaz fiyatları incelenerek, durağanlık yapısı araştırılmış, özellikle günlük fiyat veya endeks serileri gibi serilerde sıkça görülen Otoresif Koşullu Değişen Varyans (ARCH) yapısı içerip içermediği araştırılmıştır. ARCH model ailesinin özellikle günlük veya aylık serilere uygulanması literatürde sık karşılaşılan bir durum olmasına rağmen yıllık serilere çok fazla uygulanmaması, bu çalışmanın yapılmasındaki dürtülerden biri olmuştur.

ARCH modeli Engle (1982) tarafından geliştirilmiştir. Engle zaman serileri modellerinde ileri sürülen sabit varyanslılık varsayımını reddederek hataların sabit varyanslı olmadığını İngiltere enflasyon verilerini inceleyerek göstermiştir. ARCH sürecinin kullanılmaya başlamasıyla birlikte çeşitli uzantıları önerilmeye başlamıştır.

Engle 1982 yılındaki çalışmasında; geçmiş dönem gözlem değerleri kullanılarak tahmin edilen  $y_t$  rassal değişkeninin değerini temsil eden koşullu olasılık fonksiyonunu  $f(y_t / y_{t-1})$  şeklinde ifade edildiğini, bu noktadan hareketle bir sonraki dönemin tahminlenen varyansını da  $V(y_t / y_{t-1})$  şeklinde göstermiştir. Klasik ekonometri modellerinde koşullu varyans  $y_{t-1}$ 'e bağlı değildir. Fakat Engle 1982 yılındaki çalışmasında koşullu varyansın geçmiş dönem değerlerine de bağlı olduğunu göstermiştir.

#### 2. Bulgular



Şekil 1. 1967-2013 Doğal Gaz Fiyatları (\$) (Note: The graph shows data from 1970 to 2010)

Şekil 1. incelendiğinde, özellikle 1980 sonrası doğalgaz fiyatlarında artan bir trend olduğu görülmüştür.

Seride ARCH Etkisinin olup olmadığını araştırmak için ARCH-LM testi uygulanmıştır. Çeşitli gecikme değerleri için ARCH etkisini sınanan test sonuçları Çizelge 1'de verilmiştir.

Çizelge 1. Çeşitli Gecikme Uzunlukları İçin ARCH-LM Testi Sonuçları

Gecikme Uzunluğu	P değeri
1	0,00*
2	0,001*
3	0,005*
4	0,006*
5	0,006

ARCH etkisinin varlığı çeşitli gecikme uzunlukları için test edildiğinde 5 gecikmeye kadar seride ARCH etkisinin olduğu belirlenmiştir. Çalışmanın sonraki aşamalarında serinin sahip olduğu bu ARCH etkisinin modellenmesi için ARCH ve GARCH modelleri denenerek en uygun model belirlenmiştir. Ayrıca uygun olduğuna karar verilen modellerden bazıları içinde öngörü performansları karşılaştırılmış ve hangi modelin öngörü performansının daha iyi olduğu belirlenmeye çalışılmıştır.

Doğal logaritması alınmış seri için ARMA(2,1) modeli kullanılarak ARCH-GARCH modeli denemeleri yapılmış ve sonuçlar Çizelge 2. de verilmiştir.

Çizelge 2. Denenen Bazı ARCH-GARCH Modeli AIC Değerleri

Model	AIC Değeri
ARCH(4)	-2,082
GARCH(3,1)	-2,152
E-GARCH(3,2) ASD*=2	-1,870
T-ARCH(2) (T=1)	-2,070
I-GARCH(1,1)	-1,933
I-GARCH(2,2)	-1,83
P-GARCH(2,1)	-1,889

Çizelge 2 incelendiğinde son dönem etkilerini dikkate alan I-GARCH Modellerinin daha anlamlı sonuçlar verdiği görülmüştür ki, ilgili serinin Kartezyen grafiği incelendiğinde seride özellikle son dönemde daha düzensiz dalgalanmalar görülmektedir.

#### KAYNAKÇA

www.iea.org

Engle., R., F. 1982 Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation, *Econometrica*, Vol. 50, No. 4. (Jul., 1982), pp. 987-1007.

**Keywords:** Volatility, Natural Gas Price, ARCH-GARCH Models

## BİLDİRİ OTURUMLARI 1

### SESSION 1

#### Statistical Methods

## A GENETIC ALGORITHM FOR PORTFOLIO OPTIMIZATION

Derya TURFAN, Cagdas Hakan ALADAG, Ozgur YENIAY  
Hacettepe University, Faculty of Science, Department of Statistics, Ankara, Turkey

E-mails: [deryaturfan@hacettepe.edu.tr](mailto:deryaturfan@hacettepe.edu.tr) [aladag@hacettepe.edu.tr](mailto:aladag@hacettepe.edu.tr) [yeniay@hacettepe.edu.tr](mailto:yeniay@hacettepe.edu.tr)

### 1. Introduction

Portfolio optimization is one of the most important problems in the field of finance. For solving this problem, Harry Markowitz [4] developed mean-variance model as a quantitative tool which seeks to optimally allocation among different assets. He presented his famous model through quantifying portfolio return as mean and calculating variance as risk. According to the this model, an investor aims either the minimization of an objective function representing the portfolio variance (risk) for a given level of return or the maximization of an objective function representing the portfolio return for a given level of risk.

$N$  and  $R^*$  represent the number of assets available and the desired expected return, respectively. Let  $\mu_i$ ,  $\sigma_{ij}$ ,  $w_i$  be the expected return of asset  $i$  ( $i=1, \dots, N$ ), the covariance between assets  $i$  and  $j$  ( $i=1, \dots, N$ ;  $j=1, \dots, N$ ), the proportion ( $0 \leq w_i \leq 1$ ) held of asset  $i$  ( $i=1, \dots, N$ ), respectively. The general formulation of standard Markowitz's mean-variance model for the portfolio selection problem:

$$\text{Minimize } \sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_{ij} \quad (1)$$

subject to

$$\sum_{i=1}^N w_i \mu_i = R^* \quad (2)$$

$$\sum_{i=1}^N w_i = 1 \quad (3)$$

$$0 \leq w_i \leq 1, \quad i=1, \dots, N \quad (4)$$

Genetic algorithms (GAs) are a family of computational models inspired by natural evolution [2,3]. The theoretical foundations of GAs were originally developed by Holland [1]. Stancu et al. [5] stated that in GAs, an initial population containing constant number of chromosomes is generated randomly (regarding portfolio optimization, each chromosome represents the weight of an individual security) and an evaluation function is formed to evaluate the fitness of each chromosome, which defines if the chromosome represents a good solution. Using crossover, mutation and natural selection, the population will evolve towards a population that contains only the chromosomes with good fitness. The larger the fitness value is the better objective function the solution has. It is proven that better solutions are obtained by GAs in solving combinatorial optimization problems like portfolio selection. The aim is to find weights of the portfolio invested in each asset in order to maximize the portfolio return and minimize the portfolio risk.

### 2. The implementation and the obtained results

The accessible 28 stocks traded in Istanbul Stock Exchange 30 Index (ISE 30) were chosen. Therefore, each chromosome has 28 genes. The series for each stock contain weekly data between the dates 1 January 2012 and 30 June 2013. Totally 78 observations were taken as closure price for every week. The parameters of the proposed algorithm such as population size, crossover fraction, mutation fraction and iteration bound were taken as 100, 0.5, 0.25 and 200, respectively.

In this study, to provide support to the decision makers in the process of making a choice among different options, we propose an alternative solution approach which is based on genetic algorithms. The proposed genetic algorithm is introduced and it is applied to a real life problem. As a result of the implementation, 100 different solutions with different risk levels and return values are obtained and presented.

### REFERENCES

- [1] Holland J.H. (1975), Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control and artificial intelligence, Michigan: University of Michigan Press.
- [2] Holland J.H. (1976), Adaptation. In R. Rosen & F. M. Snell (Eds.), Progress in theoretical biology, Michigan: University of Michigan Press, 4, 263-293.
- [3] Goldberg D.E. (1989), Genetic algorithms in search, optimization and machine learning, Addison-Wesley Publishing Company.
- [4] Markowitz H. (1959), Portfolio selection: efficient diversification of investments, New York: Wiley.
- [5] Stancu S. and Predescu M.O. (2010), Genetic algorithm for the portfolio selection problem on the Romanian capital market, Proceedings of International Conference on Engineering and Meta-Engineering, Orlando, USA, 57-60.

### ÖZET

#### PORTFÖY OPTİMİZASYONU İÇİN BİR GENETİK ALGORİTMA

Finans alanında karşılaşılan en önemli problemlerden biri portföy optimizasyonudur. Portföy yönetiminde amaç, yatırımcının getiri beklentileri ve riske karşı tutumu doğrultusunda, portföye dahil edilecek varlıklara ve bu varlıkların oranlarına karar vermektir. Yatırımcının belirlediği kısıtlar, eldeki verilerin sayıca fazla olması gibi nedenlerden dolayı zor bir problemdir. Bu nedenle portföy optimizasyonunun çözümü için klasik yöntemlerin yanı sıra genetik algoritmalar gibi sezgisel algoritmalar da kullanılmaktadır. Bu çalışmada, optimal portföy oluşturmak için Markowitz'in ortalama-varyans modeline yeni bir genetik algoritma yaklaşımı uygulanmıştır. Yapılan uygulama sonucunda, uygun bir portföy oluşturmak için yatırımcıya, elde edilen farklı risk, getiri ve varlıkların oranlarını içeren yatırım alternatifleri sunulmuş ve yorumlanmıştır.

**Anahtar Kelimeler:** Portföy Optimizasyonu, Genetik algoritma, Risk, Getiri.

#### COMPARISON OF DISCRETIZATION METHODS FOR LEARNING BAYESIAN NETWORKS

Emre DÜNDER , Mehmet Ali CENGİZ  
Ondokuz Mayıs University, Faculty of Science, Department of Statistics, Kurupelit/SAMSUN  
[emre.dunder@omu.edu.tr](mailto:emre.dunder@omu.edu.tr) [macengiz@omu.edu.tr](mailto:macengiz@omu.edu.tr)

Bayesian Networks, also known as probabilistic belief networks or causal networks are graphical models that can efficiently represent n-dimensional probability distributions with a directed acyclic graph. Bayesian networks are used in many fields such as biology, medicine, chemistry etc. The advantage of BNs is their capacity for being used both as predictive and descriptive models. In prediction they constitute an efficient tool for solving different inference tasks. As a descriptive tool they possess the ability to efficiently represent the dependence or independence relationships among the random variables. The construction process of the Bayesian networks is called as learning. Learning Bayesian networks is known to be an NP-hard problem. The manual construction of BNs is a complex and highly time-consuming task and given the increasing availability of data in many domains, one of the areas that have seen a major activity in BN research is that of automatically learning the BN structure from data. Structure learning becomes a problematic operation with the differences of the distributions among the continuous variables.

Bayesian networks are powerful tools for handling problems which are specified through a multivariate probability distribution. In hybrid Bayesian networks, where both discrete and continuous variables appear simultaneously, the continuous variables should be normally distributed. The condition of normal distribution is solved with the discretization of the continuous variables. In this study we compared the affects of different discretization techniques to the quality of the structure. We implemented score based and hybrid learning algorithms for learning the structure of Bayesian networks with golden standard networks. We also applied structure learning algorithms both with continuous and discretized datasets. Experimental results are achieved with bnlearn package in R Project.

#### KAYNAKLAR

1. Neapolitan R. (2003), Learning Bayesian Networks, Prentice Hall Series in Artificial Intelligence.
2. Kim H. (2012), discretization, R package version 1.0-1, URL <http://cran.r-project.org/web/packages/discretization/index.html>
3. Scutari M. (2011), Measures of Variability for Graphical Models, Università degli Studi di Padova, Dipartimento di Scienze Statistiche.
4. Cobb B. R. Rumi R. and Salmeron A. (2007), Bayesian Network Models with Discrete and Continuous Variables, *Advances in Probabilistic Graphical Models*.

#### FUZZY LINEAR PROGRAMMING: AN APPLICATION

Nilüfer PEKİN ALAKOÇ<sup>a\*</sup>, Gültaç EROĞLU İNAN<sup>b</sup>, Ayşen APAYDIN<sup>c</sup>  
<sup>a</sup>Atilim University, Faculty of Engineering, Department of Industrial Engineering, 06836-İncek, Ankara, Türkiye. [npalakoc@atilim.edu.tr](mailto:npalakoc@atilim.edu.tr)  
<sup>b</sup>Ankara University, Faculty of Science, Department of Statistics, 06100-Tandoğan, Ankara, Türkiye. [geroglu@ankara.edu.tr](mailto:geroglu@ankara.edu.tr)  
<sup>c</sup>Ankara University, Faculty of Science, Department of Statistics, 06100-Tandoğan, Ankara, Türkiye. [aapaydin@ankara.edu.tr](mailto:aapaydin@ankara.edu.tr)

#### Fuzzy Linear Programming

Almost all real world problems include uncertainty. There exists many techniques for modeling or solving these problems. The fuzzy theory is one of the popular way of handling with the uncertainty. Since 1965, when the fuzzy set theory was first proposed by Zadeh [1], the theory has been improved by many scientists. Moreover, the applications of the theory has been performed in many fields.

Linear programming is the most frequently used optimization tool for solving real world problems. In any linear programming problem the aim is to minimizing or maximizing a linear objective function subject to linear constraints. The expression in Equation 1 gives the canonical form of a linear program:

$$\begin{aligned} \text{Minimize } & z = c^T x \\ \text{subject to } & Ax \geq b \\ & x \geq 0 \end{aligned} \quad (1)$$

where  $x$  represents the vector of decision variables,  $c$  is the vector of objective function coefficients and  $b$  is the vector of constraints right-hand sides and  $A$  is the matrix of technological coefficients. [2].

In any linear programming model, it is assumed that the data  $(c, A, b)$  is known exactly. However, the  $(c, A, b)$  input data of real world optimization problems usually do not known exactly. The input data are usually fuzzy because of incomplete information. Fuzzy linear programming has been suggested for these nondeterministic situations.

The fuzzy linear programming problem is obtained by replacing the crisp numbers by fuzzy numbers. In a fuzzy linear programming the vector  $c$ , the vector  $b$ , the matrix  $A$  or any combination of them might be defined as fuzzy numbers.

There are several distinct techniques for solving the problems that include fuzzy information. The most popular and commonly used approaches are Verdegay's approach, Werner's approach, Zimmermann's approach and Chanas's approach [3].

In this study, a review of the fuzzy linear programming approaches is presented. As an application of the FLP a real world problem is solved by the most suitable approach.

#### REFERENCES

- [1] Zadeh L.A. (1965), Fuzzy sets, Information and Control, 8, 338-353.

- [2] Nash S.G. and Sofer A. (1996), Linear and nonlinear programming, McGraw-Hill, New York.
- [3] Lai Y.J. and Hwang C.L. (1992), Fuzzy Mathematical Programming, Berlin-Heidelberg, Springer-Verlag.
- [4] Tanaka, H., Okuda, T. and Asai, K. (1978), On fuzzy mathematical programming, J. Cybernetics, 3, 37-46.
- [5] Zimmermann, H.J. (1983), Fuzzy mathematical programming, Comput. & Ops. Res., 10, 4, 291-298.

## ÖZET

### BULANIK DOĞRUSAL PROGRAMLAMA: BİR UYGULAMA

Günümüz dünyasında birçok gerçek hayat problemi belirsizlik içermektedir. Bulanık mantık bu belirsizliklerle mücadelede kullanılan önemli tekniklerden biridir. Zadeh'in 1965 yılında yapmış olduğu çalışmada bulanık küme teorisini açıklamıştır. Bu tarihten sonra bulanık mantık çeşitli bilim alanlarında yaygın olarak geliştirilmiş ve uygulamaları yapılmıştır. Doğrusal programlama problemlerinde de belirsizlik içeren durumlar bulanık mantık yaklaşımları kullanılarak incelenmektedir. Bu belirsizlikler bulanık sayılarla ifade edilecek, bulanık doğrusal programlama probleminin amaç fonksiyonunda, kısıtlarında, sağ yan değerlerde olabilmektedir. Bu çalışmada, bulanık doğrusal programlama ile bir gerçek hayat probleminin çözülmesi hedeflenmektedir.

**Anahtar Kelimeler:** Doğrusal programlama, Bulanık doğrusal programlama, Bulanık küme teorisi, Bulanık sayılar.

## AN APPLICATION OF SHOCK MODELS TO THE INVENTORY

Ayşe Tansu<sup>1</sup>, Adeoye Olaitan

Cyprus International University, Engineering Faculty, Industrial Engineering Department, Nicosia, Cyprus

Cyprus International University, Engineering Faculty, Industrial Engineering Department, Nicosia, Cyprus

\*ayset@ciu.edu.tr

## ABSTRACT

The real world systems are subject to online deterioration. Yeh Lam's model a new class of models called  $\delta$ -shock models in which failure was dependent on the frequency of shocks. Rangan et.al presented this model by generalizing Yeh Lam's results for renewal shock arrivals and random threshold. In this paper inventory is considered as a random. Demands occur at random and Inventory is ordered in single units.

Keywords - Inventory,  $\delta$ - model, random threshold

## 1. INTRODUCTION

This study involves the inventory application. In such a modeling approach, each order has a random lead time similar to a  $\delta$  - shock model. If the time between demands is less than  $i$ , the demand is satisfied. If the time between demands is greater than  $i$  then demands are not satisfied. In such a modeling approach, a system is subject to randomly occurring demands, each of which adds a nonnegative random quantity to the accumulated demand process  $\{W(t), t \geq 0\}$ .

Study has shown that systems are likely to experience demand if the time between two successive demand is very small. While the earlier shock models concentrated solely on the magnitude of the damage caused by the shocks, Yeh Lam and Zhang's model paid attention to the frequency of the shocks. Recently A. Rangan and A. Tansu [1] got some results on a new class of shock model by analyzing the statistical characteristics of  $\delta$  - shock model, thereby establishing an optimal replacement and repair model for deteriorating systems.

In this paper the demand is considered instead of shock arrivals. It analyzes the existing models and uses demand as a random instead of shock, and ordering inventory in single units.

## 2. NOTATION USED

Z: Random variable denoting the time between two successive demands.

$f_Z(\cdot)$ ,  $F_Z(\cdot)$ ,  $\bar{F}_Z(\cdot)$ : Probability density, Cumulative distribution, and survivor function of Z.

$i$  = Random variable denoting the random lead time.

$g_I(\cdot)$ ,  $G_I(\cdot)$ ,  $\bar{G}_I(\cdot)$ : Probability density, Cumulative distribution and survivor function of I.

$W(t)$ : Random variable denoting time between two successive orders.

$kw(t)$ ,  $Kw(t)$ ,  $\bar{K}_W(t)$ : Probability density, Cumulative distribution and survivor functions of W.

N(t): Counting variable denoting the number of ordered inventory in (0, t).

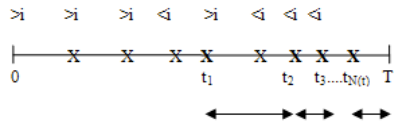
$$M(t) = E\{N(t)\}.$$

$L_f(s)$ : Laplace transform of the density function of  $f(t)$ .

### 3. THE MODEL ASSUMPTION

The model is governed by the following assumptions:

- A1. A new system is put on operation at time  $t = 0$ . The system on demand is satisfied and the times are assumed to take negligible amount of time. The system is inventory ordered at a fixed time  $T$ , which is the decision variable.
- A2. The system is subject to demands. The time between demands  $Z$  is assumed to be independently and identically distributed with distribution function  $F_z(\cdot)$ .
- A3. The demand is satisfied if the time between demands is less than random lead time  $i$ , and demands are not satisfied if the time between demands is greater than random lead time  $i$ .
- A4. The satisfying orders of systems take negligible amount of time to complete.
- A5. Threshold value  $I$  is a random variable with distribution function  $G_i$ .
- A6. The demands and random lead time are independent of each other.
- A7. A typical sample path of the operation of the system under the stated assumptions when the threshold value  $I$  is a constant  $i$  is given in the Figure below.



X- Demand

X- Ordered inventory

Fig.1. A typical sample path of the operation of the system

### 4. CONCLUSION

The paper focused on the frequency of demand to the system while the ordering is done if the demand is necessary. Demands occur at random and inventory ordered in single units. Two models are considered. In the first one the demands are assumed to be exponentially distributed while in the second model the demands are uniformly distributed. The different demand arrival distributions can be done and demand satisfaction can be analyzed according to the distributions.

### REFERENCES

- [1] A. Rangan and A. Tansu "Some Results on a New Class of Shock Models", Asia-Pacific Journal of Operational Research, vol 27, no 4 pp 503-515.

### THE POLICY DETERMINANTS OF THE REGIONAL DISPARITIES IN THE HIGHER EDUCATION IN MATHEMATICAL SCIENCES: A MULTIVARIATE CASE STUDY OF LESSONS TO BE DRAWN FROM THE PRIVATIZATION EFFORTS OF THE TURKISH HIGHER EDUCATION SYSTEM WITH SPECIAL REFERENCE TO STATISTICAL SCIENCES.

Erkan Abdülgaffar AĞAOĞLU

Yeditepe University, Faculty of Economic and Administrative Sciences, The Department of Business Administration 34755 Kayışdağı Campus, İstanbul, Turkey

[agaoglu@yeditepe.edu.tr](mailto:agaoglu@yeditepe.edu.tr); [agaoglu\\_ag@yahoo.com](mailto:agaoglu_ag@yahoo.com)

With the initiation of its first foundation university in 1985, Turkey had 31 universities in total during 1990. At the beginning of year 2011, however, there were 166 universities (including 9 private vocational schools). 102 of these were public universities (61.45%); 55 (33.50%) were initiated by foundations; and there were 9 (5.40%) private vocational schools of higher education. Student population of the foundation universities was expected to approach 10% of the total. Although the phenomenon is relatively very new and expanding on a very rapid speed, it serves as an extremely important primary academic learning exercise

This research focuses on drawing lessons from the "Mathematical Sciences Education" perspective of the said learning exercise, drawn from the "Faculties of Arts and Sciences" with programs placing students with the same type of examination scores. Within this perspective and background characteristics, this study would serve as the very first of its nature in the Turkish literature of higher education.

Student selection characteristics of the Turkish University Entrance Examination System (ÖSYS), however, is oriented to department wise (academic program wise) selections utilizing various test modules, which makes it almost impossible for all "Mathematical Sciences" to be placed within the same data matrix. We therefore pick up "Statistical Sciences" as a prototype, and place them within the cluster of programs taking student inputs with the same type of examination scores. The analysis of the science cluster programs of "Faculties of Arts and Sciences", or simply "Faculties of Sciences" is based upon Physics, Chemistry, Mathematics, Biology, and Statistics. We, on the other hand, bring programs of the Faculties of Administrative Sciences and Economics side by side, and analyze "Econometrics" which lies within these faculties with different placement criterions. In comparison to the science cluster.

We base this initial study upon bench marking "Business Administration and Economics Departments" of the universities as a non mathematical science cluster to compare and contrast "Econometrics" programs of these faculties. The advantage of such a selection is that almost all universities tend to have these favorite departments in their portfolios. Thus, it is possible to handle "Statistical Sciences" without any missing data problem; moreover such a minimum missing data approach would easily avoid any "Validity" and "Reliability" problems, which are mostly confronted in quantitative approach such as of this study. Most importantly, this approach is very much in line with the theme of this conference; intending to have a session on "Statistics Education". Such a motive will help us to initiate an inventory perspective within the Turkish Higher Education System, which would then be able to provide a strategic planning base for micro and macro analysis. Further strength of this study is that it cross checks the regional attraction clusters for the above domain.

The next very basic aim of this study is to cluster the above population of university departments in accordance with "Turkish Geographical Regions", namely, Black Sea, East Anatolia, Central Anatolia, South-East Anatolia, Mediterranean, Aegean, and Marmara regions, with reference to the "Regional Prosperity Index" of the Turkish State Planning Organization. All institutions of the Higher Education are then categorized according to the above clusters. Each iteration in the above analysis is repeated region wise; and the profile is interpreted through comprehensive



“Descriptive Statistics”, and also by “Chi-Square” test. Firstly, data for both 2003 and 2009 is subjected to a “Multiple Regression” run; further elaborated through “Logit & Probit” analysis, for grouping of Foundation and Public Universities. The general regression model is run with Entrance Scores for the considered Departments as Dependent Variable(Y); Population(X1), Turkish State Planning Organization Prosperity Index of the State(X2), and the percentage of seats filled by ÖSYM(X3) as independent variables. The run is then replicated for 0-1 dummy-dependent variable; with 0's for foundation and 1's for public universities. Results indicate an increasing tendency of student preferences for populated regions, with an increasing impact for larger “Economic Prosperity Regions”, thus rendering plenty of evidence for “Regional Disparities”. The regional implications of

such an early warning model give very important repercussions for Strategic Planning dilemma within the Turkish Higher Education System, and renders enormous re-planning as well as growth potential for further research. . Low levels of Type I, and Type II errors, with strong descriptive statistics results are encouraging, reflecting prospects for further platforms, which may include cross cultural studies. Further research could also be possible with the inclusion of TRNC Universities, Foreign Universities on the list of ÖSYM, full coverage of vocational school departments, and many other areas. The study also depicts important repercussions of “Statistics Education” and “Econometrics Education” under two different platforms. In this respect, this study can be counted as the first academic platform assess the SEI (Statistics Education Inventory).

In short, the study clearly indicate that the bottleneck do exist, and such a bottleneck exists when the competition in the higher education industry is “Imperfect” due to the presence of two groups with conflicting competitive edges. Any policy orientation what-so-ever, needs to concentrate upon this dilemma, and take corrective actions before any long term macro-based strategy can be designated for long terms, for example, beyond 2023. This study is expected to be the major theme of any other research intending to cover “Inventorial” perspective of the Turkish Higher Education System.

**Key Words:** Analysis of Higher Education and Research Institutions; Regional Economic Development and Development Disparities; Education and Research Institutions; Turkish Higher Education System; Mathematical Sciences; Econometrics; Statistical Sciences; Discriminant Analysis; Turkish Foundation Universities; Turkish Public Universities; Quality of Public and Private Education Systems; Strategic Planning of the Higher Education, Early Warning Models.

## SOME ASPECTS OF SIMULATION WITH STOCHASTIC INVESTMENT MODELS

Şule ŞAHİN, Hacettepe University, Beytepe, Ankara, Turkey, [sule@hacettepe.edu.tr](mailto:sule@hacettepe.edu.tr)  
A.David WILKIE, InQA Actuarial Software, Bournemouth, UK, [david.wilkie@inqa.com](mailto:david.wilkie@inqa.com)

### Introduction

In this paper we describe some of the theoretical and practical aspects of the Monte Carlo simulation of an investment model. We discuss the input and output set of variables of the models and different ways of simulation. We also discuss the initial conditions required defining the “neutral” initial conditions and “neutralised parameters” which are found based on the “neutral” initial condition. Finally we suggest using additional information in the first periods of the simulation to adjust the formulae parameters for a limited “select period”. We use Wilkie model as an example but most of the aspects apply to any similar model used over time.

### Wilkie Model

Wilkie model is a stochastic investment model developed by A. D. Wilkie which describes the behaviour of various economic factors such as price inflation, wage inflation, dividend yields, share dividends, long term and short term interest rates [2][3][4]. We use Wilkie model to describe the aspects of the Monte Carlo simulation of the stochastic investment models as an example.

### Practical and Theoretical Aspects

In this paper we consider the following aspects of the Monte Carlo simulation of the stochastic investment models.

#### 3.1. Monte Carlo Simulation

Monte Carlo simulation is one of the common features of stochastic investment models. The models in general cannot be treated analytically, except in limited circumstances but they are designed to be used in Monte Carlo simulation exercises [1].

#### 3.2. Initial Conditions

Every stochastic investment model needs some values of the state space at time  $t=0$ . These values are called “initial conditions”. Lee and Wilkie introduced the “neutral” initial conditions and “neutralised” parameters [1]. Neural initial conditions might be mean values for linear models and median or long run means for the non linear models.

#### 3.3. Select Period

Economic period the models cover affects the parameters of the models. For example Wilkie model forecast of inflation is based on the history of inflation itself and it has a relatively high standard deviation. The standard deviation of the forecast for the rate of inflation in the Wilkie model over one year ahead is the same as over any future year and this might be criticized because for short term forecasts the standard deviation can be considered as relatively high. To overcome this problem we propose adjusted parameters for the short term which Lee and Wilkie named as “select period”.

### REFERENCES

- [1] Lee P.J. and Wilkie A.D. (2000), A Comparison of Stochastic Asset Models, Proceedings of the 10<sup>th</sup> AFIR Colloquium, Tromsøe.
- [2] Wilkie A.D. (1986), A Stochastic Investment Model for Actuarial Use, Transactions of the Faculty of Actuaries, 39, 341-381.

[3] Wilkie A.D. (1995), More on a stochastic asset model for actuarial use, British Actuarial Journal, I, 777-964.

[4] Wilkie A.D., Şahin Ş., Cairns A., Kleinow T., (2010), Yet More on a Stochastic Economic Model: Part 1: Updating and Refitting, 1995 to 2009, Annals of Actuarial Sciences, 5(1), 53-99.

#### ABSTRACT

##### SOME ASPECTS OF SIMULATION WITH STOCHASTIC INVESTMENT MODELS

Monte Carlo simulation of an investment model defined by a set of time series formula requires consideration of a number of practical and theoretical aspects. In this paper we describe some of them, using the “Wilkie model” as an example. We discuss the variables that can form the working set, the input set and the output set, all of which may be different. There are different ways of simulating, either in a linear parallel structure or in a branching tree structure. We then discuss the initial conditions required, which may be market conditions at some date, or may be “neutral” initial conditions, which may be defined in different ways. What we call “neutralizing parameters” may have role, and we discuss how these may be found. Finally we suggest using additional information in the first periods of the simulation to adjust the formulae or parameters for a limited “select period”. Although the Wilkie model is used as an example, most of the aspects apply to any similar model used for simulation over time.

**Key Words:** Stochastic investment models, initial conditions, Monte Carlo simulation, select period, Wilkie model.

## BİLDİRİ OTURUMLARI 1

### SESSION 1

#### *Olasılık ve Stokastik Süreçler*

## BÜTÜNLEŞİK REKOR DEĞERLER

Agâh KOZAN\*, Halil TANIL

Ege Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 35100, B-Blok, Kat:2, Bornova, İzmir, Türkiye  
agah.kozan@ege.edu.tr, halil.tanil@ege.edu.tr

### Giriş

Klasik Rekor Model ile ilgili ilk çalışmayı 1952'de Chandler yapmıştır. Kendisi, rekor değerlerin stokastik davranışlarını incelemiştir. O günden bu yana bu alanda pek çok çalışma yapılmıştır.

Klasik rekor modelin tanımına göre,  $\{X_i\}_{i \in \mathbb{N}^+}$ , birbirinden bağımsız ve aynı sürekli  $F(x)$  dağılım fonksiyonu ve  $f(x)$  olasılık yoğunluk fonksiyonuna sahip rasgele değişkenler dizisi olmak üzere,  $X_j$  gözlemi, kendinden önceki tüm gözlemleri aşıyorsa buna "Üst Rekor Değer" ve kendinden önceki tüm gözlemlerden küçük oluyorsa "Alt Rekor Değer" adı verilir [1].

Klasik rekorların yanı sıra, bazı çalışmalarda, sadece o ana kadar meydana gelen en büyük ve en küçük değer ikilisi kayıt altına alınmak istenebilir. Bu şekilde toplanan değerlere ise literatürde "Güncel Rekorlar" adı verilmektedir.

$L'_n$  ve  $U'_n$ , sırasıyla güncel alt ve üst rekor değerleri gösterebilir. Başlangıçtaki güncel alt ve üst rekor değerler birbirine eşittir,  $L'_0 = U'_0 = X_1$ .  $(L'_n, U'_n)$  aralığı "Rekor Kapsamı" (Record Coverage) ve  $R_n = U'_n - L'_n$  değeri ise "Rekor Açıklığı" (Record Range) olarak adlandırılmaktadır.  $X_j$  dizisinde alt veya üst rekor meydana geldiğinde yeni bir Rekor Açıklığı meydana geleceği açıkça görülmektedir [2].

Bu çalışmada yukarıdakine benzer bir şekilde, zaman içerisinde meydana gelen alt ve üst rekor değerlerin bir arada toplanıp sıralanmasıyla oluşturulan listeye ait sıralanmış rasgele değişkenlerin ortak dağılımı ile ilgilenilmiştir.

Bir araştırmacı, ilgilendiği sistemin performansını sadece üst veya sadece alt rekor değerler listesine dayalı olarak değil de her iki listenin de bütünlük bir yapı içerisinde bulunduğu bir model ile değerlendirmek isteyebilir. Böylesi bir durumda, önerdiğimiz model geçerli olacaktır. Örneğin, bir araştırmacı kuraklık ve sel gibi yağışa bağlı olarak önem alınması gereken iki olağanüstü durumu birbirinden ayırmak isteyebilir.

**Tanım:**  $i$  zaman indeksi olmak üzere,  $X = \{X_i\}_{i \in \mathbb{N}^+}$ , birbirinden bağımsız ve aynı  $F$  dağılımlı rasgele değişkenler dizisi olsun.  $\Delta(1) = 1$ ,  $\Delta(2) = 2$  ve  $r = 3, \dots, n$  için;

$\Delta(r) = \min \{i > \Delta(r-1) : X_i < X_{1:\Delta(r-1)} \text{ veya } X_i > X_{\Delta(r-1):\Delta(r-1)}\}$  " $r$ 'nci Bütünlük Rekor Zamanı" olmak üzere,  $\{X_{\Delta(r)}\}_{r=1, \dots, n}$  rasgele değişkenleri "**Bütünlük Rekor Değerler Listesi**" olarak adlandırılmaktadır.

$Y_i = X_{\Delta(i)}$  olmak üzere  $\{Y_i\}_{i=1, \dots, n}$  listesinin sıralanmasıyla oluşturulan listeye  $(Y_{1:n} < Y_{2:n} < \dots < Y_{n:n})$  "Sıralanmış Bütünlük Rekor Değerler Listesi" adı verilir.

Bu çalışmada yukarıda tanımlanan sıralanmış bütünlük rekor değerler listesinin elemanlarının ortak olasılık yoğunluk fonksiyonu incelenecektir.

### KAYNAKLAR

[1]Arnold, B.C., Balakrishnan. N. and Nagaraja, H.N., 1998. Records. John Wiley & Sons, New York

[2]Houchens, R.L., 1984. Record value, theory and inference. Ph.D. Dissertation, University of California, Riverside.

### ABSTRACT

#### MIXED RECORD VALUES

Consider an infinite sequence of independent and identically distributed (iid) continuous random variables. We introduce the new concept of "Mixed Record Values" which can be described as ordered random variables constructed from a collection of n lower and upper record values collected in a timeline. We obtain the joint probability density function of these n Mixed Record Values.

**Key Words:** Mixed records, Record value theory, Current records

## GENELLEŞTİRİLMİŞ BETA MÜDAHALELİ YARI-MARKOV RASTGELE YÜRÜYÜŞ SÜRECİ İÇİN ASİMPOTİK SONUÇLAR

Tülay KESEMEN

KTÜ, Fen Fakültesi, MatematikBölümü, 61080, Trabzon [tkesemen@gmail.com](mailto:tkesemen@gmail.com)

Zafer KÜÇÜK

KTÜ, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, 61080 Trabzon [zkucuk@ktu.edu.tr](mailto:zkucuk@ktu.edu.tr)

Çisem ÖCAL

KTÜ, Fen Fakültesi, Matematik Bölümü, 61080, Trabzonocalcistem@gmail.com

Tahir KHANİYEYEV

TOBB Ekonomi ve Teknoloji ÜniversitesiMühendislik Fakültesi, Endüstri Mühendisliği Bölümü

06560 Söğütözü, Ankaratahirkhaniev@etu.edu.tr

### 1.Özet

Bu çalışmada, kesikli şans karışımı bir yarı-Markov rastgele yürüyüş süreci  $(X(t))$  ele alınmıştır. Bu süreç matematiksel olarak tanımlanmış ve ergodik dağılımının ilk dört momenti için kesin ifadeler elde edilmiştir. Bunlardan yararlanarak,  $\{\zeta_n\}, n \geq 0$  rastgele değişkenler dizisinin dağılımı  $s, S, \alpha, \beta > 0$  olmak üzere  $(s, S, \alpha, \beta)$  parametrelili genelleştirilmiş beta dağılımına sahip olması durumunda ve  $E(\zeta_n) \rightarrow \infty$  iken sürecin ergodik dağılımının ilk dört momenti için asimptotik sonuçlar elde edilmiştir. Bunlara ilaveten, bu asimptotik sonuçlardan yararlanarak, sürecin ergodik dağılımının çarpıklık ve basıklık katsayıları için de asimptotik açılımlar elde edilmiştir. Sonunda da elde edilen asimptotik açılımların doğruluğu Monte Carlo simülasyon metoduyla test edilmiştir.

### 2.Giriş

Fizik, kimya, biyoloji ve iktisat bilim alanlarının yanı sıra kuyruk, güvenilirlik, sigortacılık ve matematiksel finans gibi çalışma alanlarında karşılaşılan birçok ilginç problemlerin incelenmesinde ve ifade edilip çözülmesinde yenileme, ödüllü yenileme, rastgele yürüyüş süreçleri ve onların çeşitli modifikasyonları geniş bir ölçüde kullanılmaktadır. Fakat bu süreçlerin her yönüyle araştırılması teorik açıdan zordur. Bu zorluğu ortadan kaldırmak için 1900'lü yıllardan sonra, ya benzetim yöntemleri kullanarak bilgisayar yardımı ile sayısal sonuçlar elde etmek ya da integraller için asimptotik yöntemlere başvurularak yaklaşık ve sade ifadeler elde etmek üzere iki yönde araştırmalar yoğunlaştırılmıştır. Bunlarla ilgili son yıllarda asimptotik yöntemlerin uygulanmasına ait birçok değerli çalışmalar ortaya konulmuştur ([2], [3], [4]).

Simon C. Parker (1998) genelleştirilmiş beta dağılımını ABD'nin gelir vergisi verilerine uygulayarak pareto ve lognormal dağılımdan daha esnek bir yapıya sahip olduğunu göstermiştir [5]. Beta dağılımının böyle esnek bir yapıya sahip olmasından dolayı reel hayatta ortaya çıkan problemlere karşı daha fazla uyumlu olabileceğinden bu çalışmada, müdahale çeşidi genelleştirilmiş beta dağılımı olarak ele alınmıştır ve süreç matematiksel olarak inşa edilmiştir. Sürecin ergodik dağılımının ilk dört momenti için  $\gamma \rightarrow \infty$  iken üç terimli asimptotik açılımları elde

edilmiştir. Ayrıca, bu asimptotik sonuçlardan yararlanarak, sürecin ergodik dağılımının varyansı, çarpıklık ve basıklık katsayıları için de asimptotik açılımlar elde edilmiştir. Sonuçta, Monte Carlo simülasyon yöntemi uygulanarak, ergodik momentler için elde edilen asimptotik sonuçların doğruluğu test edilmiştir.

**Anahtar Kelimeler:** Yarı-Markov rastgele yürüyüş süreci, kesikli müdahale, asimptotik açılımlar, genelleştirilmiş beta dağılımı.

### KAYNAKLAR

- [1] Feller W.(1971), Introduction to Probability Theory and Its Appl. II. J. Wiley, N.Y.
- [2] Kesemen T. (2013), On the semi-Markovian Random Walk With Delay and Weibull Distributed Interference of Chance, Hacettepe Journal of Mathematics and Statistics, in press.
- [3] Khaniyev T., Aksop C. (2011), Asymptotic Results for an Inventory Model of Type (s,S) With a Generalized Beta Interference of Chance, TWMS J. App. En. Math. V.1,N.2, 223-236.
- [4] Khaniyev T.A., Mammadova Z. (2006), On the Stationary Characteristics of the Extended Model of type of (s,S) With Gaussian Distribution of Summands, Journal of Statistical Computation and Simulation, 76(10), 861-874.
- [5] Parker, Simon C. (1999), The Generalized Functions for the Size Distribution of Earnings, Economics Letters, Elsevier, 62, 2, 197-200.

### ABSTRACT

#### ASYMPTOTIC RESULTS FOR THE SEMI-MARKOVIAN RANDOM WALK WITH GENERALIZED BETA DISTRIBUTED INTERFERENCE OF CHANCE

In this paper a semi-Markov random walk process  $(X(t))$  with a generalized beta distributed of chance is considered.  $X(t)$  is constructed mathematically and exact formulas for the first four moments of the ergodic distribution of the process are obtained. Using these expressions the asymptotic expansions for the first four moments of the ergodic distribution of the process are found as  $E(\zeta_n) \rightarrow \infty$  when the random variable  $\zeta_n$  has a generalized beta distribution with parameters  $(s, S, \alpha, \beta), s, S, \alpha, \beta > 0$ . Moreover, the asymptotic expansions for the skewness and kurtosis of the ergodic distribution of the process  $X(t)$  are established. Finally, the accuracy of the approximation formula is tested by the Monte Carlo simulation method.

**Key Words:** Semi-Markov random walk process, discrete interference of chance, asymptotic expansions, generalized beta distribution.

## İKİNCİL REKOR DEĞERLER

Halil TANIL

Ege Üniversitesi, Fen Fakültesi, İstatistik Bölümü, B-blok, Kat:2, Bornova, İzmir, Türkiye. halil.tanil@ege.edu.tr

### Giriş

Sayılabilir sonsuz bağımsız ve aynı sürekli  $F$  dağılımlı zaman sıralı rasgele değişkenler dizisi  $\tilde{X} = \{X_j\}_{j \in \mathbb{N}^+}$  ile gösterilsin.  $1 \leq i \leq m$  için  $X_1, X_2, \dots, X_m$ 'lere dayalı  $i$ 'inci sıra istatistiği  $X_{i,m}$  olsun. Rekör değerlerin tanımı ve bu değerlere ait dağılım teorisi, uç (extreme) hava koşullarını modelleme motivasyonu ile ilk kez Chandler (1952) tarafından verilmiştir. Örneğin,  $X_j$  gözleminin bir üst (upper) rekör değer olması için  $X_j > X_{j-1:j-1}$  şartı sağlanmalıdır. Ayrıca,  $\tilde{X}$  dizisinin üst rekör değer zamanları  $U(1) = 1$  ve  $r > 1$  için  $U(r) = \min\{j : j > U(r-1), X_j > X_{j-1:j-1}\}$  olmak üzere ilk  $r$  adet üst rekör değer  $X_1 < X_{U(2)} < X_{U(3)} < \dots < X_{U(r)}$  şeklinde gösterilmektedir.

Diğer yandan, bazı alanlarda çalışan araştırmacılar (örneğin, yaşam sigortacılığı alanında), uç değerleri 'aykırı değerler' (outliers) olarak gördükleri için rekör değerler yerine ikinci en büyük (veya en küçük), üçüncü en büyük (veya en küçük) gibi  $k$ 'ıncı en büyük (veya en küçük) değerlerle çalışma eğilimi gösterirler. İşte bu motivasyonla literatürde ilk kez Dziubdziela ve Kopociński (1976), üst  $k$ -rekör değerler kavramını tanımlamışlar ve dağılım teorisini vermişlerdir. Literatüre bakılırsa, üst  $k$ -rekörlerin farklı şekillerde tanımlandığı görülebilir. Bizim burada vereceğimiz tanım Arnold ve diğ. (1998)'ne aittir:  $\tilde{X}$  dizisinin üst  $k$ -rekör zamanları  $U_k(1) = k$  ve  $r > 1$  için  $U_k(r) = \min\{j : j > U_k(r-1), X_j > X_{U_k(r-1)-k+1:U_k(r-1)}\}$  olmak üzere, ilk  $r$  adet üst  $k$ -rekör değer  $X_{U_k(1)-k+1:U_k(1)} < X_{U_k(2)-k+1:U_k(2)} < \dots < X_{U_k(r)-k+1:U_k(r)}$  şeklinde ifade edilir.

Bu çalışma aşağıdaki gibi bir motivasyondan hareketle ortaya çıkmıştır: "Varsayalım ki, bir araştırmacı dünya rekorları listesinin her zaman gölgesinde kalmış ve bu listeye girememiş **en başarılı** (en iyi) **dereceleri** kaydetmek ve incelemek istemektedir." Bu motivasyon "İkincil Rekör Değer" adını verdiğimiz, "önceki rekör değerler listesinde yer almayan ve güncel rekör değere en yakın değer" şeklinde tanımladığımız yeni bir kavramı ortaya çıkartmaktadır. Bir başka deyişle, tüm rekör değerleri silinmiş  $\tilde{X}$  dizisinin rekorlarına ikincil rekör değerler adını vermekteyiz. Bu tanıma göre,  $\tilde{X}$  dizisinin ikincil üst rekör değer zamanları  $S(1) = \min\{j : X_j < X_{j-1:j-1}, j > 1\}$  ve  $r > 1$  için  $S(r) = \min\{j : X_{S(r-1)} < X_j < X_{j-1:j-1}, j > S(r-1)\}$  olup, ilk  $r$  adet ikincil üst rekör değer  $X_{S(1)} < X_{S(2)} < \dots < X_{S(r)}$  şeklinde gösterilmiştir. İkincil üst rekör değer kavramını irdelemek için aşağıdaki zaman sıralı gözlem dizisini ele alalım:

$$3.0, 3.1, 2.8, 2.7, 1.5, 3.6, 0.5, 3.5, 2.9, 3.9, \dots \quad (1)$$

Yukarıdaki dizinin güncel üst rekör değerleri ve güncel üst ikincil rekör değerleri Çizelge 1'de gösterilmiştir.

**Çizelge 1. Yukarıda verilen (1) gözlem dizisinin güncel rekör ve ikincil rekorları**

Gözlem sıra no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Gözlem değeri	3.0	3.1	2.8	2.7	1.5	3.6	0.5	3.5	2.9	3.9					

Güncel üst rekör															
Güncel ikincil üst rekör															

Bu çizelge dikkate alınrsa, ilk 11 gözlemlerde güncel üst rekör değerin 4.1 ve önceki üst rekör değerlerden oluşturulan listenin  $\{3.0, 3.1, 3.9\}$  olduğu, güncel ikincil üst rekör değerin ise 3.4 olduğu görülecektir. Yani önceki üst rekör değerler listesine girememiş ve güncel üst rekör değere en yakın değerin 3.4 ile 9. gözlem olduğu açıktır.

Üst rekör değerlerin aykırı değerler olarak görüldüğü durumda ikinci en büyük değerler de aykırı değerler olarak algılanabilir (veya algılanmalıdır). Örneğin, bir sayı dizinin son iki üst rekoru aykırı değer ise, bu dizinin güncel en büyük ikinci değeri de kesin olarak aykırı değerdir. Oysa, bu dizinin güncel ikincil üst rekör değeri bir aykırı değer olabilir veya olmayabilir. İşte bu sebeple, üst rekorları aykırı değerler olarak gören araştırmacılar,  $k = 2$  için üst  $k$ -rekörler yerine önerdiğimiz ikincil üst rekör modelini tercih edebilirler.

Bu çalışmada, yukarıda tanımladığımız  $\tilde{X}$  dizisinin ilk  $r$  adet ikincil üst rekör değerlerinin ortak olasılık yoğunluk fonksiyonu yinelemeli bir yaklaşımla elde edilmiş, ayrıca geçerli üst rekör ve geçerli ikincil üst rekör değerler bilindiğinde, bir sonraki ikincil üst rekör değerinin koşullu olasılık yoğunluk fonksiyonu verilmiştir.

### KAYNAKLAR

- [1] Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1998). Records, John Wiley & Sons, New York.
- [2] Chandler K.N. (1952), The distribution and frequency of record values, J. Roy. Statist. Soc. B 14, 220–228.
- [3] Dziubdziela W. and Kopociński W. (1976), Limiting properties of the k-th record values, Appl. Math. 15, 187-190.

### ABSTRACT

#### SECONDARY RECORD VALUES

Consider an infinite sequence of independent and identically distributed (iid) continuous random variables. We introduce the new concept of "secondary record value" which is not belong to the list of previous records and closest the current record value in this sequence. We obtain the joint probability density function (pdf) of first  $r$  secondary record values, recursively. We also obtain the conditional pdf of next secondary record value given by the values of current secondary record and current record.

**Key Words:** Secondary records, Ordinary records, Distribution theory

### AÇIK JACKSON MODELİNDE MOMENTLERİN BULUNMASI

Vedat SAĞLAM<sup>1</sup>, Murat SAĞIR<sup>1\*</sup>, Erdiç YÜCESOY<sup>1</sup>, Mülgan ZOBU<sup>2</sup>, Tolga ZAMAN<sup>1</sup>

<sup>1</sup>Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Ana Bilim Dalı, 55139 ,Samsun, TÜRKİYE

<sup>2</sup>Amasya Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Ana Bilim Dalı, Amasya, TÜRKİYE

E-mail: [istatistikci\\_murat@hotmail.com](mailto:istatistikci_murat@hotmail.com)

Model  $r$  – aşamalı  $M/M/c$  tipli kuyruk sistemlerinden oluşur. Müşteriler sisteme  $\lambda$  parametrelili Poisson akımı ile gelmektedir. Her bir bekleme hattında keyfi sayıda bekleme olabilir, i. sistemin her bir kanalında hizmet süresi  $(\eta_i)$ ,  $\mu_i$  parametrelili üstel dağılıma sahiptir ve  $\eta_i$  ( $i = 1, 2, \dots, c$ ) ler bağımsızdır, her bir müşteri önce 1. sistemde daha sonra 2. sistemde ve bunun gibi  $r$ . sistemde hizmetini bitirerek genel sistemden ayrılır. Bu çalışmada  $r = 2$  ve  $c = 1$  alınarak sistem analiz edilmiş, performans ölçütleri ve bunların optimal değerleri elde edilmiştir. Bununla birlikte her aşamadaki müşteri sayılarının bağımsızlığı ve her aşamada kalma sürelerinin bağımsızlığı sırasıyla iki teorem yardımıyla ispatlanmıştır. Ayrıca Birinci ve ikinci aşamadaki toplam kalma süresinin hypoüstel dağılıma uyduğu gösterilmiş ve momentleri elde edilmiştir.

**Key Words:** Jackson Modeli, performans ölçütleri, optimizasyon, hypoüstel dağılım, momentler.

#### Finding Moments at an Open Jackson Model

**Abstract:** The model consists of  $r$ -stage  $M/M/C$  type of queuing systems. The customers arrive to system according to Poisson distribution with parameter  $\lambda$ . There are arbitrary numbers of waiting at every waiting room,  $(\eta_i)$  is the service time at every channel of  $i$ -th system with exponential distribution having parameters  $\mu_i$  ( $i = 1, 2, \dots, c$ ) are independent. Every customer has service at 1-st system and then at 2-nd system and so on finally at  $r$ -th system and leaves the main system. In this paper system is analyzed, measures of performances and their optimal values are found supposing  $r = 2$  and  $c = 1$ . In addition the independences of number of customers and independences of sojourn times at every stage are proved with the help of two given theorems. Also it is shown that the total time in system at stage 1 and at stage 2 is according to hyper-exponential distribution and moments of this distribution are obtained.

**Keywords :** Jackson model, measures of performances, optimization, hyper-exponential distribution, moments.

#### KAYNAKLAR:

- [1] Bhat ,U. N., 2008, An Introduction to Queuing Theory, Boston, 2008.
- [2] Harrison, J., M., and Ngyen, V., 1992, The QNET method for two-moment analysis of open queueing Networks, Queueing sistem , 6,1-32.
- [3] Holdijk, A., and Spieksma, F., 1992, On Ergodicity And Recurrence Properties Of a Markov Chain With an Application To an Open Jackson Network, Adv. Appl. Prob. 24,343-376.
- [4] Jackson, J.R., 1957, Network of Waiting Lines, Operation Research, 5,518-524.
- [5] Stewart, W. J., 2009, Probability, Markov Chains, Queues, Simulation, New Jersey.

### GEÇİKMELİ VE PARETO MÜDAHALELİ ÖDÜLLÜ YENİLEME SÜRECİNİN SINIR FONKSİYONELLERİNİN MOMENTLERİ İÇİN ASİMTOTİK AÇILIMLAR

İhsan ÜNVER

KTÜ, Fen Fakültesi, Matematik Bölümü, 61080, Trabzon [ihsanunver@ktu.edu.tr](mailto:ihsanunver@ktu.edu.tr)

Zafer KÜÇÜK

KTÜ, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, 61080 Trabzon, [zkucuk@ktu.edu.tr](mailto:zkucuk@ktu.edu.tr)

Şenol DEMİR

KTÜ, Fen Fakültesi, Matematik Bölümü, 61080, Trabzon [demirsennol@gmail.com](mailto:demirsennol@gmail.com)

Tülay KESEMEN

KTÜ, Fen Fakültesi, Matematik Bölümü, 61080, Trabzon [tkesemen@gmail.com](mailto:tkesemen@gmail.com)

#### 1.Özet

Bu çalışmada, gecikmeli ödüllü yenileme süreci olarak adlandırılan yarı-Markov bir model  $(X(t))$  ele alınmış ve bu modeli ifade eden stokastik süreç matematiksel olarak inşa edilmiştir. Bazı zayıf koşullar altında bu sürecin ergodik olduğu ispatlanmıştır. Bunlardan yararlanarak  $\{\zeta_n\}, n \geq 0$  rasgele değişkenler dizisinin dağılımı  $(\alpha, \lambda)$  parametrelili pareto dağılımına sahip olması halinde sürecin bazı sınır fonksiyonlarının ilk dört momentin asimptotik açılımları  $E(\zeta_n) \rightarrow \infty$  iken elde edilmiştir. Sonunda, Monte Carlo simülasyon metoduyla elde edilen bu asimptotik sonuçların doğruluğu test edilmiştir.

#### 2.Giriş

Kuyruk teorisi, stokastik finans, güvenilirlik teorisi, matematiksel sigorta, stok kontrol teorisi, fizik ve biyoloji gibi alanlarda birçok ilginç problemler yenileme ve ödüllü yenileme süreçleri ile ifade edilmektedir. Sınır fonksiyonelleri, süreçlerin karakteristiklerinin incelemesinde büyük önem taşımaktadır. Özellikle süreçlerin ergodik dağılımının incelenmesi ve bazı olasılık karakteristiklerinin bilinmesi için gereklidir. Bundan dolayı, bu çalışmada sınır fonksiyonellerinin incelenmesine geniş ölçüde yer verilmiştir.

Pareto dağılımı, iktisat, sosyal bilimler, fen ve geofizik olmak üzere geniş bir alanda uygulanabilmektedir. Bu alanların dışında matematiksel sigorta teorisinin birçok problemlerinde çözümünde pareto dağılımı aktif bir şekilde kullanılmaktadır. Bu çalışmada mevcut çalışmalardan farklı olarak gecikmeli ödüllü yenileme süreci ele alınmıştır. Müdahaleli ifade eden  $\zeta_n$  rasgele değişkeni pareto dağılımına sahip olması durumunda sürecin bazı sınır fonksiyonellerinin ilk dört momentini için  $E(\zeta_n) \rightarrow \infty$  iken asimptotik açılımlar elde edilmiştir. Sonra da, elde edilen asimptotik açılımların yardımcıyla hesaplanan moment değerlerinin, kesin değerlere ne kadar yakın olduğunu göstermek için özel bir durum ele alınmış ve bu durum için Monte Carlo simülasyon yöntemi uygulanarak asimptotik açılımların doğruluğu test edilmiştir.

**Anahtar Kelimeler:** Ödüllü yenileme süreci, sınır fonksiyoneli, ergodiklik, asimptotik açılım, simülasyon, pareto dağılımı.

İSTATİSTİKSEL DAĞILIM PARAMETRELERİ TAHMİNLERİNİN EFFICIENT GLOBAL  
OPTIMIZATION (EGO) İLE ELDE EDİLMESİ

Ali MERT\*, Olgun AYDIN

Ege Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 35040, Bornova, İzmir, TÜRKİYE.  
ali.mert@ege.edu.tr

**1. Bir Metamodelleme Yöntemi: Kriging**

Metamodel, girdi/çıkıtı arasında mevcut olduğu varsayılan bir fonksiyonel yapının simülasyon çalışması ardından tahminlendiği istatistiksel yaklaşımdır. Genellikle, ilgilenilen sistemin simülasyon modeli çalışmasının dahi maliyetli ve/veya çok fazla zaman alan olduğu durumlarda kullanılmaktadır.

Kriging yöntemi, bir maden mühendisi olan D.G. Krige tarafından geliştirilmiştir ve sıkça kullanılan metamodelleme tekniklerinden biridir. Kriging yöntemi, rasgele bir sürecin bilinmeyen çıktı değerlerini tahminleyebilmektedir. Elde edilen tahmin, rasgele süreçten gözlemlenen çıktı değerlerinin ağırlıklı lineer kombinasyonudur. Bu ağırlıklar tahmin edilmek istenen noktadaki girdi ile önceden çıktı değerleri gözlemlenen girdi noktaları arasındaki uzaklığa bağlıdır.

Çıktı değerlerinin beklenen değeri sabittir. Ağırlıklar tahmin varyansını minimum yapmak üzere elde edilir. Hali hazırda gözlemlenen bir çıktı değeri tahminlenmek istendiğinde bu değer gözlemlenmiş olan tahmin değeri elde edilecektir. [3]

Kriging yöntemi ile herhangi bir  $x^*$  noktasındaki çıktı değeri tahmin etmek istendiğinde (1)'deki denklemden yararlanılmaktadır.

$$\hat{y}(x^*) = f' \hat{\beta} + r' R^{-1} (Y - F \hat{\beta}) \quad (1)$$

$Y$  : gözlemlenmiş çıktı değerleri

$\hat{y}$  :  $x^*$  noktasındaki çıktının tahmin değeri

$f$  : girdi ve çıktı değişkenleri arasındaki ilişkinin formunu gösteren vektör

$\hat{\beta}$  :  $f$  vektöründeki değerlerin katsayıları

$r$  : tahminin gerçekleştirileceği  $x^*$  noktasındaki girdi değeri ile önceden gözlemlenmiş girdi değerleri arasındaki korelasyon vektörü

$R$  : önceden gözlemlenen girdi değerleri arasındaki korelasyon matrisi

$F$  : önceden gözlemlenen girdi değerlerinin  $f$  vektörüne uygun hallerinden oluşan matris

**2. EFFICIENT GLOBAL OPTIMIZATION (EGO)**

EGO yöntemi optimum değerinin elde edilmesinde bazı zorluklar yaşanan fonksiyonların optimizasyonunda kullanılan Kriging metamodeline dayalı bir yöntemdir. EGO yöntemine göre optimizasyon esnasında fonksiyonun kendisi yerine Kriging yöntemi yardımı ile elde edilen Expected Improvement Function (Beklenen Gelişme Fonksiyonu) (2) kullanılır. Beklenen Gelişme Fonksiyonu (2) maksimize edilir ve bu işlem önceden belirlenen durdurma kuralı sağlanana kadar tekrar edilir [1-2].

**KAYNAKLAR**

[1] Feller W.(1971), Introduction to Probability Theory and Its Applications II, John Wiley.

[2] Khaniyev T.A., Aliyev R.T., Küçük Z., Bekar N.O.(2008),On the Distributions of a Renewal Reward Process and Its Additive Functional, Mathematical and Computational Applications, 13, 1, 41-50.

[3] Aliyev R., Bekar N.O., Khaniyev T., Ünver İ. (2010),Asymptotic Expansions for the Moments of the Boundary Functionals of the Renewal Reward Process With a Discrete Interference of Chance,Mathematical and Computational Applications, 15, 1, 117-126.

**ABSTRACT**

**ASYMPTOTIC RESULTS FOR THE MOMENTS OF THE BOUNDARY FUNCTIONALS OF THE RENEWAL REWARD PROCESS WITH DELAY AND PARETO DISTRIBUTED INTERFERENCE OF CHANCE**

In this paper, semi-Markov, a model called as the renewal reward processes  $X(t)$  is considered and the stochastic process expressed by this model is constructed mathematically. Under some weak assumptions the ergodicity of this process is discussed. Using these results the asymptotic expansions for the first four moments of the boundary functionals are established as  $E(\zeta_n) \rightarrow \infty$  when random variable  $\zeta_n$  has a pareto distribution with parameters  $(\alpha, \lambda)$ . Finally, the accuracy of the approximation formula is tested by the Monte Carlo simulation method.

**Key Words:** Renewal reward process, boundary functional, ergodicity, asymptotic expansion, simulation, pareto distribution.

$$EI = \left( f_{\min} - \hat{y} \right) \Phi \left( \frac{f_{\min} - \hat{y}}{s} \right) + s \phi \left( \frac{f_{\min} - \hat{y}}{s} \right) \quad (2)$$

**3. İSTATİSTİKSEL DAĞILIMLARIN PARAMETRE TAHMİNİNDE EGO YÖNTEMİ KULLANIMI**  
İstatistiksel dağılımların parametre tahmini istatistik literatüründe önemli bir yer tutmaktadır. Bu problemi çözmek için En Küçük Kareler, En Çok Olabilirlik ve Moment tahminçileri gibi yöntemler kullanmak mümkündür. Özellikle ilk iki yöntem kullanılırken esasen bir optimizasyon problemi çözmek gerekmektedir. Bu problemleri bazı dağılımlar için parametrik olarak çözmek mümkün değildir. Bu gibi durumlarda araştırmacılar daha başka yöntemlere ihtiyaç duymaktadırlar. EGO yöntemi de bu tip çözümü zor olan optimizasyon problemlerini çözmekte kullanılmaktadır.

Bu çalışmada çarpık bir dağılım olan Genelleştirilmiş Gamma dağılımı (GGD) ve simetrik bir dağılım olan Normal dağılımın parametreleri En Çok Olabilirlik yöntemi ile elde edilen optimizasyon problemlerinin EGO ve Genetik Algoritma (GA) ile çözümü yoluyla tahminlenmeye çalışılmıştır. Bahsi geçen dağılımlardan farklı parametre değerlerinde ve farklı örneklem ölçümlerinde veri setleri türetilmiştir. Sonrasında bu veri setleri için parametre tahminleri gerçekleştirilmiştir.

GGD ve Normal Dağılım için sonuçlardan bir kısmı Çizelge 1'de gösterilmiştir. Normal Dağılımın her iki parametresi için de EGO ile hem küçük varyanslı hem de yanlılığı küçük tahminler elde edilmiştir. GGD için ise EGO l ve a parametreleri tahmininde her iki kritere göre daha iyi tahminler elde etmiştir. GGD 'nın c parametresi için ise GA daha iyi sonuçlar vermektedir.

**Çizelge 1. 30x255' lik veri seti için l=1, c=2, a=3 parametrelerine sahip GGD ve  $\mu = 10, \sigma = 9$  parametrelerine sahip Normal Dağılım için sonuçlar**

		GGD			Normal Dağılım	
		l	c	a	$\mu$	$\sigma$
GA	ortalama	1,559272	1,802621	2,323802	10,048657	8,721235
	varyans	0,4437419	0,3576834	0,6717297	2,64021273	1,140325864
	n*varyans	13,312256	10,730503	20,151891	79,2063819	34,20977591
	yanlılık miktarı	0,559272	0,197379	0,676198	0,048657	0,278765
EGO	ortalama	0,788	2,5564	3,4464	10,028473	8,748784
	varyans	0,1675	0,223087	0,174642	2,53942873	1,139819903
	n*varyans	5,025	6,692622	5,23925	76,1828618	34,1945971
	yanlılık miktarı	0,212	0,5564	0,4464	0,028473	0,251216

#### KAYNAKLAR

[1] D. R. Jones, M. Schonlau and W. J. Welch(1998), Efficient Global Optimization of Expensive Black-Box Functions, Journal of Global Optimization.

[2] M. Schonlau, (1997), Computer Experiments and Global Optimizaton PhD. Thesis. University of Waterloo,Canada.

[3] W. van Beers (2005) Kriging Metamodeling For Simulation PhD. Thesis, Tilburg University.

#### OBTAINING PARAMETERS OF STATISTICAL DISTRIBUTIONS WITH EGO

There are various approaches to obtain parameters of a statistical distribution such as Moment, Mean Square Error and Maximum Likelihood (ML) Estimations. The problem is actually an optimization problem and becomes problematic for some statistical distributions. In this study, we expressed ML functions of some statistical distributions and optimized these functions using Efficient Global Optimization (EGO) method which is based on Kriging Metamodeling technique. We generated data sets from different distributions in different sizes and we estimated parameter values of mentioned statistical distributions of data sets, by employing EGO. We investigated performance of EGO for various distributions and sample sizes.

**Key Words:** Parameter estimation, Generalized Gamma Distribution, Normal Distribution, Maximum Likelihood, Efficient Global Optimization.



## ÇOKTAN SEÇMELİ TESTLERDE KAYIP VERİ MEKANİZMASININ İNCELENMESİ: ÖBBS ÖRNEĞİ

Ergül DEMİR  
Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü  
erguldemir@ankara.edu.tr

### Giriş

Kayıp veri mekanizması, kayıp verilerin ve kayıp veri içeren değişkenlerin karakteristiğini belirlemede, bu belirlemeye bağlı olarak kullanılacak kayıp veri yöntem ve analizlerinin tercihinde anahtar rol oynamaktadır. Kayıp veri mekanizmaları üzerine bilinen ilk kavramlaşmalar, Rubin[4] tarafından yapılmıştır. Rubin, kayıp verilerin oluşmasına yol açan süreçleri, kayıp verinin 'seçkisiz kayıp (missing at random)' ve gözlenen verinin 'seçkisiz gözlenen (observed at random)' olup olmamasına bağlı olarak tanımlanmaktadır. Allison[1] ve Enders[3], Rubin'in oldukça teknik yaklaşımlarından hareketle kayıp veri mekanizmalarını daha genel düzeyde üç varsayımla açıklamaktadır. Buna göre bir veri setinde kayıp verilerin oluşmasını açıklayan mekanizma; (i) Tam Seçkisiz Kayıp (Missing Completely at Random), (ii) Seçkisiz Kayıp (Missing at Random) ve (iii) Seçkisiz Olmayan Kayıp (Not Missing at Random) olarak tanımlanan varsayımlarla açıklanabilmektedir. Bu varsayımlar doğrultusunda kayıp veri mekanizmasının açıklanması, pratikte bir veri setinde yer alan kayıp verilerin 'ihmal edilebilir (ignorable)' olup olmadığı kararının verilebilmesini sağlamaktadır. Tam seçkisiz bir biçimde oluşması durumunda kayıp verilerin ihmal edilmesi, seçkisiz oluşması durumunda ise kayıp veriler yerine olasılıklı değer ataması yapılması mümkündür. Seçkisizliğin sağlanmaması durumunda ise istatistiksel kestirimin yanlışlık içerme olasılığı yüksektir. Kayıp veri mekanizmasının belirlenmesindeki temel sorun, gerçek değerlerin bilinmiyor olmasıdır. Bu nedenle kayıp verilerle ilişkili potansiyel bir yanlışlığın incelenmesi, genellikle, bazı temel değişkenler açısından yanıtlayanlar ve yanıtlamayanlar arasında manidar bir fark bulunup bulunmadığının test edilmesi yaklaşımına dayanmaktadır[2, 5].

Bu çalışmanın amacı; çoktan seçmeli testlerde, kayıp veri oluşma olasılığı ile ilişkili olabilecek değişkenlerin belirlenmesi ve bu değişkenler düzeyinde kayıp veri mekanizmasının incelenmesidir.

### Yöntem

Bu çalışma temel araştırma türlerinden tarama modelinde bir araştırma olarak tasarlanmıştır. Kayıp veri mekanizmasına yönelik inceleme ve analizler, Millî Eğitim Bakanlığı (MEB) tarafından 2009 yılında uygulanan 'Öğrenci Başarıları Belirleme Sınavı (ÖBBS)'nin 9A kitapçığı verileri üzerinde gerçekleştirilmiştir. ÖBBS, Türkçe, Matematik, Fen Bilimleri, Sosyal Bilimler ve İngilizce alanlarında 5 alt testten oluşmaktadır. Türkiye'de MEB tarafından ortalama 5 yılda bir, tabakalı seçkisiz örnekleme ile belirlenen bir örneklem üzerinde uygulanmaktadır. ÖBBS 2009 uygulamasında her bir alt testte 15'er soru içeren 9A kitapçığı, Türkiye'nin 30 farklı ilinde toplam 16670 9. sınıf öğrencisi tarafından yanıtlanmıştır. Ham veri seti üzerinde doğru, yanlış ve kayıp veri sayılarının yanı sıra kayıp veri oluşma olasılığı ile ilişkili olabilecek cinsiyet, bölge, okul türü, anne ve baba eğitim düzeyi, kendini başarılı görme düzeyi, evde ders çalışma süresi, ders dışı destek alma süresi gibi 11 bağımsız değişken tanımlanmıştır. Bu değişkenler ile kayıp veri oluşma olasılığı arasındaki ilişkiler, her bir alt testte ayrı ayrı olmak üzere CHAID analizi ile analiz edilmiştir.

### Bulgular

Her bir alt testte, tanımlanan her bir değişken düzeyinde, kayıp veri miktarları açısından manidar fark bulunduğu görülmüştür. Kayıp veri miktarı açısından en yüksek manidar fark, bölge değişkenine göre ortaya çıkmıştır. Matematik ve İngilizce alt testleri dışındaki testlerde ikinci sırada, haftalık ders dışı destek alma süresi gelmektedir. Alt testlere göre farklılık göstermekle birlikte 2, 3 ve 4. sıralarda okul türü ve cinsiyet değişkenleri gelmektedir. Anne ve baba eğitim düzeylerine göre kayıp veri miktarları, en düşük düzeyde manidar fark oluşturmaktadır.

### Sonuç

Bu çalışmada tanımlanan 11 bağımsız değişkenin tamamının, kayıp veri oluşma olasılığı ile ilişkili olabileceği belirlenmiştir. Kayıp veri oluşma olasılığı ile en yüksek ilişkiyi gösteren değişkenler, başta bölge değişkeni olmak

## BİLDİRİ OTURUMLARI 2

### SESSION 2

#### Uygulamalı İstatistik 2

üzere, ders dışı destek alma durumu, okul türü, cinsiyet ve kendini başarılı görme düzeyidir. Kayıp veri oluşma olasılığı ile en düşük düzeyde ilişkili değişkenler ise anne ve baba eğitim düzeyleridir. Söz konusu ilişkiler, veri setindeki kayıpların TSK varsayımını sağlamadığına yönelik kanıt oluşturmaktadır. Bu değişkenler düzeyinde bir kayıp veri yanlılığının bulunma olasılığı, dolayısıyla kayıp verilerin ihmâl edilmesi ve elde edilecek istatistiksel kestirimlerin yanlılık içerme olasılığı yüksektir.

#### KAYNAKLAR

- [1] Allison, P.D. (2002). Missing Data. California: Sage Publication, Inc.
- [2] Culbertson, M.J. (2011). Is It Wrong? Handling Missing Responses in IRT. Annual Meeting of the National Council on Measurement in Education, April 2011.
- [3] Enders, C.K. (2010). Applied Missing Data Analysis. New York: The Guilford Press.
- [4] Rubin, D.B. (1976). Inference and Missing Data. Biometrika, s.63, n.3, sy.581-592.
- [5] Whipple, T.W. ve Muffo, J.A. (1982). Adjusting for Nonresponse Bias: The Case of an Alumni Survey. 22nd Annual Forum of the Association for Institutional Research, May, 16-19. Denver: Cleveland State University

#### ABSTRACT

##### MISSING DATA MECHANISM IN MULTIPLE CHOICE TESTS: THE CASE OF ÖBBS

In this study, it is aimed at define the variables related with presence of missing data and research the missing data mechanism with these variables. Analyses were conducted on ÖBBS-2009 data from 16670 students. ÖBBS (student achievement test) has been executed by Turkish Ministry of National Education every 4 or 5 years. 11 independent variables like region and gender and school type etc. as well as the amount of missing data were defined on the raw data set. In the analysis, CHAID were used. As a result, there are significant differences between amount of missing data and every independent variables level. Region is the major variable indicating a highly significant relationship with the possibility of missing data. Mother and father education level indicate a lowest relationship. All these relations show that missing data are not provide the MCAR mechanism. The missing data bias is likely to occur for statistical analysis.

**Key Words:** missing data, missing data mechanism, MCAR, MAR, NMAR

#### MEDYAN PARLATMA TEKNİĞİ VE UYGULAMASI

Gökçen EFENDİOĞLU, Nuri ÇELİK

Bartın Üniversitesi Fen Fakültesi, Bartın Üniversitesi Fen Fakültesi İstatistik Bölümü Merkez/Bartın, İstatistik Bölümü Merkez/Bartın

[gefendioglu@bartin.edu.tr](mailto:gefendioglu@bartin.edu.tr)

[ncelik@bartin.edu.tr](mailto:ncelik@bartin.edu.tr)

#### 1. Medyan Parlatma Tekniği

Olumsallık tabloları, kategorik verilerden meydana gelir ve bu tablolara dayalı temel soru satır ve sütunlarda anlamlı bir fark olup olmadığıdır.

Medyan Parlatma Tekniği basit toplanır modele dayalı olumsallık tablolarının analizinde kullanılan basit fakat dirençli bir veri çözümleme tekniğidir. Medyan Parlatma uygulanmadan önce daha verimli sonuçlar elde etmek için verilere dönüştürme işlemi uygulanır. Bu dönüştürme işlemi esas modeli çarpımsal modele dönüştüren bir algoritmadır.

İki yönlü tablo değişkenler birbirinden bağımsız olarak değişmektedir ve bağımlı değişken değişkenlerin her kombinasyonu için ayrı gözlemlenir.

Tablo 1: İki Yönlü Tablo

	j	
i	1	..... j
1	$x_{11}$	..... $x_{1j}$
.	.	.
.	.	.
.	.	.
I	$x_{I1}$	..... $x_{Ij}$

Sözü edilen toplamsal Medyan Parlatma modeliyle olumsallık tablosundaki her  $x_{ij}$  aşağıdaki formülle yazılabilir;

$$x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad (1)$$

Burada  $\mu$  genel değer,  $\alpha_i$  satır etkisi,  $\beta_j$  sütun etkisi ve  $\varepsilon_{ij}$  hata terimidir.

Medyan Parlatmanın sonucu olumsallık tablosunu daha iyi anlamamıza yardımcı olur. İdeal durumda artıkların sıfır veya sıfıra yakın olması gerekir.

#### 2. Dönüştürülmüş Veride Medyan Parlatma

Veri dönüştürülmesi veri ön işlemenin sık kullanılan bir adımdır. Analizin diğer adımlarından önce dönüşüm uygulamanın; normalleştirme, verinin belirli dağılım özelliklerine ulaşmak veya analizin sonraki adımlarında avantaj sağlamak gibi çeşitli sebepleri vardır.

Logaritma parametrik dönüşümlerin özel bir örneğidir ve güç dönüşümü olarak adlandırılır. Güç dönüşümü aşağıdaki gibi tanımlanır;

$$t_{\lambda}(x) = \begin{cases} (x^{\lambda} - 1) / \lambda & \lambda \neq 0 \\ \ln(\lambda) & \lambda = 0 \end{cases}$$

Dönüştürülecek x değerlerinin pozitif olduğu kabul edilir. Eğer veride negatif x değerleri varsa onları pozitif yapacak sabit bir değer bu negatif x değerlerine eklenir.

Bununla beraber, çalışmada yapılacak uygulamada TÜİK ten elde edilen gerçek veri setleri kullanılarak Medyan Parlatma Tekniğinin avantajları vurgulanacaktır.

#### KAYNAKLAR

- [1] Klawonn, F., Crull, K., Kukita, A., Pessler, F.: Median polish with power transformations as an alternative for the analysis of contingency tables with patient data. In He, J., Liu, X., Krupinski, E., Xu, G., eds.: Health Information Science, Berlin, Springer (2012) 25 {35
- [2] Berthold, M., Borgelt, C., HÄoppner, F., Klawonn, F.: Guide to Intelligent Data Analysis: How to Intelligently Make Sense of Real Data. Springer, London (2010)
- [3] Tukey, J. W. (1977). Exploratory Data Analysis, Reading Massachusetts: Addison-Wesley,
- [4] Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). Understanding robust and exploratory data analysis. New York: Wiley..
- [5] Fink, A. M. (1988), "How to Polish off Median Polish," SIAM Journal on Scientific and Statistical Computing, 9, 932-940.

#### ABSTRACT

Median Polish is a way of analysing contingency tables based on a simple additive model. It is a basic, however a robust method of analysing. Most of the time, the data needs to be applied a power transformation, before using the MP algorithm to get better results. A commonly used power transformation is the logarithm which essentially changes the underlying model to a multiplicative model. In this work, Median Polish method is applied by using a real data set.

**Key words:** Median Polish Method, contingency table, power transformation

#### VARYANS ANALİZİ MODELLERİNDE BAĞIMSIZLIK VARSAYIMIN BOZULMASININ ETKİSİ

Jale BALİBEYOĞLU

Gazi Üniversitesi Fen Fakültesi İstatistik Bölümü 06500Ankara Türkiye

[balibey@gazi.edu.tr](mailto:balibey@gazi.edu.tr)

#### 1.Giriş

Birçok istatistik çözümlemesinde modelde yer alan rassal hata teriminin normal dağılıma sahip ve birbirinden bağımsız oldukları varsayılmaktadır. Buna bağlı olarak yokluk hipotezini test etmek üzere test istatistikleri oluşturulur. Şayet normallik varsayımından şüpheye düşülürse veya normallik varsayımı gerçekleşmez ise parametrik olmayan istatistik tekniklerinden yararlanır.

Scarono, S ve Devenport J.M tarafından yapılan çalışmada, ikinci varsayım olan bağımsızlık varsayımının gerçekleşmemesinin etkisi tek faktör varyans analizinde incelenmiştir. Bu çalışmada, gözlem sayısından ve birimler arası ilişkinin miktarından test istatistiği olan F'nin ve buna bağlı olarak I tip hatanın değiştiği gösterilmiştir. Balibeyoğlu J.'nin çalışmasında regresyon analizi ve tekrarlı ölçmelerde bağımsızlık varsayımının etkisi incelenmiştir. Regresyon analizi için varsayımdaki bozulmanın F istatistiğinde bir değişikliğe yol açmadığı görülmüştür.

#### 2.Teorik Gösterim

Genel lineer model

$$Y = x\beta + \varepsilon$$

$$\varepsilon : N(0, I, \sigma^2)$$

biçimindedir.

Bir varyans analizi probleminde;

$I = P_1 + P_2 + P_3 + \dots + P_k$ ,  $P_i P_j = 0$  ( $i \neq j$ ) koşulu altında ve  $P_i$  izdüşüm matrisi olmak üzere, herhangi bir işleme ait kareler toplamı;

$$Y' P_i Y = K T$$

biçimde yazılabilir ve,

$$Y' P_i Y / \sigma^2 : X^2 r(P_i) \text{ dir.}$$

Aynı biçimde F istatistiğinin,

$$F = \frac{Y' P_i Y / P(P_i)}{Y' P_j Y / P(P_j)} \quad (1)$$

olduğu bilinmektedir.

$$S = \{x_1, x_2, \dots, x_k\}$$

$$Cov(x_i, x_j) = \sigma^2 \quad i = j \in N$$

$$Cov(x_i, x_j) = \rho\sigma \quad i \neq j \in N$$

olması halinde S kümesine sınıf içi korelasyonlu değişken seti denir. Birimler arasında bağımlılık var ise ;

$$V = C_1 P_1 + C_2 P_2 \dots + C_k P_k$$

$$C_i > 0$$

$$C_i = 0$$

olması halinde I=V olmaktadır.

Pi ise izdüşüm matrisi olmak üzere, (1) eşitliğini tekrar yazarsak;

$$F = \frac{C_j}{C_i} \left[ \frac{Y' P_i Y / r(P_i)}{Y' P_j Y / (P_j)} \right]$$

biçiminde olacaktır.

Görüldüğü üzere F istatistiği  $\frac{C_j}{C_i}$  gibi bir değişime uğramaktadır.

Bu çalışmada iki faktör varyans analizinde, bağımsızlık varsayımının bozulmasının etkisi, grup içi korelasyonun  $\rho_1$ , gruplar arası korelasyonun  $\rho_2$  olması durumu için incelenecektir.

#### KAYNAKLAR

- [1] Balibeyoğlu J. (1994), Regresyon analizi ve tekrarlı ölçmelerde bağımsızlık varsayımındaki bozulmanın etkisi, Gazi Üniversitesi Fen – Edebiyat Fakültesi Fen Bilimleri Dergisi C.4, 97–109.
- [2] Graybil F. (1976), An Theory and Application of The Lineer Model, Wadsworth, Publishing Company.
- [3] Kenny, D., A., Judd, C., M. (1986), Consequences of violating the independence assumption in analysis variance, Psychological Bulletin, 99(3) 422 – 431.
- [4] Winer, B. J. (1971) Statistical Principles in Experimental Design Mc Grow-Hill.
- [5] Scariano, S., Davenport, J.(1987), The effect of violation of indepenence assumption in the oneway anova, American Statistician, 41, 123-139.

#### ABSTRACT

##### THE EFFECT OF VIOLATION INDEPENDENCE ASSUMPTION IN THE ANALYSIS OF VARIANCE MODELS

In this focuses on relationship between true types I and II error probabilities and the effects of departures from independence assumption on hypothesis testing in anova.

**Key Words:** Type I Error, Correlated data, Scaled F Variable

#### BİR BOYUTLU KESME VE ANA MALZEME SEÇİMİ PROBLEMİ İÇİN İKİ AMAÇLI TAM SAYILI PROGRAMLAMA MODELİ

Nergiz KASIMBEYLİ  
Endüstri Mühendisliği Bölümü, Anadolu Üniversitesi, İki Eylül Kampüsü, 26555 Eskişehir  
nkasimbeyli@anadolu.edu.tr

#### Özet

Çalışmada bir boyutlu kesme ve ana malzeme seçimi problemi incelenmektedir. Bu problem, belli uzunluklu siparişlerin karşılanabilmesi için farklı boyutlara sahip ana malzemeler içinden toplam fire miktarını minimum yapacak optimal kesme planının belirlenmesi ve bu optimal plana karşı gelen ana malzemenin seçilerek stoklanması problemi kapsamaktadır. Benzer problemlerin modellenmesi ve çözülmesi sürecinde karşılaşılan en büyük zorluklardan biri, problemin matematiksel modelinin oluşturulmasında kesme planlarının ne şekilde oluşturulabileceğidir. Literatürde yapılmış olan araştırmaların çoğunda kesme planları matematiksel modelin parametre kümesi olarak kullanılmıştır. Fakat kesme planlarının tamamının bir parametre kümesi olarak oluşturulması oldukça zor, hatta çoğu zaman imkânsız bir problemdir, zira çok büyük olmayan problemler için bile kesme planı sayısının sonsuz denecek kadar büyük boyutlara ulaştığı biliniyor ki bu da problemin çözümü sırasında büyük güçlükler neden olmaktadır. Bu çalışmanın amacı, söz konusu problemin çözümü için kesme planlarını parametre kümesi olarak kullanmayan bir matematiksel model ve bu model için bir çözüm algoritmasının geliştirilmesidir. Çalışmada, ele alınan problem için iki amaçlı tam sayılı, kesme planlarını parametre olarak içermeyen bir matematiksel model geliştirilmiştir. Standart optimizasyon yazılımlarının modelin çözümünde zorlanabileceği göz önünde bulundurularak, bu model için özel bir çözüm algoritması da geliştirilmiştir. Geliştirilmiş olan matematiksel model ve çözüm algoritması test problemler üzerinde denenmiştir.

#### REFERENCES

- [1] Kasimbeyli, N., Saraç, T., Kasimbeyli, R., (2011), A two-objective mathematical model without cutting patterns for one-dimensional assortment problems, Journal of Computational and Applied Mathematics, Vol. 235 No 16, pp. 4663–4674.

#### ABSTRACT

##### A TWO-OBJECTIVE INTEGER PROGRAMMING MATHEMATICAL MODEL FOR ONE-DIMENSIONAL CUTTING AND ASSORTMENT PROBLEMS

This paper considers the one-dimensional cutting and assortment problem which includes the determination of the number of different sizes of standard lengths to be maintained as inventory and to be used to fulfill a set of customer orders. One of the main difficulties in formulating and solving this kind of problems is the use of cutting patterns in the mathematical model. Many mathematical programming approaches for solving the assortment problems assume the existence of the set of cutting patterns. The corresponding mathematical models use the cutting patterns as model parameters. Because of a huge number of cutting patterns to be obtained for such kind of models, this leads to computational difficulties in solving these problems. The purpose of this paper is therefore to develop a mathematical model without the use of cutting patterns as a parameter set. In this paper, a two objective integer programming model is developed for solving a one-dimensional cutting and assortment problem with two or more types of stock lengths. We suggest a special solution method. The mathematical model and the solution approach are demonstrated on test problems.

**Key Words:** Cutting Problem, One-Dimensional Assortment Problem, Multi-Objective Optimization.

## SIRALI LOJİSTİK REGRESYONDA ORANTISAL ORAN MODELİ İÇİN PARALEL DOĞRULAR VARSAYIMI

Erkan ARI\*

Cumhuriyet Üniversitesi Fen Fak. İstatistik Böl., 58140, Sivas, Türkiye, eari@cumhuriyet.edu.tr

Zeki YILDIZ

Eskişehir Osmangazi Üniversitesi Fen Fak. İstatistik Böl., 26480, Eskişehir, Türkiye, zyildiz@ogu.edu.tr

### Özet

Sıralı lojistik regresyon çözümlemesinde paralel doğrular varsayımının sağlanması önemlidir. Paralel doğrular varsayımı; sıralı lojistik regresyonda bağımsız değişkenler ile bağımlı değişken arasındaki ilişkinin bağımlı değişken kategorilerine göre değişiklik göstermediğini ifade eder. Bu varsayımın sağlanmaması durumunda Orantısal Oran modelinin kullanılması hatalı olmaktadır. Bu çalışmada, Orantısal Oran Modeli için paralel doğrular varsayımının sağlanmadığı sonucuna ulaşılmıştır.

### 1. Orantısal Oran Model (Proportional Odds Model) (PO)

Sıralı lojistik regresyonda bağımlı değişkenin sıralı olduğu ve kategoriler arasında paralellik varsayımı sağlandığı durumda çalışmalarda yaygın olarak kullanılan sıralı lojistik regresyon modeli "Orantısal Oran Model" dir (Brant, 1990; Bender and Grouven, 1998).

Orantısal Oran Model, sıralı lojistik regresyon için McCullagh (1980) tarafından tanımlanmıştır. Model yığılımlı olasılıkların dağılımına dayanır.

Orantısal Oran Model yığılımlı olasılıklar kullanılarak Eşitlik 1.'deki gibi elde edilir (Kleinbaum and Ananth, 1997).

$$P(Y \leq y_j | X) = \frac{\exp(\alpha_j - x' \beta)}{1 + \exp(\alpha_j - x' \beta)} \quad j = 1, 2, \dots, J - 1 \quad [1.]$$

Burada  $\alpha_j$  bilinmeyen parametrelerin kestiricisi ve  $J - 1$  tane kestiriciye karşılık gelen  $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_{j-1}$ ;  $\beta$ 'lar ise  $\beta = (\beta_1, \dots, \beta_k)'$   $x$ 'e karşılık gelen regresyon katsayılarının bir vektörüdür.

### 2. Paralel Doğrular Varsayımı

Sıralı lojistik regresyon modellerinde sıralı odds'a ait önemli bir varsayım bulunmaktadır. Bağımlı değişken açıklanırken, kategorilerine ait odds oranları için ifade edilen formüllerde kategoriler arasında parametre açısından herhangi bir fark yoktur. Kısacası, bağımsız değişkenler ile bağımlı değişken arasındaki ilişki bağımlı değişkenin kategorilerine göre değişmez, parametre tahminleri kesme noktalarına göre değişiklik göstermez. Sıralı lojistik regresyonda,  $J$  kategorili bağımlı değişkende,  $J - 1$  adet lojit karşılaştırması için varsayım sağlandığında  $\alpha_{j-1}$  adet kesme noktası ve yalnızca bir adet  $\beta$  parametresi bulunur. Ancak çok terimli lojistik regresyonda ise  $J - 1$  adet lojit karşılaştırması için  $\alpha_{j-1}$  adet kesme noktası bulunurken,  $J - 1$  adet de  $\beta$  parametresi vardır. İşte sıralı lojistik regresyon çok terimli lojistik regresyondan bu noktada ayrılır (Kleinbaum and Klein, 2010).

### 3. Paralel Doğrular Varsayımının Testi

Paralel doğrular varsayımının geçerliliğini kontrol etmek için temelde Olabilirlik Oran Testi, Wald Ki-Kare testi gibi testler kullanılmaktadır. Sıralı lojistik regresyonda bu testler ile

parametrelerin farklı kategorilerindeki eşitliği test edilerek paralel doğrular varsayımı sınımasna gidilir.

Orantısal Oran Modelin geçerliliği Olabilirlik Oran Testi ile incelendiğinde, model anlamlı bulunmuştur ( $p = 0,0157$ ). Orantısal Oran Model anlamlı olmasına rağmen bu modelin kullanılabilmesi için öncelikli olarak Paralel Doğrular Varsayımının test edilmesi gerekir. Paralellik varsayımı için parametrelerin farklı kategorilerdeki eşitliği test edilerek paralel doğrular varsayımının sağlanıp sağlanmadığı test edilir. Çözümleme için oluşturulan hipotez Eşitlik 2.'deki gibidir.

$$H_0 = \beta_{1k} = \beta_{2k} = \dots = \beta_{(j-1)k} \quad j = 1, 2, \dots, J - 1 \quad [2]$$

$k$ 'nci bağımsız değişkene ait katsayıların birbirine eşit olup olmadığı olabilirlik oran testi ile incelenmiş ve ( $\chi^2 = 20,83$ ;  $p = 0,0003$ ) değeri ile sıfır hipotezi reddedilmiştir ve paralel doğrular varsayımının ihlal edildiği sonucuna ulaşılmıştır. Brant'ın Wald testine göre de paralel doğrular varsayımı sağlanmamıştır. Bu test paralel doğrular varsayımının hangi değişken veya değişkenler tarafından bozulduğu noktasında bilgi verir. Paralellik varsayımının sağlanmamasında bir değişkenin varsayımı sağlamaması yeterlidir.

### 4. KAYNAKLAR

[1]Bender, R. and Grouven, U., 1998, Using Binary Logistic Regression Models for Ordinal Data with Non-Proportional Odds, J. Clin Epidemiol Vol. 51, No.10, 809-816 p.

[2]Brant, R., 1990, Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression, Biometrics, Vol. 46, No. 4, 1171-1178 p.

[3]Kleinbaum, D.G. and Ananth, C.V., 1997, Regression Models for Ordinal Responses: Review of Methods and Applications, International Journal of Epidemiology, Vol 26, No.6, 1323-1333 p.

[4]Kleinbaum, D.G. and Klein M., 2010, Logistic Regression. A Self- Learning Text, Third Edition, Springer, 590 p.

[5]McCullagh, P., 1980, Regression Models for Ordinal Data, Journal of the Royal Statistical Society, Series B (Methodological), 42, No. 2, 109-142 p.

### ABSTRACT

#### The Parallel Line Assumption for Proportional Odds Model in Ordinal Logistic Regression

It is significant to ensure parallel line assumption in ordinal logistic regression analysis. Parallel line assumption points out that the relation between independent and dependent variable does not differ in terms of dependent variable categories. In case this assumption is not ensured, the use of proportional odds model is a mistake. In the study, it is concluded that parallel line assumption for proportional odds model is not provided.

**Key Words:** Proportional Odds Model, Parallel Slope Assumption.

## ARALIK DEĞERLİ ZAMAN SERİLERİ ÖNGÖRÜ YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Arş. Gör. Ebrucan İSLAMOĞLU, Prof. Dr. Faruk ALPASLAN

Ondokuzmayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Atakum-Samsun

e-posta: [ebrucantiring@omu.edu.tr](mailto:ebrucantiring@omu.edu.tr)

### 1. Aralık Değerli Zaman Serileri

Aralık değerli zaman serileri literatürde yeni araştırılan bir konudur. Gözlemleri gün içinde değişen veriler aralık değerli zaman serileri olarak bilinir. Bu seriler açılış-kapanış veya en düşük - en yüksek değerlerden oluşmaktadır. Aralık değerli zaman serileri ekonomi, mühendislik, tıp, meteoroloji ve iktisat gibi çeşitli alanlarda görülebilir. Söz konusu seriler; piyasa verileri, fiyat endeksleri, para - banka verileri, bilanço verileri ve dış ticaret endeksleri gibi verilerdir.

Çalışmamızda zaman serilerinin çeşitli tahmin yöntemlerinin farklı kombinasyonlarını karşılaştırılan deneysel bir araştırma yapılmıştır. Kullanılan seriler T. C. Merkez Bankası internet sitesinden (Elektronik Veri Dağıtım Sistemi-EVDS) elde edilmiştir. Bu seriler Altın Borsası İşlemleri I - İstanbul (İş günü, TL/ABD Doları) en düşük (ABD/Ons) günlük verileri, Altın Borsası İşlemleri II - İstanbul (İş günü, TL/ABD Doları) en düşük (TL/Kg) günlük verileri ve Geçinme Endeksi (Ücretliler) (1995/100) en düşük günlük verilerinden oluşmaktadır. Bu seriler kullanılarak farklı yaklaşımlar oluşturulmuş ve modelleme teknikleri karşılaştırma için kullanılmıştır. En düşük ve en yüksek serileri çözümleyen Yaklaşım1, merkez ve açıklık serilerini çözümleyen Yaklaşım3 oluşturulmuştur. Karma Otoregresif Bütünleşik Hareketli Ortalama Süreci (Auto Regressive Integrated Moving Average - ARIMA), Yapay Sinir Ağları (Artificial Neural Networks - ANN), Holt Üstel Düzleştirme Yöntemi (Holt Exponential Smoothing Method) ve Melez Modeller (Hybrid Models) kullanılan modelleme teknikleridir. Öngörü sonuçlarının doğruluğunu ölçen performans ölçütleri Hata Kareleri Ortalamasının Karekökü (Root Of Mean Square Error-RMSE), Mutlak Hata Oranları Ortalaması (Mean Absolute Percentage Error- MAPE) ve Yön Doğruluğu (Direction Accuracy - DA) dur.

Çizelge1. Aralık değerli zaman serileri için elde edilen sonuçlar

		RMSE	MAPE	DA
ALTIN BORSASI İŞLEMLERİ-I	Yaklaşım1	ARIMA(0, 0, 1) = <b>93,47</b> <b>2004</b>	YSA (1-2-1) = <b>0,0006</b> <b>2004</b>	YSA (7-1-1) = <b>0,3878</b> <b>2001</b>
	Yaklaşım3	ARIMA(0, 0, 0) = <b>0,552</b> <b>2000</b>	YSA (10-5-1) = <b>0,0085</b> <b>2001</b>	MelezYSA = YSA (3-2-1) = <b>0,1837</b> <b>2000</b> <b>2000</b>
ALTIN BORSASI İŞLEMLERİ-II	Yaklaşım1	Holt Üstel Düzleştirme= <b>1032,2590</b> <b>2000</b>	YSA (3-2-2)= <b>0,0014</b> <b>2000</b>	MelezYSA= <b>0,5102</b> <b>2000</b>
	Yaklaşım3	ARIMA (1, 0, 0)= <b>20,9430</b> <b>2000</b>	YSA (5-6-1)= <b>0,0046</b> <b>2000</b>	YSA (10-2-1)= <b>0,2041</b> <b>2001</b>
GEÇİNME ENDEKSİ	Yaklaşım1	MelezYSA= <b>4,3817</b> <b>2000</b>	YSA (9-3-2)= <b>0,0000185</b> <b>2004</b>	YSA (2-2-1) = YSA (9-4-1) = YSA (1-6-1) = YSA (9-3-1) = <b>0,0204</b> <b>2001</b> <b>2004</b> <b>2002</b>
	Yaklaşım3	YSA (1-1-1)= MelezYSA= <b>1,2937</b> <b>2000</b> <b>2000</b>	YSA (1-1-1)= <b>0,0000</b> YSA(1-1-1)= <b>2000</b> <b>2001</b>	YSA (1-1-1)= MelezYSA= <b>0,000</b> <b>2000</b> <b>2000</b>

Çalışmamızda 03 Ocak 2000 ve 31 Aralık 2004 tarihleri arasındaki üç zaman serisi için öngörü yöntemi olan ARIMA, Holt Üstel Düzleştirme, Yapay Sinir Ağları ve Melez Modeller karşılaştırılmıştır. Altın fiyatları ve vatandaşların geçim gereksinimini karşılayabilmesi için yaptıkları harcamalar tahminlenmiştir. Elde ettiğimiz sonuçlara göre aralık değerli zaman serilerinin tahmininde yapay sinir ağı modelinin diğer zaman serisi tahmin

## BİLDİRİ OTURUMLARI 2

### SESSION 2

#### Zaman Serileri 2

modellerine göre daha üstün sonuçlar verdiği görülmüştür. Bu nedenle yapay sinir ağlarının öngöründe etkin bir yöntem olduğu söylenebilir.

#### KAYNAKLAR

- [1] Akaike H. (1974), A new look at the statistical model identification, IEEE Trans. Autom, Control 19716-723.
- [2] Beheshti M., Berrached A., De Korvin A., Hu C. and Sirisaengtaksin O., (1998), On interval weighted three-layer neural networks, in : SS'98; Proceedings of the 31st Annual Simulation Symposium, IEEE Computer Society, Washington, DC,USA.
- [3] Billard L. and Diday, E. (2000), Regression analysis for interval-valued data. Data Analysis, Classification and Related Methods, in: Proceedings of the Seventh Conference of the International Federation of Classification Societies(IFCS'00), Springer, Belgium.
- [4] Lima Neto E. A. , De Carvalho F. A. T. (2008), Centre and range method for fitting a linear regression model to symbolic interval data, Comput, Stat, Data Anal. 52, 1500-1515.
- [5] Maia A. L. S. and De Carvalho, F. A. T. (2010), Holt's Exponential Smoothing and Neural Network Models For Forecasting Interval-Valued Time Series. International Journal of Forecasting, 02, 012.

#### ABSTRACT

##### THE COMPARE OF INTERVAL-VALUED TIME SERIES FORECASTING METHODS

In this study, we aim to provide an introduction of interval-valued time series. Different approaches have been used to analyze interval-valued time series. We have introduced central tendency and dispersion measures. We study on Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Networks (ANN), Holt Exponential Smoothing Method and Hybrid Models. These are model estimation methodologies for the analysis of interval-valued time series. We used Root Of Mean Square (RMSE), Mean Absolute Percentage Error (MAPE) and Direction Accuracy (DA). The significant result of study is that the results of application demonstrated the superiority of the Artificial Neural Networks (ANN) in the interval-valued time series.

**Key words:** Interval-valued time series, Forecasting, Artificial Neural Networks, ARIMA, ANN, MAPE, DA, RMSE.

#### DEĞİŞTİRİLMİŞ ÜSSEL DÜZLEŞTİRME YÖNTEMİ

Güçkan YAPAR ,Hanife TAYLAN  
Dokuz Eylül Üniversitesi, Fen Fakültesi, İstatistik Bölümü  
e-mail: [guckan.yapar@deu.edu.tr](mailto:guckan.yapar@deu.edu.tr) ve [hanife.taylan@deu.edu.tr](mailto:hanife.taylan@deu.edu.tr)

#### Değiştirilmiş Üssel Düzleştirme Yöntemi

Bu çalışmada üssel düzleştirme yöntemine alternatif bir yöntem önerilecektir ve değiştirilmiş üssel düzleştirme yöntemi olarak adlandırılacaktır çünkü bu yöntem geleneksel üssel düzleştirme metodunda düzleştirme katsayısında ufak bir değişiklik yapılarak elde edilecektir. Bu yöntemde başlangıç değeri ve optimum düzleştirme katsayısı aynı anda elde edilecektir. Bu yöntemle geleneksel yöntemdeki aynı düzleştirme katsayısında başlangıç değerine ve uzak geçmişe daha az ağırlık verilerek buradan elde edilen fazlalık yakın geçmişteki verilere üssel olarak dağıtılacaktır. Önerilen bu yöntem geleneksel üssel düzleştirme yöntemiyle aynı düzleştirme katsayısında karşılaştırıldığında eğer gelecek yakın geçmişin bir tekrarı olacaksa çoğu zaman geleceğin tahmininde daha iyi sonuç verecektir.

Son yüzyıl da zaman serisi analizleri için çok sayıda yöntem önerilmiştir. Son 25 yılda zaman serisi analizlerine ilişkin çalışmaların değerlendirilmesi De Gooijer ve Hyndman (2005) tarafından yapılmıştır. Literatürde birçok zaman serisi analiz yöntemi geliştirilmiştir fakat en başarılı yöntemlerin temeli üssel düzleştirme yönteminden gelmektedir. Bu yöntemin ilk ve önemli çalışmaları Holt, 1957; Brown, 1959; Winters, 1960 tarafından verilmiştir. Üssel düzleştirme yönteminin zaman serisi analizlerinde en popüler yöntem olmasının birçok sebebi vardır bunlardan ilki çok basit ve uygulamasının kolay olması fakat bu yöntemin popülaritesinin esas nedeni geleceğin tahminlenmesinde diğer çok k yöntemler karşısında tartışılmaz üstünlüğüdür (Gardner, 1985; Makridakis and Hibon, 2000).

Üssel düzleştirme yönteminin temel mantığı geleceğin tahminlenmesinde yakın geçmişteki gözlemlere daha fazla ağırlık vererek üssel bir şekilde azalarak geçmişteki gözlemlere ağırlık verme prensibidir. Üssel düzleştirme yöntemi bir modeller ailesidir ve bu yöntem zaman serisinin üç tane bileşeninin olduğunu varsayar bunlar düzey, eğim ve mevsimsellik ve amaç bu üç bileşenin son halini bulmak ve bunları kullanarak geleceğin tahminlenmesidir. Üssel modellemeye beş farklı trend ve üç farklı mevsimsellik bileşeni ile bugün 15 farklı model oluşturulabilmektedir Gardner (1985).

Üssel düzleştirme yönteminin uygulamasında takip edilmesi gereken adımlar uygun modelin kararı daha sonra ise bu modelin bileşenlerine ilişkin düzleştirme katsayılarının belirlenmesi ve düzleştirme istatistiklerinin başlangıç değerlerinin belirlenmesi. Modelin belirlenmesi verinin görsel incelenmesi ile olabilmektedir fakat düzleştirme katsayıları ve başlangıç değerlerinin belirlenmesi için birçok yöntem mevcuttur, Chatfield and Yar (1988). Düzleştirme katsayıları ve başlangıç değerleri en çok olabilirlik veya hata karalar toplamını minimize etme yöntemiyle elde edilebilir. Başlangıç değerlerinin belirlenmesinde uygulanan temel yöntemler ilk gözlemin, ilk birkaç gözlemin veya tüm gözlemlerin ortalamasını almak ve daha sonra bu başlangıç değerlerine göre optimum düzleştirme katsayısının belirlenmesidir. Bu yöntemler pratikte yaygın bir şekilde kullanılmaktadır fakat düzleştirme katsayısı küçük ve gözlem sayısı az olduğunda yanlış bir başlangıç değeri tahminlenmenin sonucunu olumsuz yönde etkilemektedir. Üssel düzleştirme yönteminin başarısı en son gözleme verilen ağırlık bir başka değişle düzleştirme katsayısının ve başlangıç değerlerinin belirlenmesine bağlıdır.

Bu çalışmada önerilen yöntemin geleneksel üssel düzleştirme yöntemiyle karşılaştırmasını kolaylaştırmak için model olarak eğim ve mevsimselliğin olmadığı tek parametrelili basit üssel düzleştirme yöntemiyle yapılacaktır. Önerilen yöntemin geleneksel üssel düzleştirme yönteminin sahip olduğu özelliklerin sağlandığı teorik olarak gösterilecek ayrıca tahmin performansları uygulamalı olarak gerçek veri setleri üzerinde gösterilecektir.

#### Kaynaklar

- [1] Brown, R.G., (1962). Smoothing, Forecasting and Prediction of Discrete Time Series. Prentice-Hall, Englewood Cliffs, N.J.
- [2] Gardner, Jr.E.S., (1985). Exponential smoothing: the state of the art. Journal of Forecasting, 4: 1-28.
- [3] Holt C., (1959) Forecasting seasonal and trends by exponentially weighted moving averages. ONR Research Memorandum 52.
- [4] Makridakis, S. and Hibon, M., (2000), The M3-Competition: results, conclusions and implications. International Journal of Forecasting, 16: 451-476.
- [5] Winters, P.R., (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6: 324-342.

#### ABSTRACT

##### MODIFIED EXPONENTIAL SMOOTHING METHOD

In this study, a new exponential smoothing framework will be introduced as an alternative to traditional exponential smoothing method. This new framework will be called modified exponential smoothing (MES) since it will be obtained from traditional exponential smoothing method by modifying the smoothing parameter. By means of this modification, initial value and optimum smoothing constant will be obtained simultaneously. The main concept in this modification is giving less weight to initial value and old past observations and transferring this extra weight to recent past observations to obtain more accurate and more adaptive forecasting results.

**Key Words:** Time series, exponential smoothing, smoothing parameter, initial value

#### HAVALİMANI YOLCU TALEBİNİN BOX-JENKİNS YÖNTEMİ İLE ÖNGÖRÜLENMESİ: ANTALYA HAVALİMANI ÖRNEĞİ

Hakan DEMİRGİL, Hakan BOZDAĞ

Süleyman Demirel Üniversitesi İ.İ.B.F.Ekonometri Bölümü, Süleyman Demirel Üniversitesi İ.İ.B.F.Ekonometri Bölümü Mail: hakandemirgil@sdu.edu.tr, hakanbozdag@sdu.edu.tr

#### Giriş

Öngörü, her planlama yapılan alanda olduğu gibi hava yolu şirketlerinin karlarını belirlemede kritik bir rol oynamaktadır. Bu nedenle havayolu yolcularının talebinin modellenmesi ve öngörülenmesi geçmişte ve günümüzde önemli ve stratejik bir araştırma konusu olmuştur.

Bu çalışmada tek değişkenli zaman serisi yöntemi olan Mevsimsel Box- Jenkins (SARIMA) yöntemi yardımıyla Eylül 2013 - Ağustos 2014 dönemi için Antalya Havalimanı' na yönelik kısa dönem uluslararası yolcu talebinin öngörülenmesi amaçlanmıştır. Araştırmada, havalimanının yolcu talebinin ölçütti olarak havalimanından giriş yapan toplam uluslararası yolcu sayısı alınmış ve Ocak 2004-Ağustos 2013 döneminde Antalya Havalimanı dış hatlar terminalinden giriş yapan aylık uluslararası yolcu sayısı verilerinden yararlanılmıştır.

#### 2. Yöntem

Zaman serisi modeli oluşturmada Box-Jenkins (1976) yöntemi geleneksel ekonometrik modellere göre, durağanlığı, deterministik bileşen bilgisini ve geleceğe ilişkin tahminleri bir arada ortaya koyduğu için sık tercih edilen bir yöntemdir.

Yöntem belirlenen birçok model arasından en iyi modeli seçerek zaman serisinin gelecekte alacağı değerleri tahmin etmeye yöneliktir. Bir değişkene ilişkin yapılacak tahmin, kendi gecikmeli değerleri, hata terimleri ya da her ikisinin birleşimi ile yapılmaktadır. Yani değişken, kendi dinamikleri ile açıklanmaya çalışılmaktadır.

Box-Jenkins yaklaşımın da temel fikir, cimrilik (tutululuk) prensibine dayanmaktadır. Cimrilik (seyreklik-azlık anlamında) prensibi zaman serisi verilerinin özelliklerini ortaya koyan optimal (minimum sayıda parametre veya serbestlik derecesini gözönünde tutan) bir model kurmayı amaçlar. Box-Jenkins, tutumlu modellerin çok sayıda parametre içeren modellere göre daha iyi öngörü ürettiklerini öne sürer. Tutumlu bir modelin verilere uyumu, gereksiz herhangi bir parametrenin eklenmesinden daha iyidir.

Yaklaşımdaki temel adımlar genel hatlarıyla, zaman serisi modelinin belirlenmesi (tanımlanması), model parametrelerinin tahmin edilmesi, ayırt edici kontrol (test) ve ileri yönelik tahmin (önraporlama-öngörü) olarak dört adımda özetlenebilir.

Durağan bir zaman serisinin t-dönemindeki gözlemlerinin, bir önceki yılın aynı dönemine karşı gelen dönemine ait gözlemlerinin artı rassal şokun doğrusal bir fonksiyonu olarak ifade edildiği durumda sürecin mevsimsel otoregresif süreç olduğu ifade edilmektedir. Durağan-dışı zaman serisinin mevsimsel AR ve mevsimsel MA süreçlerine sahip olması durumunda oluşturulan modeller, Mevsimsel Bütünleşik Otoregresif-Hareketli Ortalama Modeli olarak adlandırılmakta ve SARIMA(P;D;Q)veya ARIMA(P;D;Q)s biçiminde gösterilmektedir. Burada kullanılan P;D;Q simgeleri sırası ile mevsimsel AR bileşeni mertebesini, mevsimsel fark alma derecesini ve mevsimsel MA bileşeni mertebesini göstermektedir. Bu modellerde zaman serisi gözlemleri sadece mevsimsel gecikmelerde ve katlarında bağımlı olmakta, mevsim içindeki gözlemler bağımsız olmaktadır. Mevsim etkisi taşıyan seriler, hem mevsimsel olmayan hem de mevsimsel olan bileşen olarak modelde yer alacakları için modellerde bu ayrımı uygun olarak mevsimsel olan parametreler ve mevsimsel olmayan parametreler olarak ifade edilmeleri gerekmektedir.



ZAMAN SERİLERİ ANALİZİNDE YENİ BİR YAPAY SİNİR AĞLARI ÇALIŞMASI VE  
KARŞILAŞTIRMALI ANALİZ

Araş.Gör. Selim DÖNMEZ

Eskişehir Osmangazi Üniversitesi F1 blok Meşelik Kampüsü posta kodu:26480 sdonmez3@gmail.com

KAYNAKLAR

- [1] Çuhadar, M., Güngör, İ., ve Gökso, A., (2009 C: 14), Turizm Talebinin Yapay Sinir Ağları ile Tahmini ve Zaman Serisi Yöntemleriyle Karşılaştırmalı Bir Analizi : Antalya İline Yönelik Bir Uygulama, Süleyman Demirel Üniversitesi İ.İ.B.F. Dergisi.
- [2] Box, G. E .P., Jenkins S. G., Rreinsel G. C. (1994), Time Series Analysis: Forecasting and Control., New Jersey, Prentice-Hall Inc., Third Editions.
- [3] Montgomery, D.C., Jennings, L.A. ve Külahcı, M., (2008), Introduction To Time Series Analysis And Forecasting, New Jersey, John Wiley & Sons. Inc.
- [4] Grubb, H., Mason, A.,(2001 Volume: 17), Long Lead Time Forecasting Of UK Air Passengers By Holt Winters Methods With Damped Trend, Journal of Air Transport Management.
- [5] Abdelghany, A., ve Guzhva, V. S., (2010 Volume: 5), A Time-Series Modeling Approach For Airport Short-Term Demand Forecasting, Airport Management.

ABSTRACT

FORECASTING AIRPORT PASSENGER DEMAND BY BOX-JENKINS : A CASE STUDY OF ANTALYA AIRPORT

In this study it is aimed to build Seasonal Box-Jenkins model (SARIMA) and with this model to foresight the short term international passenger demand of Antalya International Airport for September 2013 and August 2014. In this research it is used the total number of international passenger arrivals as a measure of inbound international passengers demand and monthly international passenger arrivals to Antalya in the period of January 2004 – August 2013 data were utilized to build appropriate model. As a consequence of several attempts it has been observed that SARIMA(P;D;Q) model has presented best performance and by the means of this model it has been forecasted the monthly inbound tourism demand to Antalya Airport for September 2013 and August 2014.

**Key Words:** Time Series, SARIMA Models, Airport Passangers Demand, Forecast

Yapay Sinir Ağları Ve Zaman Serileri Analizi

Yapay sinir ağları metodolojisi, 20. yüzyılın ortalarından bu yana önemli bir ölçüde gelişmiştir. Başlarda McCulloch ve Pitts(1943) tarafından temelleri atılan yapay sinir ağları modelleri daha sonra Rosenblatt(1958), Hopfield(1982) gibi öncüler tarafından modernleştirilmiştir. Daha sonraki yıllarda yapay sinir ağları çeşitli alanlarda uygulamaya başlanmıştır. Bunların başında fonksiyon yaklaşımı, örüntü sınıflandırması ve zaman serileri analizi gelmektedir. Zaman serileri analizi üzerine yazılan makalelerden bugüne kadar 2418'i yapay sinir ağları metodolojisinin kullanımını içermektedir([http://apps.webofknowledge.com/summary.do?SID=X1b8jaLfhaKeP39A1E4&product=WOS&qid=1&search\\_mode=GeneralSearch](http://apps.webofknowledge.com/summary.do?SID=X1b8jaLfhaKeP39A1E4&product=WOS&qid=1&search_mode=GeneralSearch)). Zaman serileri analizinde yapay sinir ağları haricinde kullanılan yöntemlerin arasında ARIMA metodolojisi, Bayeşçi istatistik metodolojisi, bulanık mantık metodolojisi ve sağlamcı istatistiksel yaklaşımlar bulunmaktadır. Bu yöntemlerin belli avantajları olduğu gibi belli dezavantajları bulunmaktadır. Bu durumda araştırmacının bir zaman serisini analiz ederken yapılabileceği en iyi şey, bu yöntemleri karşılaştırarak bir sonuca varmaktır. Zaman serileri analizinde kullanılan yapay sinir ağları modelleri, pek çok alanda parametrik olmayan bir model çeşidi olarak son derece işlevsel olabilmektedir ancak istatistiksel bir teoriye dayanmadıkça bunu zaman serileri analizinde uygulamak son derece zordur. Bu zorluğun temel kaynağı, ideal bir yapay sinir ağı modelinde sinaptik ağırlıkların durumundan kaynaklanır. Bu konu Aladağ(2011) tarafından incelenmiş ve tabu algoritması kullanarak en iyi yapay sinir ağı tasarımını bulmaya çalışmıştır. Bu verimli sonuç veren çalışmaya rağmen yapay sinir ağları tasarımının, zaman serilerinde öngörü problemin tamamını özetlemediğini düşünmekteyiz. Herşeyden önce istatistiksel bakış açısıyla bakıldığında her zaman serisi verisi ARIMA modelleriyle veya daha az bilinen ama klasikleşmiş hal-uzayı modelleriyle(ingilizcesi state-space model) ifade edildiği görülür. Bu durumda öngörü problemi, modelleme probleminin tamamen olmasa da kısmen bağımsız olarak ele alınmalıdır. Bunun yanında bu durum, yapay sinir ağlarında sabit miktarda girdi nöron sayısının bulunabileceğini göstermektedir ki bu çalışmada bu durum kanıtlanmaya çalışılacaktır. Geliştirilen yöntemde, önce veride gerçekleştirilen standartlaştırma ile yöntemin işlemesi kolaylaştırılacaktır. İkinci adımda girdi tabakasındaki nöron miktarı istatistiksel olarak belirlenecek ve de sinaptik ağırlıklar için en küçük kareler tahmin yöntemiyle güven aralıkları geliştirilerek çözüm arama uzayı belirtilecek ve orda çözüm aranacaktır. Örneğin istatistiksel zaman serileri arasında meşhur olan gaz çıkışı(gas furnace data) verisinde model aşağıdaki şekilde oluşturulup uygulandığında aşağıdaki sonuçlar çıkmıştır:

$$y_{i+1} = \mu + a_1 * y_i + b_1(x_{i+1} - b_2x_i) \quad (1)$$

EğitimRMSE	TestRMSE	Mu katsayısı	a1 katsayısı	b1 katsayısı	b2 katsayısı
0.0056755	0.0046889	0.2512	0.815	0.0464	-0.0725

Bunun yanında Bass ve Clarke(1972)'nin satış oranları ile reklamcılık oranları arasında ilişki arayan makalesindeki veriyi aynı modelle analiz ettiğimizde aşağıdaki sonuçlarla karşılaştık:

**Çizelge2.**

EğitimRMSE	TestRMSE	Mu katsayısı	a1 katsayısı	b1 katsayısı	b2 katsayısı
0.015467	0.0095021	0.1622	0.4691	0.2794	0.1676

Bu sonuçlar her ne kadar veri standartlaştırması sayesinde elde edilmiş gözükse de veri standartlaştırması olmadan da iyi sonuçlar elde etmek de mümkündür. Örneğin sinaptik ağırlıkların aralıklarını Tiku'nun yöntemiyle oluşturursak Bass ve Clarke(1972)'nin verisinden aşağıdaki sonuçları elde edebiliriz:

**Çizelge3.**

EğitimRMSE	TestRMSE	Mu katsayısı	a1 katsayısı	b1 katsayısı	b2 katsayısı
15.371	5.4129	7.7432	0.5499	0.0673	0.0807

Görüldüğü gibi yöntem oldukça iyi sonuç verebilmektedir. Bunun yanında işlem açısından büyük kolaylık da sağlanmaktadır. Bu nedenle bu yöntemin zaman serileri analizinde başarılı sonuçlar elde edeceğini görebiliriz.

**KAYNAKLAR**

[1] Bass, F. M., and Clarke, D. G.(1972), Testing distributed lag models of advertising effect, J. Marketing Research 9, 298-308.

[2] Hopfield J.J.(1982), Neural networks and Physical Systems with emergent collective computational abilities, Proceedings of the National Academy of Sciences, 79, 2554-2558

[3] McCulloch W.S. and Pitts W.(1943), A Logical calculus of the Ideas Immanent in Nervous Activity, Proceedings of the National Academy of Sciences, 79, 2554-2558

[4] Rosenblatt F.(1958), The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Psychological Review, 65, 6, 386-408

[5] Aladağ Ç.H.(2011), A new architecture selection method based on tabu search for artificial neural networks, Expert systems with applications, 38, 3287-3293

**BİLDİRİ OTURUMLARI 2**

*SESSION 2*

*Veri Madenciliği*

### KÜMELEMeye DAYANAN İŞBİRLİKÇİ FİLTRELEME - KARŞILAŞTIRMALI BİR ÇALIŞMA

Özkan ASLAN Nevin GÜLER

Anadolu Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İki Eylül Kampüsü, ESKİŞEHİR  
Muğla Sıtkı Koçman Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Kötekli Kampüsü, MUĞLA,  
[ozkanaslan@anadolu.edu.tr](mailto:ozkanaslan@anadolu.edu.tr), [nguler@mu.edu.tr](mailto:nguler@mu.edu.tr)

#### İşbirlikçi Filtreleme

İnternet, intranet ve ticari sistemlerin hızla gelişmesiyle birlikte, kullanılabilir çevrimiçi bilgi miktarı da önemli ölçüde artmıştır. Buna bağlı olarak, kullanıcılar için gerekli bilgilerin hızlı ve etkin bir şekilde bulunması oldukça zor hale gelmiştir. Bu problemin çözümü için çok sayıda bilgi filtreleme tekniği geliştirilmiştir. İşbirlikçi Filtreleme (İF) [1] bu tekniklerin başında gelmektedir. İF, filtreleme işlemi için çevrim içi kullanıcıların önceki tercihleri ve ilgi alanlarından yararlanmaktadır. Buna bağlı olarak özellikle e-ticaret alanında kullanıcılara çok sayıda web sayfası, doküman, kitap vs gibi öğeler arasında kullanıcıların ihtiyaçları doğrultusunda önerilerde bulunmak amacıyla kullanılmaktadır. İF'deki temel amaç, "geçmişte benzer ilgi alanlarına sahip kullanıcıların gelecekte de benzer davranış göstereceği" varsayımdan yola çıkarak bir grup kullanıcının tercihlerine göre "aktif kullanıcı" olarak adlandırılan kullanıcının tercihlerini tahmin etmektir. Geleneksel IF algoritması şu şekilde özetlenebilir: n tane kullanıcı  $(k_1, k_2, \dots, k_n)$  ve m tane öğeden oluşan  $(o_1, o_2, \dots, o_m)$  oluşan bir veri tabanı oluşturulur. Her kullanıcı  $(k_i)$  için görüşlerini bildirdiği veya oyladığı bir öğe kümesi belirlenir. Derece skoru matrisi oluşturulur. Derece skoru matrisi kullanıcıların öğelere belli bir aralıkta verdiği oylardan  $(d_{ij})$  oluşan  $(n \times m)$  boyutunda bir matristir. Ancak her kullanıcı tüm öğeleri oylamamaktadır ve dolayısıyla derece skoru matrisinin bazı elemanları eksiktir. İF'deki amaç, diğer kullanıcılar ile benzerliği ölçülerek aktif kullanıcı için görüşünü belirtmediği öğelere ilişkin oyu tahmin etmektir. Aktif kullanıcı ile diğer kullanıcılar arasındaki benzerliği ölçmek için çeşitli benzerlik ölçütleri kullanılır. Bu çalışmada aşağıdaki gibi tanımlanan Pearson Korelasyon kullanılacaktır:

$$PK(k_x, k_y) = \frac{\sum_{h=1}^{n'} (d_{k_x, o_h} - \bar{d}_{k_x})(d_{k_y, o_h} - \bar{d}_{k_y})}{\sqrt{\sum_{h=1}^{n'} (d_{k_x, o_h} - \bar{d}_{k_x})^2} \sqrt{\sum_{h=1}^{n'} (d_{k_y, o_h} - \bar{d}_{k_y})^2}} \quad (1)$$

Burada  $\bar{d}_{k_x, o_h}$ , x. kullanıcının, h. öğeye verdiği oyu,  $\bar{d}_{k_x}$ , x. kullanıcının y. kullanıcı ile ortak oyladığı öğelerin derecelerinin ortalamasını,  $n'$  ise x. kullanıcı ile y. kullanıcının ortak oyladığı öğe sayısını gösterir. Aktif kullanıcıya benzer kullanıcıların belirlenmesi için, bu kullanıcının tüm kullanıcılar ile benzerliğinin hesaplanması gerekmektedir. İF veritabanlarının genellikle çok büyük boyutta olduğu dikkate alındığında bu şekilde bir hesaplama oldukça maliyetlidir. Kümelemeye dayanan İF, geleneksel İF yöntemlerinin bu dezavantajını ortadan kaldırmak amacıyla önerilmiştir. İF'de kümeleme analizi uygulanmasındaki amaç, veri setini birbirine benzer kullanıcılardan oluşan alt kümelere ayırmak ve böylece aktif kullanıcı hangi kümeye daha yakın ise o kümede yer alan kullanıcılar ile arasındaki benzerliği hesaplamaktır. Benzerlikler hesaplandıktan sonra tahmin işlemi için aşağıdaki eşitlik kullanılmaktadır:

$$\hat{p}(k_a, o_a) = \bar{d}_{k_a} + \frac{\sum_{h=1}^{m'} PK(k_a, k_h)(d_{k_h} - \bar{d}_{k_h})}{\sum_{h=1}^{m'} |PK(k_a, k_h)|} \quad (2)$$

Burada  $m'$   $o_a$  öğesini oylayan diğer kullanıcıların sayısını göstermektedir.

#### Uygulama

Bu çalışmada işbirlikçi filtreleme alanındaki performanslarını değerlendirmek amacıyla hiyerarşik kümeleme yöntemleri[2], K-Ortalamalar (KO) [2][3], Bulanık C-Ortalamalar[2][3], Gustafson-Kessel[2][3], Olabilirlikli C-Ortalamalar (OCO) [3], İyileştirilmiş Olabilirlikli C-Ortalamalar [3] ve Ağırlıklandırılmış Olabilirlikli C-Ortalamalar [3] kümeleme algoritmaları Jester [4] veri setine uygulanacaktır. Burada ilk olarak veri seti 5 ayrı alt kümeye ayrılacaktır. Her veri setinin %80'i eğitim, %20'i ise test seti olarak kullanılacak ve test setinde yer alan kullanıcıların öğelere verdiği oylar tahmin edilmeye çalışılacaktır. Karşılaştırma kriteri olarak ise aşağıdaki gibi tanımlanan Ortalama Mutlak Yüzde Hata (OMYH) ve Ortalama Hata Kareler (OHK) kriterleri kullanılacaktır:

$$OMYH = \sum_{a=1}^n \left| \frac{p(k_a, o_a) - \hat{p}(k_a, o_a)}{p(k_a, o_a)} \right| * 100 \quad (3)$$

$$OHK = \sum_{a=1}^n \frac{(p(k_a, o_a) - \hat{p}(k_a, o_a))^2}{n} \quad (4)$$

Burada n test veri setinde oylanmamış öğe sayısını göstermektedir.

#### KAYNAKLAR

- [1] Su, X., Khoshgoftaar, T.G., "A Survey of Collaborative Filtering", Adv. Artif. Int., 2009.
- [2] Güler, N., Bulanık Kümeleme Analizi ve Bulanık Modellemeye Uygulamaları, Y.Lisans Tezi, Muğla Üniversitesi, Fen Bilimleri Enstitüsü, Muğla, 2006.
- [3] Güler, N., "Olabilirlikli ve Bulanık Kümelemeye Dayanan Modelleme Teknikleri ve Yazılım Güvenilirliğinin Tahminine Uygulanması", Muğla Sıtkı Koçman Üniversitesi, Fen Bilimleri Enstitüsü, 2012.
- [4] <http://goldberg.berkeley.edu/jester-data/>

#### ABSTRACT

#### COLLABORATIVE FILTERING BASED ON CLUSTERING – A COMPARATIVE STUDY

Collaborative filtering (CF) is widely used recommender systems. CF recommends items to active user by considering relationships between users. For this aim, it exploits similarities between users. However, in case of the number of users is large, calculating the similarities is very time-consuming. Clustering techniques is used to alleviate this drawback of CF. In this study, the performance of various clustering algorithms in the field of CF are compared.

**Key Words:** Recommender Systems, Collaborative Filtering, Clustering Techniques

**BANKACILIKTA PAZARLAMA SONUÇ TAHMİNİ İÇİN BİR ÖRÜNTÜ  
SINIFLANDIRMA UYGULAMASI**

Selver Ezgi YALNIZ\*, Hasan OĞUL

Başkent Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü,06810, Ankara, TÜRKİYE,  
seyalniz@baskent.edu.tr, hogul@baskent.edu.tr

Ülkemizde son yıllarda bankacılık sektörü zor bir dönem yaşamaktadır. Özellikle 2001 yılında meydana gelen ekonomik kriz sonrasında bankalar müşteri ağını genişletebilmek amacıyla müşteri odaklı hizmetlere önem vermeye başlamıştır. Birçok bankanın rekabet ortamında başarılı olabilmeleri için etkin pazarlama teknikleri geliştirmeleri ve dolayısıyla müşterileri ile doğrudan iletişime geçmeleri gerekmektedir. Bu çalışmada, vadeli hesap açtırma üzerine yapılan bir kampanya süreci sonunda, kampanya bilgileri ve müşteri profilinden çıkarılacak bazı özelliklere göre herhangi bir müşterinin hesap açtırıp açtırmayacağını kestiren bir akıllı sistem alt yapısı üzerinde çalışılmıştır. Bu amaçla çok sayıda öğreticili sınıflandırma algoritması özellik azaltma yöntemleriyle birlikte değerlendirilmiş ve karşılaştırılmalı sonuçları raporlanmıştır.

Uygulamada bir bankanın pazarlama kampanyasında kullandığı gerçek hayattan alınmış veriler kullanılmıştır. Kullanılan yöntemler ise; Temel Bileşen Analizi (Principal Component Analysis), Fisher Doğrusal Ayırıcı (Fisher Linear Discriminant Analysis), En Yakın K Komşu Algoritması (K Nearest Neighbor Algorithm), Naive Bayes Sınıflandırıcısı (Naive Bayes) ve Destek Vektör Makinası (Support Vector Machine) metodlarıdır. Uygulanan çeşitli metotlara göre veri kümesi üzerinde değerlendirme kriterleri belirlenmiş ve bu kriterler üzerinden karşılaştırma yapılarak en iyi sonuca ulaşmaya çalışılmıştır.

Uygulamada kullanılan veri kümesi UCI- Machine Learning Repository’ den alınan Bank Marketing isimli veri kümesidir. Bu veri kümesi Mayıs 2008 ve Kasım 2010 yılları arasında bir Portekiz bankasının 17 farklı pazarlama kampanyasında kullanılması için gerçek kişilerden toplanmıştır [3]. Veriler toplamda iki set halinde sunulmuştur. Bu iki setten birincisi 45211 örneği içeren orijinal veri kümesi iken, diğeri bu orijinal veri kümesinin %10 azaltılmış bir versiyonu olarak

4521 örneği içermektedir. Uygulamada makine öğrenme algoritmalarının hesaplarında zaman probleminin yaşanmaması için 4521 örnek içeren küçük veri kümesi kullanılmıştır. 4521 örnek içinde 4000 tanesi vadeli mevduat açılmayan müşterileri belirlerken, diğer 521 tanesi vadeli mevduat hesabı açılmış olan müşterileri göstermektedir. Veri kümesi toplamda 14 özellik (13 özellik + 1 sınıf bilgisi) ile ifade edilmiştir. Buna göre veriler sürekli (continuous), sayısal (numeric) ve ikili (binary) özelliklere sahiptir Tablo.1’de veri kümesindeki özellikler, tanımları ve türleri verilmiştir.

**Tablo 1: Verilerin Özellikleri**

Özellik Adı	Açıklama	Türü
<b>Kişisel Müşteri Bilgileri</b>		
Yaş	Müşterinin yaşı	Sayısal
İş	Müşterinin iş durumu (sistem yöneticisi, işsiz, yönetici, v.b.)	Sürekli Kategorik
Evlilik Durumu	Müşterinin evli olup olmadığı	İkili (Evli(1) veya Değil(0))
Eğitim Düzeyi	Müşterinin eğitim düzeyi (ilkokul, ortaokul, lise, üniversite mezunu)	Sürekli Kategorik
Kredi Durumu	Müşterinin güncel kredisinin olup olmadığı	İkili (Var(1) veya Yok(0))
Hesap Bakiyesi	Müşterinin yıllık ortalama bakiye bilgisi	Sayısal
Konut Durumu	Müşterinin ikamet ettiği konutun kira olup olmadığı	İkili (Kira(1) veya Değil(0))
Borç Durumu	Müşterinin bireysel borcunun olup olmadığı	İkili (Var(1) veya Yok(0))
<b>Güncel Kampanya Bilgileri</b>		
Konuşma Süresi	Müşteri ile görüşmenin süre bilgisi	Sayısal
Görüşme Sayısı	Müşteri ile kampanya için kaç kez görüşüldüğü	Sayısal

<b>Önceki Kampanya Bilgileri</b>		
Gün Sayısı	Müşteri ile bir önceki kampanyada son görüşmenin üzerinden kaç gün geçtiği.	Sayısal
Toplam Görüşme Sayısı	Müşteri ile daha önceki kampanyalar için toplamda kaç kez görüşüldüğü	Sayısal
Başarı Durumu	Bir önceki kampanyanın başarıya ulaşip ulaşmadığı	İkili (Başarılı(1) veya Başarısız(0))

Veri kümesinde sürekli olarak belirtilen müşterinin iş durumunu gösteren özellik, farklı gelir düzeylerine göre belirlenmiştir. Buna göre -1 ile 8 arasında sayılar farklı durumlarının maaşlarına göre artan bir sırayla atanmıştır. Yine sürekli bir özellik olarak belirtilen müşterinin eğitim düzeyi, farklı eğitim seviyelerine uygun olarak verilmiştir.

Buna göre 4 farklı eğitim düzeyi için 0-4 arası değerler müşterilere atanmıştır (Örneğin üniversite mezunu olan kişiye 4 değeri atanmıştır.). Eğitim düzeyi bilinmeyen müşteri için girilen değer -1 değeridir. Veri kümesine öğreticili öğrenme yöntemleri uygularken 2 aşamalı çapraz doğrulama (cross validation) uygulanmıştır. Bu şekilde eğitim kümesinden matematiksel olarak bir model hesaplanıp, modelin doğruluğu test kümesinde denetlenmektedir. Yapılan deneylere göre (Tablo 2), en iyi sonuç özellik azaltım olmadan SVM uygulamasıyla elde edilmiştir. Alınan sonuçlar böyle bir yaklaşımın bankacılık karar destek sistemlerinde uygulanabilirliğini göstermektedir.

**Tablo 2: Uygulama Sonuçları**

Yöntem	Doğruluk	Kesinlik	Duyarlılık	Özgüllük	MKK
FLDA	86%	43%	61%	89%	44%
FLDA/PCA	68%	12%	27%	74%	1%
k-NN (k=25)	89%	56%	10%	99%	71%
k-NN/PCA	88%	0%	0%	100%	0%
NB	88%	49%	38%	95%	36%
NB/PCA	88%	15%	1%	99%	1%
SVM	92%	60%	82%	93%	66%
SVM/PCA	81%	37%	90%	80%	50%
SVM	94%	66%	96%	94%	76%

**KAYNAKLAR**

- [1] Ling, X. and Li, C., “Data Mining for Direct Marketing: Problems and Solutions”. 4th KDD conference, AAAI Press, 73–79, 1998.
- [2] Page, C. and Luding, Y., “Bank manager’s direct marketing dilemmas – customer’s attitudes and purchase intention”. International Journal of Bank Marketing, 147–163, 2003.
- [3] Moro S., Laureano R., Cortez P. “Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology”, Euro. Sim. Model. Conf. (ESM’2011) pp. 117-121, 2011

**Key Words:** Principal Component Analysis, PCA, Fisher Linear Discriminant Analysis, FLDA, K Nearest Neighbor Algorithm, k-NN, Naive Bayes, NB, Support Vector Machine, SVM.

## LSA VE PLSA İLE TÜRKÇE METİNLERİN ANLAMSAL BENZERLİĞİNİN ÖLÇÜLMESİ

Volkan UZUN, Engin YILDIZTEPE,

Dokuz Eylül Üniversitesi, Kaynaklar Yerleşkesi, Fen Fakültesi İstatistik Bölümü, 35160, Buca/İzmir/TÜRKİYE  
[volkanuzun87@gmail.com](mailto:volkanuzun87@gmail.com) , [engin.yildiztepe@deu.edu.tr](mailto:engin.yildiztepe@deu.edu.tr)

### 1. Giriş

Günümüzde internet kullanımı sayesinde her geçen gün elektronik ortamda saklanan veri miktarı artmaktadır. Teknolojik imkânlar sayesinde veriye erişim kolaylaşmıştır. Ancak farklı türdeki bu veri yığınları arasından istenilen bilgiye hızlı ve doğru erişmek giderek daha fazla önem kazanmaktadır. Metin halindeki verilerden yararlı ve anlamlı bilgilerin edinilmesi için kullanılan yöntemler metin madenciliği olarak tanımlanabilir. Metin madenciliği, doğal dil işleme, bilgi geri kazanımı, istatistik gibi bilim dallarından yararlanır. Metinlerin anlamsal benzerliklerinin ölçülmesi, anlamsal ilişkilerinin belirlenmesi, metin özetleme gibi konular metin madenciliğinin çalışma alanlarındandır. Bu çalışmada Türkçe metinlerin benzerliklerinin ölçülmesi için Gizli Anlam Analizi (Latent Semantic Analysis – LSA) ve Olasılıksal Gizli Anlam Analizi (Probabilistic Latent Semantic Analysis – PLSA) yöntemleri incelenmiştir. LSA ve PLSA metin benzerliklerinin belirlenmesi konusunda önerilen yöntemlerin başında gelmektedir. İkinci ve üçüncü bölümlerde bu iki yöntem hakkında detaylı bilgi verilmiş, uygulama bölümünde Türkçe ve İngilizce metinler üzerindeki sonuçlar tartışılmıştır.

### 2. Latent Semantic Analysis – LSA

LSA, 1990'ların başında metinlerin anlamsal ilişkilerinin belirlenmesi için önerilmiş bir yöntemdir (Derweester vd, 1990). LSA, Tekil Değer Ayrıştırması (Singular Value Decomposition) kullanarak, metinlerin anlamları arasındaki ilişkileri belirlemeye çalışan bir yöntemdir. Metinler bazı metin temsil yöntemleri ile matris halinde ifade edildikten sonra Tekil Değer Ayrıştırması metodu ile üç temel matrise ayrıştırılır daha sonra uygun boyut indirilmesi yapılarak sonuç matrisi elde edilir. Elde edilen bu sonuç matrisine uygun benzerlik hesaplama yöntemleri uygulanarak benzerliğin derecesi hesaplanabilir. LSA, günümüzde daha çok metin özetleme ve metin benzerliğinin bulunması konularında kullanılmaktadır.

### 3. Probabilistic Latent Semantic Analysis - PLSA

PLSA, LSA metodundan farklı olarak Expectation-Maximization (EM) algoritmasından faydalanır. Metinlerin ve kelimelerin hangi olasılıklarla hangi konulara ait olduğunu belirlemeye çalışır (Hofmann, 2001). Başlangıçta, kelime ve doküman olasılıkları rassal olarak, konu olasılıkları ise konu sayısına göre eşit olarak belirlenir. Ardından expectation adımında bu olasılıklar kullanılarak yeni değerler belirlenir ve bu yeni değerler ile maximization adımında yeni olasılıklar hesaplanır. Yeni olasılıklar değerleri ile bir önceki olasılıklar arasında fark önceden belirlenen eşik değerden küçük oluncaya kadar işlem devam ettirilir. PLSA sadece metin benzerliklerinin belirlenmesinde değil ses ve görüntü verilerinin sınıflandırılmasında da kullanılmaktadır.

### 4. Uygulama

Bu çalışmada Türkçe metinler arasındaki benzerliğin istatistiksel olarak ölçülmesinde LSA'nın ve PLSA'nın performansı araştırılmıştır. İlk bölümde, kullanılan yöntemlerin teorisi ve bu yöntemleri Türkçe metinlerde uygulayabilmek için gereken işlemler anlatılmış, uygulama kısmında ise haber ajanslarından elde edilen farklı konulardaki metinlerin anlamsal olarak benzerliklerinin ölçümleriyle ilgili sonuçlar verilmiştir. Elde edilen sonuçlara göre LSA'nın ve PLSA'nın Türkçe metinlerde anlamsal benzerliğin ölçümü için kullanılabileceği görülmüştür.

Anahtar Kelimeler: Gizli anlam analizi, Olasılıksal Gizli Anlam Analizi, Metin madenciliği

## KAYNAKLAR

- [1] Deerwester S., Dumais S., Furnas G., Landauer T., Harshman R. (1990), Indexing by Latent Semantic Analysis. Journal of the American Society for Information Sciences, 41(6), 391-401.
- [2] Hofmann, T., (2001), Unsupervised Learning by Probabilistic Latent Semantic Analysis, Machine Learning, 42, 177-196
- [3] Landauer, T.K., Foltz, P.W., Laham, D. (1998), Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.
- [4] Nakov, P., Popova, A., Mateev. P., "Weight functions impact on LSA performance", The EuroConference Recent Advances in Natural Language Processing RANLP'01, Bulgaristan, 2001.
- [5] Wang, Z., Tsim, Y.C., Yeung, W.S., Chan, K.C., Jinlan, Liu., (2007), Probabilistic latent semantic analysis (PLSA) in bibliometric analysis for technology forecasting, Journal of Technology Management Innovation, 2, 11-24.

## ABSTRACT

### MEASURING SEMANTIC SIMILARITY IN TURKISH TEXTUAL DATA USING LSA AND PLSA

Text summarization and semantic measurement are current text mining topics. Latent Semantic Analysis (LSA) is an important method to determine semantic similarity between texts. Probabilistic Latent Semantic Analysis (PLSA) is another method to determine semantic relations between texts. In this paper, LSA and PLSA performances were researched in measuring semantic similarity in Turkish textual data. Twelve news articles were chosen from three different news agency and LSA and PLSA were used to measure semantic similarity. According to results, PLA and LSA can be used to measure semantic similarity in Turkish textual data.

**Key Words:** Latent Semantic Analysis, Probabilistic Latent Semantic Analysis, Text Mining

## EKG'DEKİ NORMAL VE ERKEN KARINCIK KASILMASI VURULARININ SINIFLANDIRILMASI

Öğr. Gör. Yasin KAYA  
Karadeniz Teknik Üniversitesi Enformatik Bölümü  
yasin@ktu.edu.tr

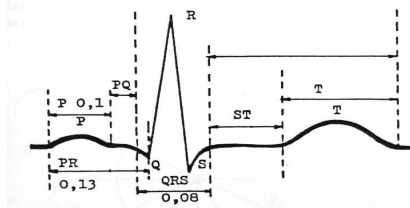
### 1. Giriş

Günümüzde ölüme neden olan en önemli problemlerden biri kalp rahatsızlıklarıdır. Kalp rahatsızlıklarının erken teşhisi ve tedavisi ani ölümleri önleyebilir[3]. İnsan vücudu üzerinden algılanan ve kalbin aktivitesi gösteren elektriksel işaretlere elektrokardiogram (EKG) adı verilir. İnsan ölümlerinin büyük bir yüzdesi kalp rahatsızlıklarından oluşmaktadır. Bu yüzden kalbin çalışması sırasındaki bozuklukların iyi bir göstergesi olan ve hastaya zarar vermeden, vücut üzerinden kolaylıkla elde edilebilen EKG işaretleri, işleme ve yorumlama açısından büyük önem taşımaktadır[3]. Bu çalışmada Normal ve Erken Karıncık Kasılması vurularının hızlı bir şekilde ve yüksek doğrulukta sınıflandırılması için sınıflandırma algoritmalarının sonuçları irdelenmiştir.

### 2. Materyal ve Metot

#### 2.1. EKG verileri

Normal EKG işareti, kalbin dinlenme durumundaki taban seviyesi üzerine sıralanan belli başlı P, Q, R, S ve T adları verilen dalgalardan oluşur. Bazen T dalgasını takiben küçük genlikli bir U dalgası da olabilir[2].



Şekil 1. Normal bir EKG vurusundaki önemli dalgalar

Normal bir kalp atımı, dalganın şekli, süresi, RR aralığı gibi temel parametreler içerir. Bu parametrelerdeki değişimler kalpte bir rahatsızlık olduğunu gösterir. Bu düzensiz dalga fazı aritmi olarak isimlendirilir ve aritmiler hasta için tehlikelidir[2-3].

#### 2.2. EKG veritabanı

Çalışmada Massachusetts Teknoloji Enstitüsü – Beth İsrail Hastanesi (Massachusetts Institute of Technology – Beth Israel Hospital MIT-BIH) aritmi veritabanından alınan 48 EKG kaydı kullanılmıştır[1].

### 2.3. EKG verilerinin sınıflandırılması

EKG verilerinin işlenmesinde karar verme işlemi genellikle üç aşamada gerçekleştirilmektedir: Ön işleme ve normalleştirme işlemi, Öznitelik çıkartma işlemi, Sınıflandırma işlemi.

### 3. Yapılan Çalışmalar

Özellik çıkarma ve sınıflandırma işlemlerinden önce giriş sinyali ilk olarak pencere genişliği 3 olan bir ortalama filtresi kullanılarak sinyal üzerindeki gürültüler azaltılmıştır. Sinyalde bulunan daha büyük sapmaları gidermek için (özellikle ana çizgiden sapmalar - baseline drift) sinyal filtrelenmiştir. Burada ana çizgiden sapmanın giderilmesi için 2Hz'in altındaki frekans bileşenleri sinyalden alçak geçiren (low pass) filtre kullanılarak çıkarılmıştır. Sınıflandırma işlemi için özellik vektörü olarak bir kalp vuru sinyalinin zaman serisi kullanılmıştır. Bir vuru sinyalini elde etmek için sinyalde R tepeleri tespit edilmiş ve 100 örnek önce ve 150 örnek sonra olmak üzere 250 örneklilik bir vektör oluşturulmuştur. Sınıflandırma için iki ayrı sınıfta (N-Normal vuru, V-Erken karıncık kasılması) yapılmıştır.

Sınıflandırma işlemlerinde kullanılan sınıflandırma algoritmaları ve başarımları Çizelge 1'de verilmiştir.

Çizelge1. Sınıflandırma sonuçları

Sınıf /Algoritma	NN	KNN	Bayes	DT	ID3	SWM
N	99,6	99,76	92,53	99,23	98,71	99,50
V	93,6	94,04	88,82	92,52	90,64	59,39

### KAYNAKLAR

- [1] Moody, G. B. and Mark R. G.(1990), The MIT-BIH arrhythmia database on CD-ROM and software for use with it. Computers in Cardiology. IEEE Computer Society Press, p. 185-188.
- [2] Yazgan, E. ve Korürek, M. (1996), Tıp Elektronik, İTÜ yayınları, ISBN:975-561-073-1
- [3] Dokur, Z. (1998), Yapay Sinir Ağları ve Genetik Algoritmalar Kullanılarak EKG vurularının Sınıflandırılması, Doktora Tezi, Fen bilimleri Enstitüsü, İstanbul Teknik Üniversitesi

### ABSTRACT

#### THE CLASSIFICATION OF NORMAL AND PREMATURE VENTRICULAR CONTRACTION IMPULSES OF ECG

Today, one of the most important problems that cause death is heart disorders. Early detection and treatment of heart disease may prevent sudden death. Detected through the human body and seen as a result of activity of the heart's electrical signals is called electrocardiogram (ECG). Decision-making process of ECG data processing is generally full fill in three stages: Pre-processing and normalization process, the feature extraction process, classification process. In this study compared the results of the classification algorithms for the classification of Normal and Premature Ventricular Contraction impulses of ECG data.

**Key Words:** EKG, NN, SWM, Bayes, Karar Ağaçları

## GÖĞÜS KİTLELERİNİN YOĞUNLUKLARININ MAMOGRAFİ GÖRÜNTÜLERİNDEN İSTATİSTİKSEL OLARAK ELDE EDİLMESİ

Tolga BERBER, Uğur ŞEVİK

Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü

tberber@ktu.edu.tr, usevik@ktu.edu.tr

Adil ALPKOÇAK

Dokuz Eylül Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü

alpkoçak@cs.deu.edu.tr

Pınar BALCI

Dokuz Eylül Üniversitesi Tıp Fakültesi Radyodiyagnostik Anabilim Dalı

Pinar.balci@deu.edu.tr

### 1. Giriş

Günümüzde mamografi görüntüleri, göğüs kanseri teşhisinde kullanılan en önemli araçlardan biridir. Meme kanserlerinin başarılı bir şekilde tedavi edilmesi her şeyden önce erken tanıya bağlıdır. Mamografi bu noktada, yani meme kanserinin erken tanısında, çok önemli bir rol oynar. Son yıllarda, bilgisayar destekli tanı (BDT) alanında birçok çalışma otomatik veya yarı otomatik olarak göğüs kitlelerini algılama ve tanıma üzerine yoğunlaşmıştır. Bu çalışmaların birçoğunda temel olarak istatistiksel ve matematiksel yöntemler kullanılmıştır. Özellikle, görüntülerden istatistiksel öznitelik çıkarımı; bu özniteliklere dayalı makine öğrenmesi sistemlerinin geliştirilmesi ve bu sistemlerin kullanımı ile otomatik olarak kitle sınıflandırılması en sık kullanılan yöntemlerdir.

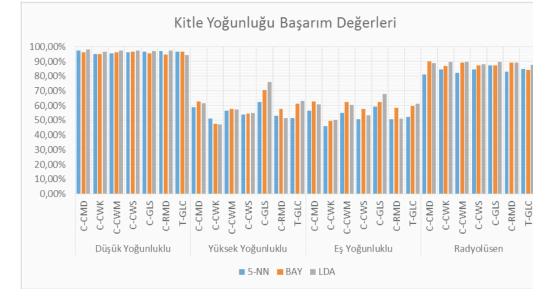
### 2. İstatistiksel Öznitelik Çıkarımı

Dokuz Eylül Üniversitesi Radyodiyagnostik Anabilim Dalı Mamografi biriminden elde edilen 260 göğüs kitlesi görüntüsü kullanılarak, kitlelerin yoğunluk özellikleri istatistiksel olarak tahmin edilmeye çalışılmıştır. Veri setimizde 119 adet yüksek yoğunluklu, 111 eş yoğunluklu, 6 adet düşük yoğunluklu ve 24 adet radyolüsen kitle bulunmaktadır [1]. Öznitelik vektörlerinin %10'u eğitim amacıyla kullanılmış ve sistem 10 aşamalı çapraz doğrulama ile sınanmıştır. Çalışmada gri seviye birlikte görülme matrisinden (Gray Level Co-occurrence Matrix; GLCM) ayıklanan öznitelikler [2] ile tarafımızdan önerilen ve kitlenin polar gösteriminin istatistiksel özellikleri kullanılmıştır [3]. Bu öznitelikler birlikte kullanılarak 3 adet sınıflandırıcı eğitilmiş ve başarımları sınanmıştır.

### 3. Yöntem

Mamografi görüntülerinden ayıklanan öznitelikler, literatürde sıklıkla kullanılan 3 adet gözetimli öğrenme yöntemleri kullanılarak sınıflandırıcılar geliştirilmiş ve bu sınıflandırıcıların göğüs kitlelerinin özelliklerini sınıflandırma başarımı sınanmıştır. Çalışmamıza; Saf Bayes [4], Doğrusal Ayırıcı Analizi [5] ve k-En Yakın Komşu öğrenme yöntemleri kullanılmıştır. Elde edilen sonuçlar doğruluk ölçüğü kullanılarak ölçülmüş ve sonuçlar Grafik 1'de verilmiştir.

### 4. Sonuçlar



Grafik 1. Yoğunluk özelliği sınıflandırma sonuçları.

Sonuçlara göre düşük yoğunluklu ve radyolüsen kitleler yüksek yoğunluklu ve eş yoğunluklu kitlelerden daha başarılı bir şekilde sınıflandırılmışlardır. Özellikle düşük yoğunluklu kitleler polar gösteriminden ayıklanan istatistiksel öznitelikler, doğrusal ayırıcı analizi sınıflandırıcısıyla birlikte kullanıldığında %98,06 oranında bir başarımları elde edilmiştir.

### KAYNAKLAR

- [1] Akçay O., Alpköçak A., Balci P., and Dicle O., (2009), "MuDMaDs: A Multipurpose Digital Mammography Reference Dataset," Proceedings of 6th National Medical Informatics Conference, Antalya, Turkey.
- [2] Haralick R. M., Shanmugam K., and Dinstein I. H., (1973), "Textural features for image classification," IEEE Transactions on Systems, Man and Cybernetics, 3(6), pp. 610-621.
- [3] Mierswa I., Wurst M., Klinkenberg R., Scholz M., and Euler T., (2006), "YALE: Rapid prototyping for complex data mining tasks," Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, pp. 935-940.
- [4] Zhang H., (2004), "The Optimality of Naive Bayes," Proceedings of the 17th International FLAIRS conference, AAAI Press, Menlo Park, CA.
- [5] Mika S., Ratsch G., Weston J., Scholkopf B., and Mullers K., (1999), "Fisher discriminant analysis with kernels," Neural Networks for Signal Processing IX, Proceedings of the 1999 IEEE Signal Processing Society Workshop, IEEE, pp. 41-48.

### ABSTRACT

#### IDENTIFICATION OF BREAST MASS DENSITIES FROM MAMMOGRAPHY IMAGES USING STATISTICAL METHODS

In this study, we measured performance of statistical low-level image features on breast mass density classification task. According to the results, LDA classifier with statistical properties of polar mass representation feature obtains the highest accuracy score with 98.06%.

**Key Words:** Breast cancer, Mass density, Machine Learning, Statistical Feature Extraction

FISHER INFORMATION IN PROGRESSIVE TYPE II CENSORED ORDER STATISTICS AND THEIR CONCOMITANTS FOR SOME COPULA MODELS

Tuğba YILDIZ, Burcu ÜÇER\*

Dokuz Eylül University, Department of Statistics, Faculty of Science, 35160 Buca, Izmir, TURKEY,  
tugba.ozkal@deu.edu.tr, burcu.hudaverdi@deu.edu.tr

Introduction

Fisher information takes part in statistical inference through the Cramer-Rao inequality and its association with the asymptotic properties of the maximum likelihood estimators. The problem of finding the efficient estimators caused the researchers to take the evaluation of corresponding Fisher information into concern. The Fisher information is closely related to sufficiency and efficiency. When computing the relative efficiency of the estimator, especially finding the optimal estimator, Fisher information can be used. Fisher information from censored samples is a useful tool for planning life testing experiments and for evaluating the performance of estimators and tests based on censored samples.

Progressive type II censoring model have an important role in the reliability and life tests. Let  $N$  identical units be placed on a lifetime test at time zero. Immediately following the  $i$ th failure,  $R_i$  surviving units are randomly withdrawn from the test,  $1 \leq i \leq n$ . Thus,  $n$  failures are observed then  $R_1 + R_2 + \dots + R_n$  units are progressively censored, since  $N = n + R_1 + \dots + R_n$ . The resulting  $n$  ordered values that are obtained result of this type of censoring are referred to as progressive type II censored order statistics and expressed as  $X_{1:n:N}^{\tilde{R}}, X_{2:n:N}^{\tilde{R}}, \dots, X_{n:n:N}^{\tilde{R}}$ . The censoring scheme is denoted by  $\tilde{R} = (R_1, R_2, \dots, R_n)$ . When  $\tilde{R} = (0, 0, \dots, 0)$ , the progressive type II censoring scheme reduces to the case of no censoring and usual order statistics are obtained.

In our study, we consider Fisher information in progressive type II censored order statistics and their concomitants. Let  $(X_i, Y_i)$ ,  $i = 1, 2, \dots$  be a sequence of independent and identical distributed bivariate random variables  $(X, Y)$  with an absolutely continuous distribution function  $F_{X,Y}(x,y)$ . The  $Y$  value associated with  $X_{r:n}$  is denoted by  $Y_{[r:n]}$ .  $Y_{[r:n]}$  is called the concomitant of the  $r$ th order statistics. They are also known as induced order statistics in the literature. The concomitants are of interest in selection and prediction problems based on the ranks of the  $X$ 's. Concomitants of order statistics used mostly in selection procedures, when  $k (< n)$  individuals are chosen on the basis of their  $X$ -values. Then the  $Y$  values represent performance of an associated characteristic. As an example, if the  $k$  out of  $n$  candidates as judged by their scores in a screening test, then  $Y_{[n-k+1:n]}, \dots, Y_{[n:n]}$  are the scores on a final test. Or, if  $k$  rams from sample size  $n$  is selected for breeding in according to their genetic make up, then the quality of the wool of one of their female offspring might be represented by  $Y_{[n-k+1:n]}, \dots, Y_{[n:n]}$ . For the details of concomitants can be found in Bhattacharya (1984), David (1993), David and Nagaraja (1998).

Let  $(X, Y)$  is absolutely continuous with joint c.d.f.  $F_{X,Y}(x,y; \theta)$  and p.d.f  $f_{X,Y}(x,y; \theta)$ .  $X_{r:n}$  be the  $r$ th order statistic and  $Y_{[r:n]}$  be its concomitant obtained from random sample of size  $n$  from  $f$ , for  $1 \leq r \leq n$ . David and Nagaraja (1998) gives the joint p.d.f. of  $(X_{r:n}; Y_{[r:n]})$  as

$$f^{X_{r:n}, Y_{[r:n]}}(x, y; \theta) = kf(x, y; \theta) [F_1(x; \theta)]^{r-1} [1 - F_1(x; \theta)]^{n-r}$$

where  $k = \frac{n!}{(r-1)!(n-r)!}$ .

In this study, the Fisher information about dependence parameter of a bivariate copula model for progressive type II censored order statistics and their concomitants is given. Also the values of Fisher information for different values of  $N$ ,  $n$  and for different censoring schemes are computed.

BİLDİRİ OTURUMLARI 2

SESSION 2

Statistics Theory 1



## REFERENCES

- [1] Abo Eleneen, Z.A. (2008), Fisher information in Type II Progressive censored samples. Commun. Statist. Theor. Meth. 37:682-691.
- [2] Amini, M., Ahmadi, J. (2007), Fisher information in record values and their concomitants about dependence and correlation parameters. Statistics&Probability Letters 77: 964-972
- [3] Bhattacharya, P.K. (1984), Induced order statistics. Theory and application. In: Krish-naiah, P.R., Sen, P.K. (Eds.), Handbook of Statistics, vol. 4 North-Holland, Amsterdam, pp. 383-403.
- [4] David, H.A. (1993), Concomitants of order statistics: review and recent developments. In: Hoppe, F.M. (Ed.), Multiple Comparisons, Selection and Application in Biometry. Dekker, New York, pp. 507-518.
- [5] David, H.A., Nagaraja, H.N. (1998), Concomitants of order statistics. In: Balakrishnan, N., Rao, C.R., (eds) Handbook of statistics, 16. Elsevier, Amsterdam, pp 487-513.

## FISHER INFORMATION IN PROGRESSIVE TYPE II CENSORED ORDER STATISTICS AND THEIR CONCOMITANTS FOR SOME COPULA MODELS

We study on the Fisher information about the dependence parameters for a bivariate copula of progressive type II censored order statistics and their concomitants. The relative efficiency of the estimator of the dependence parameter for progressive type II censored order statistics and their concomitants is obtained. Also, some numerical results are presented.

**Key Words:** Fisher Information; progressive type II censored order statistic; concomitant; copula

## A NEW RESTRICTED ESTIMATOR

Gülesten ÜSTÜNDAĞ ŞIRAY, Selma TOKER\*

University of Cukurova, Faculty of Science and Letters, Department of Statistics, Adana, Turkey,  
gustundag@cu.edu.tr, stoker@cu.edu.tr

### Modified and Restricted Two Parameter Ridge Estimators

The standard multiple linear regression model is

$$y = X\beta + \varepsilon, \quad \varepsilon \sim (0, \sigma^2 I_n)$$

where  $y$  is an  $n \times 1$  vector of observations on the response variable;  $X$  is a  $n \times p$  model matrix of full column rank of observations on  $p$  non-stochastic explanatory variables;  $\beta$  is a  $p \times 1$  vector of unknown parameters and  $\varepsilon$  is an  $n \times 1$  vector of disturbances. For the standardized variables in a multiple linear regression model  $C$  and  $r$  denote correlation matrix of the regressors and vector of correlations of the dependent variable with  $r = X'y$ ,  $C = X'X$ .

In the presence of the multicollinearity, biased estimators alternative to the ordinary least squares estimator (OLSE) are proposed as a remedy for this problem. The most popular estimator is the ridge regression estimator (RE) which is proposed by Hoerl and Kennard (1970). Although the RE is the most common estimator for dealing with multicollinearity, it has also negative drawbacks. These negative drawbacks led Lipovetsky and Conklin (2005) to introduce the two parameter ridge estimator (TRE).

Another alternative technique to remedy the problem of multicollinearity is specifying restrictions on the parameter space. Suppose that there are  $m$  linear equality restrictions on the parameter vector  $\beta$  as  $H\beta = h$  where the  $H$  is  $m \times p$  and of full row rank  $m < p$ ,  $h$  is an  $m \times 1$  vector and both  $H$  and  $h$  are known.

Swindel (1976) introduced the modified ridge regression estimator by combining the prior information with the ridge regression.

By shrinking towards the shortest vector  $\beta_0 = H'(HH')^{-1}h$  which satisfies the linear restrictions  $H\beta = h$ , Gross (2003) proposed the restricted ridge estimator.

In this paper, by following Gross (2003) we introduce two new estimators, the modified two parameter ridge estimator (MTRE) and restricted two parameter ridge estimator (RTRE) as

$$\hat{\beta}_q(k, b) = q(C + kI)^{-1}(r + \frac{k}{q}b)$$

$$\hat{\beta}_q^r(k) = \hat{\beta}_q(k, \beta_0) - (C + kI)^{-1}H'[H(C + kI)^{-1}H']^{-1}(H\hat{\beta}_q(k, \beta_0) - h).$$

where  $k$  and  $q$  are two new constant parameters,  $b$  is a given non stochastic vector which should be chosen to represent the prior information on  $\beta$  and  $\hat{\beta}_q(k, \beta_0)$  is obtained by writing  $\beta_0$  in the MTRE. It is easily seen that

A NEW APPROACH FOR LEARNING STRUCTURE IN HYBRID BAYESIAN NETWORKS

Emre DÜNDER , Mehmet Ali CENGİZ  
Ondokuz Mayıs University, Faculty of Science,,Departmet of Statistics, Kurupelit/SAMSUN

[emre.dunder@omu.edu.tr](mailto:emre.dunder@omu.edu.tr)

[macengiz@omu.edu.tr](mailto:macengiz@omu.edu.tr)

Bayesian networks are the probabilistic graphical models that represent causal relations in a multivariate dataset. The main task is constructing the graphical model with an automated method. These methods are called as structure learning in Bayesian networks. In recent years many researchers have focused on learning the structure of Bayesian networks. Some real problems are more naturally modeled by hybrid Bayesian networks that consist of mixtures of continuous and discrete variables with their interactions described by equations and continuous probability distributions. When mixed data is used, the continuous part of the data should be normally distributed. The failure of this assumption can occur to the construction of wrong causal relations in Bayesian networks. Because of this limitation the discretization methods are used for rendering continuous variables to the discrete ones. Generally, the continuous variables are discretized to solve the distributional problem for learning the structure. The use of different discretization techniques can be expected to affect the quality of the learned structure. Due to the loss of information, the discretization process causes the errors.

In this paper we introduce a new constraint based approach for learning hybrid Bayesian networks from data. Instead of classical statistical independence tests, generalized linear modeling (GLM) methods are replaced to learn the Bayesian network structure. There is no need to use any discretization method in structure learning processes together with this approach so the loss of information about the data is prevented. The implementations are performed in R project statistical programming language.

**Key Words:** Hybrid Bayesian Networks, Structure Learning, Generalized Linear Modelling

**KAYNAKLAR**

1. Neapolitan R. (2003), Learning Bayesian Networks, Prentice Hall Series in Artificial Intelligence.
2. Faraway J. J. (2006), Extending the Linear Model With R, Boca Raton London New York
3. Scutari M. (2011), Measures of Variability for Graphical Models, Università degli Studi di Padova, Dipartimento di Scienze Statistiche.

$\hat{\beta}_q^r(k)$  includes the RRE for  $q = 1$  and the restricted least squares estimator (RLSE) for  $k = 0$  and  $q = 1$  as special cases.

We obtain a criterion for the choice of  $q$  which makes the coefficient of multiple determination ( $R^2$ ) maximum as

$$q = \frac{r' M_k r + h' A' C M_k r}{r' M_k C M_k r},$$

where  $M_k = (C + kI)^{-1} - (C + kI)^{-1} H' [H(C + kI)^{-1} H']^{-1} H(C + kI)^{-1}$  and  $A = (C + kI)^{-1} H' [H(C + kI)^{-1} H']^{-1}$ .

We find the matrix mean square errors of the MTRE and RTRE as follows:

$$\begin{aligned} \text{MSE}(\hat{\beta}_q(k, b)) &= q^2 \sigma^2 (C + kI)^{-1} C (C + kI)^{-1} \\ &+ [(q(C + kI)^{-1} C - I)\beta + (C + kI)^{-1} kb]' \\ &\times [(q(C + kI)^{-1} C - I)\beta + (C + kI)^{-1} kb] \end{aligned}$$

$$\begin{aligned} \text{MSE}(\hat{\beta}_q^r(k)) &= q^2 \sigma^2 M_k C M_k \\ &+ [(q M_k C - I)\beta + (I - M_k C)\beta_0]' [(q M_k C - I)\beta + (I - M_k C)\beta_0]. \end{aligned}$$

The formula of  $\text{MSE}(\hat{\beta}_q^r(k))$  is so complicated therefore the comparison of this estimator with the other estimators can not be easily seen theoretically. So, we examine the theoretical results with a data set about Total National Research and Development Expenditures (Gruber, 1998). It is emphasized that specifying restrictions on parameters decreases the estimated scalar mean square error values of the OLSE, RE and TRE. Besides this, the RTRE outperforms the RRE and RLSE under given restrictions and prior information.

**KAYNAKLAR**

- [1] Hoerl, A. E., Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. Technometrics, 12, 55-67.
- [2] Gross, E. (2003). Restricted Ridge Estimation. Statistics and Probability Letters, 65, 57-64.
- [3] Lipovetsky, S. and Conklin, W. M., (2005). Ridge Regression in Two Parameter Solution. Applied Stochastic Models in Business and Industry, 21: 525-540.
- [4] Swindel, B. F., (1976). Good estimators based on prior information, Communications in Statistic, 5, 1065-1075.
- [5] Gruber, M.H.J. (1998). Improving Efficiency by Shrinkage. Marcel Dekker, Inc, New York.

**Key Words:** restricted two parameter ridge estimator, modified two parameter ridge estimator, two parameter ridge estimator, restricted ridge estimator, ridge estimator

UNCERTAINTY ASSESSMENT OF THE RESPONSE SURFACE MODEL PARAMETERS WITH  
BAYESIAN AND FUZZY APPROACHES

Özlem TÜRKŞEN

Ankara University, Faculty of Science, Statistics Department, 06100, Tandoğan, Ankara, Turkey  
turksen@ankara.edu.tr

**Introduction**

It is well-known that the first stage for the solution of response surface problems is modeling of the unknown response. Regression analysis has been used as a powerful tool for investigating the relationship between response and input variables. This relationship is defined by the unknown model parameters which are uncertain and specific to the problem. In this study, the uncertainty of response surface model parameters are evaluated by using Bayesian and fuzzy approaches. The probability simply describes the uncertainty in the Bayesian framework while the possibility is used for representing uncertainty in the fuzzy framework. HPD credible interval and alpha-cut values are used to analyse the uncertainty of model parameters for Bayesian and fuzzy approaches, respectively.

**Bayesian Modeling Approach**

A key feature of Bayesian statistics is combining two separate sources of information, prior information and data. The result is posterior distribution. Linear regression methods can be thought of as Bayesian posterior inference based on a non-informative prior distribution for the parameters of the normal linear model. In linear regression model, the observations consist of a response variable in a vector  $\mathbf{Y}$  ( $n \times 1$ ) and one or more input variables in a matrix  $\mathbf{X}$  ( $n \times p$ ). The parameters, regression coefficients and error variance of fitted model, are denoted as  $\boldsymbol{\beta}$  ( $p \times 1$ ) and  $\sigma^2$  ( $> 0$ ). The model that relates observations and parameters is written as

$$\mathbf{Y} | \boldsymbol{\beta}, \sigma^2, \mathbf{X} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n). \quad (1)$$

Once the model is specified, the Bayesian analysis seeks the posterior distribution for the parameters. A non-informative prior distribution, commonly used for linear regression, can be defined as

$$h(\boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-1}, \quad \boldsymbol{\beta} \in R^p, \quad \sigma^2 > 0. \quad (2)$$

The posterior distribution of  $\boldsymbol{\beta}$  given  $\sigma^2$  is

$$\boldsymbol{\beta} | \sigma^2, \mathbf{Y} \sim N_p(\hat{\boldsymbol{\beta}}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}) \quad (3)$$

in which  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$ . The marginal posterior distribution of  $\sigma^2$  is  $\sigma^2 | \mathbf{Y} \sim GI\left(\frac{k}{2}, \frac{kS^2}{2}\right)$

and the marginal posterior distribution of  $\boldsymbol{\beta}$  given  $\mathbf{Y}$  is obtained as  $\boldsymbol{\beta} | \mathbf{Y} \sim t_p\left(k; \hat{\boldsymbol{\beta}}, S^2 (\mathbf{X}'\mathbf{X})^{-1}\right)$  where  $k = n - p$  and  $S^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})/k$ . In this study, the Bayesian confidence interval (HPD credible interval), which provides an actual probability statement based only on the observed data, is used to analyse the uncertainty of the model parameters. The  $100\gamma\%$  HPD credible interval, for each  $\beta_j$ ,  $j = 1, 2, \dots, p$ , is defined as

$$\hat{\beta}_j \pm mS\sqrt{d_j} F_{t(k)}^{-1}\left(\frac{1-\gamma}{2}\right), \quad j = 1, 2, \dots, p \quad (4)$$

in which  $d_j$  is a diagonal element of  $(\mathbf{X}'\mathbf{X})^{-1}$  and  $F_{t(k)}^{-1}$  is a quantile function.

**ACKNOWLEDGEMENTS**

The author thanks to Prof. Dr. Ayşen Apaydın for her informative guidance about fuzzy approach. The author also thanks to Prof. Dr. Maria Antonia Amaral Turkman for her valuable contributions and helpful suggestions which led to improvements on Bayesian approach.

**REFERENCES**

- [1] Carlin B.P. and Louis T.A. (2009), Bayesian Methods for Data Analysis, Third edition, Chapman and Hall.
- [2] Gelman A., Carlin J.B., Stern H.S. and Rubin D.B. (2003), Bayesian Data Analysis, Second edition, Chapman and Hall.
- [3] Paulino C.D., Turkman, M.A.A. and Murteira, B. (2003), Estatística Bayesiana, Fundação Galouste Gulbenkian.
- [4] Türkşen, Ö. (2011), Fuzzy and Heuristic Approach to the Solution of Multi Response Surface Problems, PhD Thesis, Ankara.

**ABSTRACT**

**UNCERTAINTY ASSESSMENT OF THE RESPONSE SURFACE MODEL PARAMETERS WITH BAYESIAN AND FUZZY APPROACHES**

Modeling of the unknown response is a crucial step to get knowledge about the relationship between the response and input variables. The predicted response model involves many parameters which are uncertain simply because of the lack of knowledge. In this study, Bayesian and fuzzy approaches are used to analyse the uncertainty of the response surface model parameters. In Bayesian approach, the uncertain quantity of parameters are referred as random variable. Besides, the uncertainty of the parameters is represented as a fuzzy number in fuzzy approach. The interested approaches are illustrated on a multi response experimental data set from the literature.

**Key Words:** Bayesian regression, fuzzy regression, uncertainty of parameters

OPTİMAL REPLACEMENT MODEL FOR SYSTEMS WITH SINGLE SPARE AND LEAD TIME

Ayşe TANSU

Cyprus International University, Engineering Faculty, Industrial Engineering Department, Nicosia, Cyprus

[ayset@ciu.edu.tr](mailto:ayset@ciu.edu.tr)

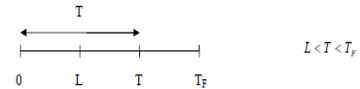
1.Introduction

In the replacement model of [1] the system was replaced at periodic instants of time  $T, 2T, \dots$ . These periodic replacements imply the availability of the replacement system. In practice this implication poses difficulties. For instance when the system is expensive like military hardware or unwieldy like a generator in a gas turbine, the assumption of a continuous supply of replacements is suspect. It is reasonable in commercial industries that only one spare unit which can be delivered upon order is available for replacement. Also in each of the above cases a random lead time for delivering the spare unit could not be ignored. Thus it becomes essential and more practical to analyze the effects of single spare and its lead time for delivery on the replacement times. Nakagawa and Osaki [2] developed an age replacement policy with random lead time. Osaki and Yamada [3] discussed an age-replacement policy with random lead time and with cost for stock and system downtimes. Sheu and Griffith [4] considered a general age-replacement policy with age dependent minimal repair and random lead time. Keywords: optimal replacement, spare, lead time.

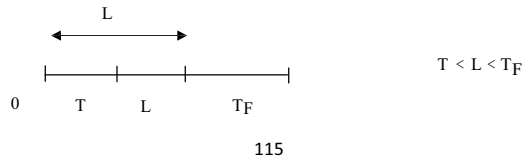
2. Model

A replacement model for a system based on the frequency of shocks with a single spare is developed. The delivery of the spare is subject to a random lead time. Using the long run average cost per unit time the optimal replacement time  $T^*$  is obtained explicitly. The model is governed by the following assumptions: Assumption1. At the beginning, the system is put into operation which is conveniently taken to be the time origin  $t=0$ . The system is replaced whenever it reaches the age  $T$  by a spare if available. The spare unit is assumed to be as good as new. Assumption2. The system is subjected to shocks and frequency of shocks dependent failures so that the distribution function of the failure time  $T_F$  is  $K(t)$ . Assumption3. An order for a spare unit is placed at each of the instant of the installation of a new/spare unit. Assumption4. The lead time  $L$  for the delivery of the spare unit is a random variable with distribution function  $H(t)$  having a finite mean. Assumption5. The shock arrival times, the threshold times and the lead time are all mutually independent random variables. From the foregoing assumptions it is clear that the time between two successive system replacements forms a cycle. In computing the length of a cycle and the cost involved in a cycle we have to consider four mutually exclusive cases which are listed below:

Case 1: If the ordered spare arrives and no failures occur before time  $T$ , then the delivered spare unit is put into stock until  $T$  when the working unit is replaced by the spare. In this case the length of the cycle is  $T$  and the relevant costs are  $\eta_1$ , the cost of replacement and  $c_s$  the cost rate for stocking the spare unit for a period  $(T - L)$ .

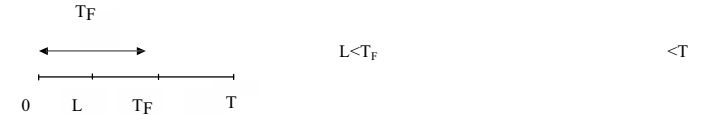


Case 2:



If the ordered spare arrives after time  $T$  and no failure occurs before the arrival of the ordered spare unit, then the working unit is replaced by the spare as soon as it is delivered. In this case the length of the cycle is  $L$  and the relevant cost is the cost of replacement  $\eta_2$

Case 3:



If the ordered spare arrives before failure and the failure occurs before the replacement time  $T$ , then the delivered unit is put into stock and the working unit is replaced by the spare upon failure. In this case the length of the cycle is failure time  $T_F$  and the relevant costs are the cost of replacement  $\eta_3$  and the cost rate for stocking the spare unit  $c_s$  for a period  $(T_F - L)$ .

Case 4:



If the ordered spare arrives after failure, then the system is shut down and replaced by the spare as soon as the spare is delivered. In this case the length of the cycle  $L$  and the relevant costs are the cost of replacement  $\eta_4$  and the cost resulting from system downtime  $c_p$  for a period  $(L - T_F)$ .

[1]A.Rangan and A.Tansu (2010), *Some Results on a New Class of Shock Models*, Asia-Pacific Journal of Operational Research, Vol. 27 (4).

[2] Nakagawa, T., & Osaki, S. (1974), *Optimum replacement policies with delay*, Journal of Applied Probability, 11, 102-110.

[3] Osaki, S., & Yamada, S. (1976), *Age replacement with lead time*, IEEE Transactions on Reliability, 25, 344-345.

[4] Sheu, S.H., & Griffith, W.S. (2001). *Optimal age-replacement policy with age-dependent minimal-repair and random-leadtime*, IEEE Transactions on Reliability, 50, 302-309.

## BİLDİRİ OTURUMLARI 2

### SESSION 2

#### Çok Değişkenli İstatistik

## AÇIMLAYICI FAKTÖR ANALİZİNDE FAKTÖR ÇIKARMA METODUNUN SEÇİMİNE İLİŞKİN BİR SİMÜLASYON ÇALIŞMASI

Algin OKURSOY\*

Kıvanç YÜKSEL

ADÜ Söke İşletme Fakültesi Yönetim Bilişim  
Sistemleri Bölümü  
09200 Aydın / Türkiye  
aokursoy@adu.edu.tr

EÜ İlaç Geliştirme Ve Farmakokinetik Araştırma-  
Uygulama Merkezi  
35100 İzmir / Türkiye  
kianc.yuksel@ege.edu.tr

### Giriş

Araştırmacının, ölçme aracının ölçtüğü faktörlerin sayısı hakkında bir bilgisinin olmadığı durumlarda ve belli bir hipotezi sınamak yerine, ölçme aracıyla ölçülen faktörlerin doğası hakkında bir bilgi edinmeye çalıştığı inceleme türlerine açımlayıcı faktör analizi (exploratory factor analysis) denir (Tavşancıl,2010, 46). Açımlayıcı faktör analizi bir veri azaltma tekniğinden ibarettir. Açımlayıcı faktör analizinde, maddelerin ilgili faktörler altında çıkması ve yüksek faktör yük değerleri ile elde edilmesi arzu edilir. Maddelerin gerçekte hangi faktör altında ölçme yaptıkları bilinmediği durumlarda bu analiz yöntemi keşfedici olarak kullanılabilir (Çokluk vd., 2010, 189).

Bu çalışmada, açımlayıcı faktör analizinin uygulanmasının adımları ayrıntılı bir şekilde ele alınacak ve her bir adım için uygulamada karşılaşılan problemlere ilişkin öneriler getirilecektir. Ayrıca faktörlerin elde edilmesinde kullanılan yöntemler olan temel bileşenler, temel eksenler ve en çok olabilirlik metodları verilerin normal dağılıma sahip olduğu ve normal dağılımdan sapma gösterdiği durumlar için simülasyon yöntemi kullanılarak incelenecektir. Hangi metodun hangi durumda kullanılmasının faydalı olacağı araştırılacaktır.

### Faktör Analizinde Örneklem Büyüklüğü, Çoklu Bağlantı ve Normallik

Veri setinin faktör analizine uygunluğunun sınanması için örneklem büyüklüğü, normallik, doğrusallık, çoklu bağlantı ve tekliğin kontrol edilmesi analizin doğru sonuçlar vermesine yardımcı olacaktır. Örneklem genişliğinin ölçme aracında yer alan değişken sayısına göre belirlendiği durumlar faktör analizinde gözlem sayısının belirlenmesinde daha sıklıkla başvurulan bir yöntemdir. Genel kural olarak her değişken için 10 gözlem alınması önerilmektedir (Velicer ve Fava, 1998, 232). Faktör analizinde normal dağılım varsayımının yerine getirilemediği durumlarda ise çözüm alternatifleri azaltmakla birlikte veri setinden yararlı sonuçlar elde edilebilir (Çokluk vd., 2010, 208). Eğer faktör analizi veri seti içerisindeki faktör sayısını belirlemek ve değişkenler arasındaki ilişkileri özetlemek için kullanılıyorsa değişkenlerin normal dağılıma sahip olma şartı aranmayabilir (Tabachnick ve Fidell, 2007, 613). Faktör analizi sürecinde değişkenler arasındaki doğrusallık analizin sonuçlarının güvenilir olduğuna işaret eder. İki değişkene ait ilişkinin doğrusal olduğu hesaplanan korelasyon katsayısının kritik değer ile karşılaştırılmasıyla sınanabilir. Kritik değerlerin hesaplanmasında aşağıdaki formülden yararlanılabilir (Spiegel ve Stephens, 1999, 317,336).

$$r^* = \frac{t}{\sqrt{n-2+t^2}} \quad (1)$$

Çoklu bağlantı sorununun var olup olmadığını anlamaya yönelik olarak kullanılan yöntemlerden bir tanesi olan Koşul Endeksidir (CI – Condition Index). Veri matrisi içerisindeki doğrusal bağlılığı tespit etmeye yönelik olarak kullanılan bir yöntemdir (Wissmann, Toutenburg ve Shalabh, 2007, 3). Koşul Endeksi (k), en büyük özdeğerin en küçük öz değere bölümünün karekökü olarak tanımlanabilir (Rao et al., 2008, 97; Gujarati, 2004, 362). Eğer k > 30 ise şiddetli bir çoklu bağlantıdan bahsedilebilir.

## Faktör Çıkarma Metodunun Seçimi

Faktör analizinde değişkenlerin normal dağılıma sahip olduğu durumlarda göre bu iki faktör çıkarma metodunun yerine en çok olabilirlik metodunun önerilmiştir. Fabrigar, Wegener, MacCallum ve Strahan (1999, 277), faktör analizi sonrasında istatistiksel hipotez testleri uygulanacak ve güven aralıkları tahmin edilecekse faktör çıkarma metodu olarak En Çok Olabilirlik yönteminin (Maximum Likelihood) kullanılmasının daha doğru olacağını belirtmişlerdir. Alan yazında, normal dağılım varsayımı sağlamadığı durumlarda ise, faktör elde edilmesinde PAF metodunun kullanılmasının daha doğru olacağı vurgulanmıştır (Fabrigar et al., 1999, 283; Costello ve Osborne, 2009, 134). PAF metodu PCA metoduna göre popülasyon yüklerinin elde edilmesinde daha uygun sonuçlar vermektedir (Widaman, 1993, 282).

Çalışma verileri SAS programında IML dili ve Factor prosedürü kullanılarak türetilmiştir. Çok değişkenli normal dağılıma  $N(\mu_i, \sigma^2_i)$  uygun olarak oluşturulan değişkenler hedeflenen popülasyon korelasyon matrisine (R) bağlı olarak rastgele türetilmiştir. Verilerin türetilmesinde temel matris ayırma yöntemi (basic matrix decomposition procedure) kullanılmıştır (Kaiser, H. F., and K. Dickman, 1962). Normal olmayan çok değişkenli veriler ise önceden belirlenmiş çarpıklık ve kuyruk kalınlığı değerlerine bağlı olarak edilmiş çok değişkenli normal dağılıma ait değişkenlerin dönüştürülmesi ile elde edilmiştir (Vale and Maurelli, 1983).

## KAYNAKLAR

- [1] ÇOKLUK, Ö., ŞEKERCİOĞLU, G., BÜYÜKÖZTÜRK, Ş. (2010). Sosyal Bilimler İçin Çok Değişkenli İstatistik : SPSS ve LISREL Uygulamaları. Ankara: Pegem Akademi.
- [2] FABRİGAR, L. R., WEGENER, D. T., MACCALLUM, R. C., & STRAHAN, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299.ss.
- [3] Kaiser, H. F., and K. Dickman. 1962. "Sample and Population Score Matrices and Sample Correlation Matrices from an Arbitrary Population Correlation Matrix." *Psychometrika* 27:179-182.

## ABSTRACT

### A SIMULATION STUDY FOR SELECTION OF FACTOR EXTRACTION METHODS IN EXPLORATORY FACTOR ANALYSIS

**Key Words:** Exploratory factor analysis, Factor Extraction Method

The exploratory factor analysis is one of the data reduction techniques. In the exploratory factor analysis, it is desired that items are appeared under the relevant factors and they have high factor loading values. In factor analysis, sample size, normality and multicollinearity affect the results of analysis. In this paper, steps of the exploratory factor analysis are detailed and suggestion will be given for encountering possible problems. Moreover, Normality effects will be discussed for factor extraction method such as Principal Component Analysis, Principal Axis Factoring and maximum likelihood by simulation study. Simulation will be run for normal distribution and non-normal distribution.

## MANOVA VARSAYIMLARININ SAĞLANMAMASI DURUMUNDA

### UZAKLIK ANALİZİ YÖNTEMİNİN KULLANIMI

B. Barış ALKAN<sup>1</sup>, Cemal ATAKAN<sup>2</sup>

<sup>1</sup>Sinop Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Sinop

E-mail:bbalkan@sinop.edu.tr

<sup>2</sup>Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara

Email:atakan@science.ankara.edu.tr

## 1. Giriş

Veri kümesinde heterojenlik, çarpıklık, kayıp veri gibi problemler söz konusuysa veya değişkenlerin sayısı gözlemlerin sayısından daha büyük ise, çok değişkenli tek yönlü varyans analizi (MANOVA) gibi geleneksel çok değişkenli teknikler, gruplar arasındaki farklılıklar belirleneceği zaman yüksek derecede kuşku sonuçlar verebilmektedir. Böyle durumlarda tek yönlü MANOVA'ya alternatif bir çıkarımsal metod olarak Gower ve Krzonowski (1999) tarafından önerilen ve uzaklık ölçülerine dayanan Uzaklık Analizi (UA) yönteminin kullanımı araştırmacıların daha doğru ve güvenli bulgulara ulaşmalarını sağlayacaktır. Gower ve Krzonowski (1999) çalışmalarında Temel Koordinat Analizi (TKA) ve Uzaklık Analizi (UA) olarak adlandırılan iki yaklaşımı ele almışlardır. TKA, Temel Bileşenler Analizine (TBA) paralel bir yaklaşım olup, analiz için gözlem çiftleri arasındaki uzaklıklardan oluşan matrisi kullanır. UA ise MANOVA'ya paralel bir yaklaşım olup, kareler toplamlarından daha çok gözlemler arasındaki uzaklıkların parçalanması temeline dayanır (Fenty 2004).

## 2. Uzaklık Analizi

n tane birimin g tane grup içine  $n_1, n_2, \dots, n_g$  büyüklüklerinde bölündüğünü varsayalım. Gruplandırma  $n \times g$  boyutlu  $\mathbf{G}$  matrisi ile gösterilebilir. i birimi r grubunda yer alıyorsa  $g_{ir} = 1$ , diğer durumda  $g_{ir} = 0$  biçimindedir.  $\mathbf{N} = \text{diag}(n_1, \dots, n_g)$ ,  $n_1, n_2, \dots, n_g$  grup birim sayılarından oluşan köşegen bir matristir.

$\mathbf{m} = (n_1, \dots, n_g)$  ise, grup birim sayılarını içeren bir vektördür.

Grup ortalamalarının koordinatları ise,

$$\bar{\mathbf{X}}_{g \times p} = \mathbf{N}^{-1} (\mathbf{G}_{n \times g})^T \mathbf{X}_{n \times p} \quad (1)$$

eşitliği ile elde edilir.

## 3. Uzaklık Analizi ile Biplot Kullanımı

Uzaklık analizi ile Biplot birlikte oluşturulabilir. UA Biplot, değişken sayısı, gözlem sayısından fazla ise etkileyici boyut indirgeme özelliklerine sahiptir (Gardner vd. 2009). Büyük veri kümelerinin analizlerinde MANOVA varsayımlarının sağlanması zor olduğundan, UA kullanımı bu tür veri kümeleri için önemlidir. UA, grup ortalamaları arasındaki uzaklıkları maksimize eder. Grup yapısının grafiksel olarak incelenmesi için UA Biplot kullanılır.

#### 4. Uygulama

Gerrild ve Lantz (1969) tarafından elde edilen ham petrol örneklerini içeren veri kümesi, üç kumtaşı bölgesinden (Bölge I: Wilhelm, Bölge II: Aşagi-Mulinia, Bölge III: Yukarı-Mulinia) alınan 56 ham petrol örneklerinin vanadyum, demir, berilyum, doymuş hidrokarbonlar, aromatik hidrokarbonlar kimyasal özelliklerini içermektedir. Veri kümesine R programında yer alan kütüphane ve fonksiyonlar yardımıyla oluşturulan program ile UA Biplot uygulanmış elde edilen bulgular yorumlanmıştır.

#### 5. Sonuç

Bu çalışmada, Temel Koordinat Analizi (TKA), Uzaklık Analizi (UA) ve Biplot yöntemleri incelenmiş ve gerçek bir veri kümesi üzerinden MANOVA varsayımlarının sağlanmaması durumunda uzaklık analizi kullanımının önemi vurgulanmıştır. Çalışmada ayrıca varsayımların sağlanmaması durumunda MANOVA'ya alternatif kullanılabilen diğer yaklaşımlar da kısaca ele alınmış ve uzaklık analizine göre avantaj ve dezavantajları tartışılmıştır.

#### KAYNAKLAR

- [1] Fenty J. (2004), Analyzing distances, The Stata Journal, Vol. 4(1), pp. 1-26.
- [2] Gardner-Lubbe S., Roux N.J.L., Maunders H., Shah V. and Patwardhan S. (2009), Biplot methodology in exploratory analysis of microarray data, Statistical Analysis and Data Mining, Vol. 2(2), pp. 135-145.
- [3] Gerrild P.M. and Lantz R.J. (1969), Chemical Analysis of 75 Crude Oil Samples from Pliocene Sand Units, ELK Hills Oil Field., U.S. Geological Survey Open-File Report, California.
- [4] Gower J.C. and Krzanowski W.J. (1999), Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance, Applied Statistics, Vol. 48, pp. 505-519.

#### ABSTRACT

#### THE USE OF DISTANCE ANALYSIS WHEN ASSUMPTIONS OF MANOVA ARE VIOLATED

Multivariate analysis of variance (MANOVA) is used for testing the null hypothesis of equal mean vectors for  $g$  groups. MANOVA assumes homogeneity of variance-covariance matrices and multivariate normality. MANOVA can not be used when its assumptions are violated. In this study, we used analysis of distance biplot as an alternative method of MANOVA. As a real data application the mean vectors of the crude oil samples dataset for tree regions are examined.

**Key Words:** Analysis of distance, biplot, multivariate analysis of variance, principal coordinates Analysis

#### KANONİK KORELASYON ANALİZİNDE BOYUT SAYISINA KARAR VERİLMESİNDE PARALEL ANALİZ YAKLAŞIMI

Gülhayat GÖLBAŞI ŞİMŞEK\*, Fatma NOYAN, Selahattin AYDOĞDU  
Yıldız Teknik Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, 34220, İstanbul, TÜRKİYE  
gulhayat@yildiz.edu.tr, fnoyan@yildiz.edu.tr, saydogdu23@gmail.com

#### Giriş

Hotelling (1936) tarafından geliştirilmiş olan Kanonik Korelasyon Analizi (KKA) iki değişken seti arasındaki ilişkileri analiz etmek için kullanılan çok değişkenli istatistik yöntemlerden biridir. KKA her bir setteki orijinal değişkenlerden, aynı set içinde korelasyonsuz olacak ve setler arasındaki korelasyonu maksimum yapacak şekilde, orijinal değişkenlerin lineer kombinasyonları olan yeni değişkenleri bulmaya çalışır. Bu yeni değişken çiftlerine kanonik değişkenler ve kanonik değişkenler arasındaki korelasyona da kanonik korelasyon denir[1].

$p$  sayıda değişkenli  $y$  ve  $q$  ( $p \leq q$ ) sayıda değişkenli  $x$  değişken seti sırasıyla  $\mu_y$  ve  $\mu_x$  ortalamalı ve

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \quad (1)$$

kovaryans matrisli  $(p+q)$  değişkenli normal dağılımdan gelmek üzere,  $y$  ve  $x$  değişken setleri arasındaki kanonik korelasyonlar  $1 \geq \rho_1 \geq \dots \geq \rho_p \geq 0$  olup,  $\Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}$ 'nin özdeğerlerinin pozitif karekökleridir. KKA'da sıfır olmayan kanonik korelasyon sayısı boyut sayısı olup, aynı zamanda üzerinde çalışılacak kanonik değişken sayısıdır. Boyut sayısına karar vermek için öncelikle  $\Sigma_{12} = 0$  sıfır hipotezi,

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_p = 0 \quad (2)$$

test edilir.  $H_0$  reddedilirse, hangi kanonik değişkenlerin sıfır olduklarının belirlenmesi için,  $H_0$ 'ı takip eden hipotez

$$H_d : \rho_{d+1} = \rho_{d+2} = \dots = \rho_p = 0, \quad d = 1, 2, \dots, p-1 \quad (3)$$

test edilir.  $H_d$  hipotezlerinin testinde Lawley-Hotelling iz istatistiği, Wilks'in  $\lambda_d$  olabilirlik oran istatistiği ve Bartlett-Nanda-Pillai iz istatistiği gibi MANOVA'da kullanılan istatistikler önerilmiş olup, Bartlett veya Lawley tarafından modifiye edilen Wilks'in  $\lambda_d$  istatistiği yaygın olarak kullanılmaktadır[2].

Paralel Analiz (PA) faktör veya temel bileşen sayısına karar vermek için kullanılan bir tekniktir. İlk olarak Horn (1965) tarafından önerilmiş olan bu teknikte  $n \times p$  orijinal veri setiyle aynı sayıda değişken ve gözlem içerecek şekilde birim kovaryans matrisli çok değişkenli normal dağılımdan  $n \times p$  boyutlu rastgele veri setleri üretilmiştir

sonra, orijinal veri setine ait korelasyon matrisinden elde edilen özdeğerler ile, rastgele veri setlerinden elde edilen korelasyon matrislerinin özdeğerleri karşılaştırılır[3]. Uygulamada genellikle orijinal veri setinden elde edilen i.özdeğer, rastgele veri setlerinden üretilen i.özdeğerlerin %95, %99 gibi istenilen belli bir yüzdesinden veya kısaca 95. veya 99. persentilden büyük olduğunda i.temel bileşen veya faktörün analize katılmasına karar verilir[4].

Bu çalışmada PA metodu KKA'ya modifiye edilerek, KKA'daki boyut sayısına karar vermek için PA'nın kullanılabilirliği araştırılmaktadır. Bu amaçla sıfırdan farklı iki kanonik korelasyona sahip

$(\rho_1 = .9, \rho_2 = .6, \rho_3 = \rho_4 = \rho_5 = 0)$ ,  $p=7$  ve  $q=5$  olan varsayımsal bir anaktüleden farklı büyüklükte örnekler çekilerek, klasik hipotez testlerinden ve PA'dan bulunan boyut sayıları karşılaştırılmıştır.

#### KAYNAKLAR

- [1] Hotelling H. (1936). Relations between two sets of variates. Biometrika, 28:321-377.
- [2] Caliński T., Krzyśko M. and Wotyński W. (2006). A comparison of some tests for determining the number of nonzero canonical correlations. Communication in Statistics-Simulation and Computation, 35:727-749.
- [3] Horn J. L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30:179-185.
- [4] O'Connor B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behavior Research Methods, Instruments, & Computers, 32(3): 396-402.

#### ABSTRACT

##### PARALLEL ANALYSIS APPROACH TO DETERMINE DIMENSIONALITY IN CANONICAL CORRELATION ANALYSIS

Canonical Correlation Analysis (CCA) considers the relationships between two sets of variables, maximizing the correlation between two new sets which are linear transformations of the original variables from both sets and not correlated within the sets. The correlation coefficients between new sets are called canonical correlations. The number of nonzero canonical correlations is called dimensionality in CCA. Parallel Analysis (PA) is a simulation based method for determining the number of factors or principal components to retain. In this study, PA method is adapted to CCA in order to investigate the applicability of PA as a dimensionality determination method in CCA.

**Key Words:** Canonical correlations, dimensionality, parallel analysis

#### ÇOK DEĞİŞKENLİ ARŞİMEDYEN KAPULALARDA ÜRETİCİ FONKSİYONUN VE AİLE PARAMETRESİNİN TAHMİN EDİLMESİ

Dr. Sıddık ARSLAN<sup>a</sup> Prof. Dr. Fikri ÖZTÜRK<sup>b</sup> Prof. Dr. Salih ÇELEBİOĞLU<sup>c</sup>

<sup>a</sup>: Gazi Üniversitesi Bankacılık ve Sigortacılık Yüksekokulu Bankacılık Bölümü ANKARA  
sarslan@gazi.edu.tr

<sup>b</sup>: Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü ANKARA

<sup>c</sup>: Gazi Üniversitesi Fen Fakültesi Bölümü ANKARA

#### Özet

Latince bir kelime olan copula Türkçe yazında kapula olarak kullanılmaktadır. Kapula parçalar arasında bağlantı veya ilişki kurma anlamına gelmektedir. İstatistikte kapula terimi bir rastgele değişken vektörünün ortak dağılım fonksiyonu ile bu dağılımın marjinalleri arasında bağlantı kuran çok değişkenli "özal" bir fonksiyon anlamında kullanılmaktadır.

Tanım kümesi  $I^n$ , değer kümesi  $I$  olan

$$C : I^n \rightarrow I \quad (1)$$
$$\mathbf{u} \rightarrow C(\mathbf{u})$$

fonksiyonu için; temelli fonksiyon, b)  $\mathbf{a} \leq \mathbf{b}$  olan  $\forall \mathbf{a}, \mathbf{b} \in I^n$  için  $V_H([\mathbf{a}, \mathbf{b}]) \geq 0$  ve  $\mathbf{u}$ 'nun bir tane  $u_k$  hariç tüm bileşenleri 1 ise  $C(\mathbf{u}) = u_k$  koşulları sağlandığında  $C$ 'ye  $n$  boyutlu kapula denir.

$\psi : [0, \infty) \rightarrow [0, 1]$ , sürekli kesin azalan ve konveks fonksiyon ve  $\psi^{-1}(t) = \phi(t) = \inf \{x \in R : \psi(x) \geq t\}$  olmak üzere,

$$C_{\psi, d}(u_1, u_2, \dots, u_d) = \psi\{\psi^{-1}(u_1) + \psi^{-1}(u_2) + \dots + \psi^{-1}(u_d)\} \quad (2)$$

biçiminde yazılan kapulaya Arşimedyen kapula ve  $\psi$  fonksiyonuna kapulanın üretici fonksiyonu denir

Arşimedyen kapulayı kullanışlı kılan özellik, çok değişkenli bir dağılım fonksiyonunu (kapulanın kendisi) tek değişkenli bir fonksiyon (üretici fonksiyon) yardımıyla ifade etmesidir.

Bu çalışmanın birinci amacı, ortak dağılım fonksiyonu  $H$  olan  $\mathbf{X} = (X_1, \dots, X_d)$  rastgele vektörünün, dayandığı kapulanın Arşimedyen olduğu varsayımı altında, üretici fonksiyonun bir tahmin edicisinin elde edilmesidir. Çünkü Arşimedyen kapulayı tahmin etme ile üretici fonksiyonu tahmin etme eşdeğer problemlerdir. İkinci amaç ise ortak dağılım fonksiyonunun kapulasının, Arşimedyen kapulaların belirli bir parametrik ailesi olduğu varsayımı altında aile parametresinin tahmin edicilerinin elde edilmesidir.

Bu çalışmada "istatistiksel sonuç çıkarımı" kavramı, Arşimedyen kapulalarda üretici fonksiyonun tahmin edilmesi olarak ele alınmaktadır. Yani  $\mathbf{X} = (X_1, \dots, X_d)$  rastgele vektörünün dayandığı kapulanın Arşimedyen olduğu varsayımı altında, üretici fonksiyonun parametrik olmayan bir tahmin edicisinin elde edilmesidir.

Çok boyutlu Arşimedyen kapulalarda sonuç çıkarımı, üretici fonksiyon için, üretici fonksiyonun tam  $d$  - monoton olmasına bağlı olarak radyal dağılım üzerinden Williamson  $d$  - dönüşümü aracılığı ile yapılmaktadır.

Bir rastgele vektörün dayandığı kapulanın Arşimedyen olduğu varsayımında bulunmak, istatistiksel sonuç çıkarımı açısından önemlidir. Arşimedyen kapulaların değişme özelliğinin bir sonucu olarak değişkenler arasındaki ikiye uyum ölçüsü katsayılarının birbirine yakın olması, bu değişkenlerin kapulasının Arşimedyen olduğu varsayımı için bir kriter olarak kullanılabilir. Uygulama verilerinin seçiminde bu durum dikkate alınmıştır.



Veri kümesinin ait olduğu Arşimedyen kapula ailesine karar verilirken, aile parametresinin Kendall'ın  $\tau$ 'suna dayalı tahmin edicisi kullanılarak parametre tahmini yapılır. Bu tahmin değeri kullanılarak Kendall dağılımına dayalı Kolmogorov-Smirnov uyum iyiliği testi kullanılarak aileye karar verilmektedir. Uygulama olarak, Türkiye'deki beş bankanın 25.09.2009-15.10.2012 tarihleri arasındaki günlük getirileri kullanılmıştır.

## KAYNAKLAR

- Arslan, A. 2013. Arşimedyen Kapulalar Üzerine Bir Çalışma. Doktora tezi (basılmamış). Ankara Üniversitesi 111s, Ankara.
- Genest, C., Neslehova J. and Ziegel J. 2011a. Inference in Multivariate Archimedean Copula Models Test, 20, 223-256.
- McNeil, A. J. and Neslehova J. 2009. Multivariate Archimedean Copulas, D- Monotone Functions and  $I_1$  - Norm Symmetric Distributions. The Annals of Statistics, 37, 3059-3097.
- Nelsen, R. B., 2006. An Introduction to Copulas. Springer, 270 s., New York.

## ABSTRACT

### ESTIMATING GENERATING FUNCTION AND FAMILY PARAMETERS IN MULTIVARIATE ARCHIMEDEAN COPULAS

Statistical inferencing on Archimedean copulas estimation is the estimation of generating function of Archimedean copulas itself. In this study, it is focused especially on the inference for the multivariate Archimedean copulas. Four different estimators, the estimator for the generator function, the estimator for the family parameter based on the Kendall's tau, the OLS estimator based on the generator function and the OLS estimator based on the Kendall's distribution are discussed. The estimators developed in this study are compared by a simulation study through the three copula family, namely, Clayton, Frank and a new Archimedean copula family proposed by Arslan (2013). As a case study, the generator functions and parameters of the considered three families were estimated for the daily stock market returns data set.

**Key Words:** Copulas, Archimedean copulas, generator function, coefficient of agreement, concordance, Kendall distribution, Kendall's  $\tau$ , radial distribution

## ÇOK DEĞİŞKENLİ SIRA İSTATİSTİKLERİ ÜZERİNE

Sinan ÇALIK  
Fırat Üniversitesi Fen Fakültesi İstatistik Bölümü 23119 Elazığ/ TÜRKİYE  
scalik@firat.edu.tr

## 1.Giriş

Çok değişkenli sıralı istatistikler; marjinal sıralamalar göz önüne alınarak tanımlanabilir. Çok değişkenli sıralı istatistiklerdeki inceleme, tek değişkenli sıralı istatistiklerdeki kadar ayrıntılı olmaz. Bu çalışmada sıralı istatistiklerde, bağımsız fakat aynı dağılımlı olması gerekmeyen tesadüfi değişkenlerin bağımsız ve aynı dağılımlı tesadüfi değişkenler türünden veren teorem, çok değişkenli hale genelleştirilmiştir.

## 2. Çok değişkenli Sıra İstatistikler

Bu çalışmadaki ilişkiler ve aritmetik işlemler daima bileşenler türünden incelenecektir.

$\mathbf{x} = (x_1, x_2, \dots, x_m)$  ve  $\mathbf{y} = (y_1, y_2, \dots, y_m)$  verildiğinde

$$\mathbf{x} \leq \mathbf{y} \text{ olması } i = 1, 2, \dots, m \text{ için } x_i < y_i \quad (1)$$

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \dots, x_m + y_m)$$

olarak ele alınacaktır.

$\xi_1, \xi_2, \dots, \xi_n$  m-boyutlu n tane tesadüfi vektör olsun. Burada  $\xi_i = (\xi_{i,1}, \xi_{i,2}, \dots, \xi_{i,m})'$  dir.

$\xi_{1,j}, \xi_{2,j}, \dots, \xi_{n,j}$ , j-inci bileşenlerin sıralanmış değerleri

$$X_{1n}^{(j)} \leq X_{2n}^{(j)} \leq \dots \leq X_{nn}^{(j)}$$

olarak ifade edilebilir.  $Z_{r:n}$  dönüşümü kullanılarak

$$X_{r:n}^{(j)} = Z_{r:n}(\xi_{1,j}, \xi_{2,j}, \dots, \xi_{n,j})$$

olarak yazılabilir. Ayrıca

$$\mathbf{X}_{r:n} = (X_{r:n}^{(1)}, X_{r:n}^{(2)}, \dots, X_{r:n}^{(m)})$$

olarak ifade edilebilir. (1)'de tanımlandığı gibi

$$\mathbf{X}_{1n} \leq \mathbf{X}_{2n} \leq \dots \leq \mathbf{X}_{nn}$$

olur. Burada,

$$\mathbf{X}_{1n} = (X_{1n}^{(1)}, X_{1n}^{(2)}, \dots, X_{1n}^{(m)})$$

m-değişkenli örnek minimumu ve

$$X_{n:n} = (X_{n:n}^{(1)}, X_{n:n}^{(2)}, \dots, X_{n:n}^{(m)})$$

m- değişkenli örnek maksimumdur.

**Teorem 1.1.** (i.n.n.i.d. tesadüfi değişkenler)

$$X_{1:n} \leq \dots \leq X_{n:n}, \xi_1, \dots, \xi_n \text{ i.n.n.i.d. tesadüfi değişkenlerinin sıralı istatistikleri ve } F_i, \xi_i \text{ 'nin d.f.'}$$

si olsun. O zaman her B Borel cümlesi için

$$P\{(X_{1:n}, \dots, X_{n:n}) \in B\} = \sum_{k=1}^n (-1)^{n-k} \frac{k^n}{n!} \sum_{|S|=k} P\{(X_{1:n}^S, \dots, X_{n:n}^S) \in B\}$$

olur. Burada toplam,  $\{1, 2, \dots, n\}$  'nin k elemanlı tüm S alt kümeleri üzerinden alınır. Bununla birlikte

$$X_{1:n}^S \leq \dots \leq X_{n:n}^S,$$

$$F^S = |S|^{-1} \sum_{i \in S} F_i$$

ortak dağılım fonksiyonlu n tane i.i.d. tesadüfi değişkenin sıralı istatistiklerini göstermektedir. [Reiss, 1989].

#### KAYNAKLAR

[1] Barnett, V. (1976), The Ordering of Multivariate Data, Journal Roy. Statist. Soc., Ser. A, 139, 318-344.

[2] Reiss, R.D. (1989), Approximate Distributions of Order Statistics, Springer-Verlag, New York Inc., USA.

[3] Guilbaud, O. (1982), Functions of non-i.i.d. random vectors expressed as functions of i.i.d. random vectors, Scandinavian Journal of Statistics, 9, 229-233.

#### ABSTRACT ON THE MULTIVARIATE ORDER STATISTICS

In the study, the definition of multivariate order statistics is given furthermore, the theorem related to distribution of order statistics of i.n.n.i.d. (independent not necessarily identically distributed) random variables to that of order statistics of i.i.d. (independent and identically distributed) random variables is generalized for multivariate case.

**Key Words:** Multivariate order statistics, independent not necessarily identically distributed, independent and identically distributed.

### BİLDİRİ OTURUMLARI 3

#### SESSION 3

#### Bulanık Teori

### TİP-1 BULANIK FONKSİYON YAKLAŞIMININ ZAMAN SERİLERİNDE BİR UYGULAMASI

Ali Zafer DALAR<sup>1\*</sup>, Erol EĞRİOĞLU<sup>1</sup>, Ufuk YOLCU<sup>2</sup>, Çağdaş Hakan ALADAĞ<sup>3</sup>, İsmail Burhan TÜRKŞEN<sup>4</sup>

<sup>1</sup>Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 55139, Samsun, TÜRKİYE E-mail: \*alizafer.dalar@omu.edu.tr, erole@omu.edu.tr<sup>2</sup>Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06100, Ankara, TÜRKİYE, E-mail: varyansx@hotmail.com<sup>3</sup>Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800, Ankara, TÜRKİYE, E-mail: aladag@hacettepe.edu.tr<sup>4</sup>TOBB Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, 06560, Ankara, TÜRKİYE, E-mail: bturksen@etu.edu.tr

#### Giriş

Bulanık küme teorisi ilk olarak Zadeh [1] tarafından ortaya atılmış ve bu teori birçok alanda uygulanmıştır. Bu teoriden yola çıkılarak birçok çıkarım sistemi ortaya koyulmuştur. Bulanık çıkarım sistemleri tahmin problemleri için yaygın kullanılmakla birlikte, zaman serisi tahmin problemi için yeterince yaygın kullanılmamaktadır. Zaman serisi tahmininde bulanık küme teorisinin en sık kullanıldığı yöntemler bulanık zaman serisi yaklaşımlarıdır. Bulanık zaman serisi yöntemleri, klasik zaman serisi yöntemleri için geçerli olan kısıtlara sahip olmaması nedeniyle zaman serisi öngörüsü elde etmek için son yıllarda literatürde sıklıkla kullanılmaya başlanmıştır. Birçok bulanık zaman serisi yaklaşımı, klasik bulanık çıkarım sistemleri gibi kural tabanlı çalışmaktadır. Kuralların belirlenmesi çıkarım sistemi içinde önemli bir problem olmakla birlikte, yöntemlerin performansını da etkileyen önemli bir faktördür.

Bulanık çıkarım sistemlerindeki kural tabanlarına gerek duymayan bulanık fonksiyon yaklaşımı Türkşen [2] tarafından ortaya konulmuştur. Bulanık fonksiyon yaklaşımları, bulanık küme teorisine dayalı olarak regresyon ve kümeleme problemleri için önerilmiştir. İlerleyen yıllarda Türkşen'in [2] bulanık fonksiyon yaklaşımı, farklı yapay zekâ yöntemleri ve farklı bulanık küme tipleri kullanılarak geliştirilmiştir (Çelikyılmaz, [3]). Bulanık fonksiyon yaklaşımları zaman serisi tahmini için, regresyon problemi çerçevesinde diğer zaman serilerinin eş anlı kullanılması ile çözümlenmeye çalışılmıştır. Ancak zaman serisi öngörüsü elde etmede, çoğunlukla zaman serisinin çözümlenmesinde diğer zaman serilerinin eş anlı değişkenleri yerine, çözümlenecek zaman serisinin gecikmeli değişkenlerini kullanmak başarılı sonuçlar vermektedir. Bu çalışmada zaman serisi öngörüsü için standart bulanık C-ortalama kümeleme yöntemini kullanan tip-1 bulanık fonksiyon (T1BF) yaklaşımı, öngörülecek zaman serisinin gecikmeli değişkenleri kullanılarak, bir gerçek hayat zaman serisine uygulanmıştır ve elde edilen sonuçlar literatürdeki diğer bazı yöntemler ile karşılaştırılmıştır.

#### Uygulama

Yöntemin performansının değerlendirilmesi amacıyla literatürde sıklıkla kullanılan bir gerçek hayat zaman serisi kullanılmıştır. Zaman serisi 1956-1994 yılları arasında çeyreklik gözlemlenen Avusturya bira tüketimi zaman serisidir ve mega litre cinsinden değerler olarak verilmiştir. Toplam 148 gözlem içeren zaman serisinin son 16 gözlemi test verisi olarak, diğer gözlemler ise eğitim verisi olarak kullanılmıştır. Yöntemin performansının karşılaştırılması amacıyla, bu yöntemin yanı sıra literatürde sıklıkla kullanılan bazı yöntemler

SARIMA- Mevsimsel Otoregresif Bütünleşik Hareketli Ortalama Modeli

WMES- Winters'ın Çarpımsal Üstel Düzleştirme Yöntemi

L&NL ANN- Lineer ve Lineer Olmayan Yapay Sinir Ağları (Yolcu vd., [4])

MNM-FTS-Çarpımsal Nöron Modele Dayalı BZS Yaklaşımı (Aladağ, [5])

ile zaman serisi çözümlenme sonuçları kullanılmıştır. Yöntemin uygulanmasında gecikme sayısı (m) 2 ile 16 arasında 1 artırılarak, küme sayısı (ks) 3 ile 10 arasında 1 artırılarak ve alfa kesim katsayısı ( $\alpha$ ) 0 ve 0.1 alınarak toplamda 240 farklı çözüm elde edilmiştir. Bulunan bu farklı çözümler arasında en iyi uygunluk değerine sahip olan çözüm, m=12, ks=10 ve  $\alpha=0.1$  olduğu durumda elde edilmiştir. Tüm yöntemlerden elde edilen öngörüler

için hata ölçütleri (Hata Kareler Ortalaması Karekökü-HKOK, Hatanın Mutlak Yüzdelik Ortalaması-HMYO, Yön Doğruluğu-YD) Çizelge 1'de verilmiştir.

#### Çizelge 1. Tüm Yöntemlerden Elde edilen Öngörüler ve Performans Ölçütleri

Test Data	SARIMA	WMES	L&NL ANN	MNM-FTS	T1BF
HKOK	47.0367	53.3295	18.7888	29.1381	17.8515
HMYO	0.0949	0.1072	0.0357	0.0532	0.0350
YD	0.7333	0.6667	1.0000	0.9333	1.0000

#### KAYNAKLAR

- [1] Zadeh L.A. (1965). Fuzzy Sets, Inform and Control, 8, 338-353.
- [2] Turksen I.B. (2008). Fuzzy function with LSE, Applied Soft Computing, 8, 1178-1188.
- [3] Celikyilmaz A. (2005). Fuzzy Functions with support vector machines. M.A.Sc. Thesis, Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada.
- [4] Yolcu U., Aladag C.H. and Egrioglu E. (2013). A New Linear & Nonlinear Artificial Neural Network Model for Time Series Forecasting, Decision Support System Journals, 54, 1340-1347.
- [5] Aladag C.H. (2013). Using multiplicative neuron model to establish fuzzy logic relationships, Expert Systems with Applications, 40(3), 15, 850-853.

#### ABSTRACT

##### AN APPLICATION of TYPE-1 FUZZY FUNCTION APPROACH For TIME SERIES

Fuzzy inference systems have been used for prediction problems in the literature. Classical fuzzy inference systems are rule based systems. The determination of the rules is important and difficult problem. Fuzzy function approaches were proposed as a good alternative for fuzzy inference systems. These approaches are not rule based. This is a big advantage for them. Fuzzy function approaches were applied to obtain forecasting by using simultaneous variables of other time series as covariates. In this study, type-1 fuzzy function (T1FF) approach has been applied to obtain forecasts of Australian beer consumption time series. The performance of T1FF approach has been compared with some recent methods in the literature.

**Key Words:** Type-1 Fuzzy Function Approach, Forecasting, Fuzzy C-means, Fuzzy Time Series, Artificial Neural Networks.

**BULANIK SİSTEM MODELLEME**

\*Elif Burcu DİLDEN, I.Burhan TÜRKŞEN

TOBB Teknoloji ve Ekonomi Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü  
Ankara /Türkiye  
ebdilden@etu.edu.tr , bturksen@etu.edu.tr

**Bulanık Regresyon Modelleme**

Klasik regresyondan farklı olarak, bulanık regresyonda üyelik değerleri Bulanık C- Ortalamalar algoritması kullanılarak veri kümesinden çekilmekte ve regresyon modeli içerisinde girdilerle beraber dahil edilmektedir. Analiz içerisinde kullanılan veri setine ait girdi ve çıktı bilgileri aşağıda gösterilmektedir:

X1	X2	X3	X4	X5	X6	X7	X8	Y
Cement	Blast Furnace Slag	Fly Ash	Water	Superplas Ticizer	Coarsa aggregate	Fine Aggregate	Age	Concrete Compressive Strength

**Çizelge 1. Beton Sertliği Dayanıklılığı Veri Setine Ait Değişkenler**

Öncelikle veri setine ait girdilerden regresyon modeline seçilebilmesi kapsamında korelasyon analizi yapılmıştır. Modelde bulunması gereken etkin değişkenler x1, x5 ve x8 olarak belirlenmiştir. FCM öbekleme algoritması uygulanarak, öbek sayısı c değeri altı ve bulanıklık derecesi m parametresi ise 1.8 olarak bulunmuştur. Bulanık c- ortalama algoritması sonucunda bulunan üyelik dereceleri ile klasik regresyondan farklı olarak girdi matrisi X aşağıdaki gösterilen bazı örneklerle değerlendirilir:

$$X' = [1, \mu_i, X], \quad X_i'' = [1, \mu_i^2, X], \quad X_i''' = [1, \mu_i^2, \mu_i^m, \exp(\mu_i), X] \quad (1)$$

Burada girdi matrisi X ve üyelik değerlerinin çeşitli transformasyonlarını  $\mu_i$ 'ler ile modeller kurularak belirleme katsayısı  $R^2$  ve ortalama karesel hata RMSE'ye göre değerlendirilir. Bulanık regresyon yapılırken üyeliklere ait transformasyonların örneğin;  $\mu, \mu^2, \log((1-\mu)/\mu)$  gibi değerlendirilerek uygun bir yapı belirlenir.

Buna bağlı olarak yapılan bulanık regresyona ait  $R^2$  ve RMSE değerleri aşağıda verilmiştir:

	öbek 1	öbek 2	öbek 3	öbek 4	öbek 5	öbek 6
RMSE	0,07185	0,07797	0,08315	0,10018	0,07985	0,06875
R2	88,2	86,1	84,2	77,1	85,5	89,2

**Çizelge 2. Bulanık regresyona ait R2 ve RMSE değerleri**

Sonuç olarak ilişkilerin belirsizlik altında, gerçeği daha iyi yansıtabilmesi bakımından bulanık regresyon ile yapılan modellemenin diğer yaklaşımlara göre, daha üstün performans verdiği açıkça görülmektedir.

**KAYNAKLAR**

- [1] L.A.Zadeh, (1965), Fuzzy Sets, Information and Control, 8(3), pp.338-353  
[2] L.A.Zadeh, (1965), Fuzzy Sets and Systems, In:Fox J.ed.,System Theory.Polytechnic Press,Brooklyn, Newyork, pp.29-39.  
[4] A.Celikyilmaz and I.B. Turksen, (2009), Modeling Uncertainty with Fuzzy Logic, Studies in Fuzziness and Soft Computing, Volume 240

**ABSTRACT**

**FUZZY SYSTEM MODELLING**

With application of FCM algorithm, membership values are calculated and fuzzy regression analysis is conducted. It is shown that the performance of the "fuzzy" regression analysis is better than the other benchmark strategies.

**Key Words:** Fuzzy logic, fuzzy regression, fuzzy clustering, prediction.

**KAPI KASASI İMALATININ FULL TİP-2 BULANIK YAKLAŞIMLA MODELLENMESİ**

M. Bahar BAŞKIR<sup>a</sup>, İ. Burhan TÜRKŞEN<sup>b</sup>

<sup>a</sup>Bartın Üniversitesi Fen Fakültesi İstatistik Bölümü 74100 Merkez/BARTIN

<sup>b</sup>TOBB Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği Bölümü 06560 Söğütözü/ANKARA

[mbaskir@bartin.edu.tr](mailto:mbaskir@bartin.edu.tr); [bturksen@etu.edu.tr](mailto:bturksen@etu.edu.tr)

Kapı Kasası İmalat Sürecinin Full Tip 2 Bulanık Öbeklenmesi

**Full Tip 2 Bulanık Öbekleme Algoritması**

Gerçek yaşama ait bir sistem birimlerinin benzerlik veya benzememe durumlarına göre öbeklenme yapılarını incelemeye Bulanık Öbek Ortalamaları (FCM) ([1]) algoritması sıkça kullanılmaktadır. Öte yandan, Tip 2 bulanık sistem yapısını ortaya koymak amacı ile Türkşen ([2]) tarafından önerilen Full Tip 2 Bulanık Öbekleme (FT2FCM) algoritması bir en iyileme problemi olarak (1)'deki gibidir.

$$\min J'(U'(U), W) = \sum_{k=1}^n \sum_{l=1}^{c'} (u'_{lk}(u_a))' (\|u'_{lk} - w_{lk}\|)_A \quad (1)$$

$$0 \leq (u'_{lk}(u_a)) \leq 1, \forall i, k; \sum_{l=1}^{c'} (u'_{lk}(u_a)) \leq 1, \forall k; 0 \leq \sum_{k=1}^n (u'_{lk}(u_a)) \leq n, \forall i, l = 1, 2, \dots, c'; k = 1, 2, \dots, n$$

Burada,  $X = \{x_1, x_2, \dots, x_n\}$  p-boyutlu uzayda n elemanlı veri seti,  $W = [w_{1k}, w_{2k}, \dots, w_{ck}]$ , ( $c'$ -öbek sayısı,  $c' \in (1, n)$ ) öbekleri temsil eden merkez vektörü,  $u'_{lk}$ ' lar  $u_{ik}$ ' nin ( $u_{ik}, x_k \in X$ 'in öbeğe ait olma derecesi)  $w_{lk}$ ' ya ait olma derecesi olmak üzere  $U'(U) = [(u'_{lk}(u_a))]_{c' \times n}$  üyelik matrisinin üyelikleri;  $m'$  uygun bulanıklık mertebesidir.  $J'$  minimize edilen amaç fonksiyonunu,  $\|\cdot\|_A$  A=I ile Öklid normunu göstermektedir. (1) ile verilen en iyileme probleminin çözümü Çizelge 1'deki gibidir.

**Çizelge 1. FT2FCM algoritmasının en iyileme problemi çözümleri.**

Üyelik Matrisinin Üyelikleri	$(u'_{lk}(u_a)) = \left[ \sum_{l=1}^{c'} \left( \frac{\ u'_{lk} - w_{lk}\ _A}{\sum_{l=1}^{c'} \ u'_{lk} - w_{lk}\ _A} \right)^{\frac{2}{2/(m')-1}} \right], i^* \neq j$
Merkezler	$w_{lk} = \frac{\sum_{k=1}^n (u'_{lk}(u_a))' u_{lk}}{\sum_{k=1}^n (u'_{lk}(u_a))' }, \forall l = 1, 2, \dots, c'$

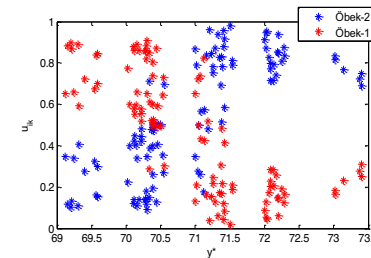
**Kapı Kasası İmalat Sürecinin Full Tip 2 Bulanık Öbeklenme Yapısı**

Bir Kapı Doğrama Fabrikasında birim kapı imalatında yapıyı gerçekleştirilen altı ana kısım; Kapı Kasası, Kapı Kanadı, Ayarlı Pervaz, Panel Kapı Kanadı, Cam Çıtalar ve Havalandırma Menfezinden oluşmaktadır (bkz. [3]). Birim Kapı Kasası imalatı toplamda 17 alt aşamadan oluşmaktadır. Bu alt aşamaların 6 tanesinde insan gücünden yararlanılmaktadır. Kapı kasası alt aşamalarında geçen ortalama süreler bakımından problemin çözümü açısından en etkili olanları Pareto analizine göre belirlenmiştir (bkz. Çizelge 2).

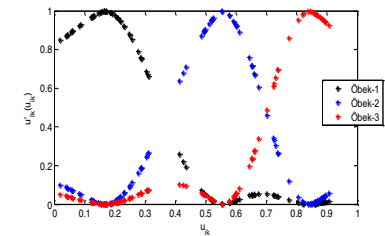
**Çizelge 2. Kapı kasası alt aşamalarında geçen ortalama süreler için Pareto analizi sonuçları.**

	Aşamalar					
	11.	13.	14.	15.	16.	17.
Ort. Süre	15.23	15.00	10.02	10.02	9.33	9.33
Yüzde (%)	22.1	21.8	14.5	14.5	13.5	13.5
Toplam %	22.1	43.9	58.4	<b>72.9</b>	86.5	100.0

Çizelge 2'ye göre, Kapı kasasına ilişkin Çok Girdili Tek Çıktılı (ÇGTÇ) sistemde 11., 13., 14. ve 15. alt aşamanın gerçekleşme süreleri (sırasıyla  $x_1^*$ ,  $x_2^*$ ,  $x_3^*$  ve  $x_4^*$ ) girdileri; 6 adet alt aşamanın toplam gerçekleşme süresi ( $y^*$ ) çıktısı oluşturmaktadır. Toplam süreye taşıma süreleri de dahildir. Buna göre, Kapı kasası için oluşturulan 100x5 boyutlu ÇGTÇ sistem ile FCM algoritmasından ve Kim-Ramakrishna ([4]), Çelikiylmaz-Türkşen ([5]) öbek geçerlik indekslerinden yararlanılarak elde edilen en iyi ( $c^*=2$ ,  $m^*=1.8$ ) ikilisi için öbeklenme yapısı Şekil 1'de gösterilmektedir. Şekil 1'de gösterilen Öbek-2'ye ait üyeliklerin FT2FCM algoritması sonucu öbeklenme yapısı Şekil 2'deki gibidir.



**Şekil 1. En iyi ( $c^*=2, m^*=1.8$ ) ikilisi için kapı kasası ÇGTÇ sistemin bulanık öbeklenmesi.**



**Şekil 2. En iyi ( $c^*=3, m'=2.2$ ) ikilisi için Öbek-2 üyeliklerinin bulanık öbeklenmesi.**

**KAYNAKLAR**

- [1] Bezdek, J.C. (1981), Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York.
- [2] Türkşen, İ.B. (2013). Full Type 2...Full Type n Fuzzy System Models, the 13th Annual Conference of the European Network for Business and Industrial Statistics (ENBIS-13), Ankara (15-19 September).
- [3] Başkır, M.B. (2006), Kapı-Doğrama Sürecinde Altı Sigma Yaklaşımı, 5. İstatistik Günleri Sempozyumu Bildiriler Kitabı, 71-78, Antalya (24-27 Mayıs).
- [4] Kim, M. and Ramakrishna, R.S. (2005), New indices for cluster validity assessment, Pattern Recognition Letters, Cilt 26, No 15, 2353-2363.
- [5] Çelikiylmaz, A. and Türkşen, İ.B. (2008a), Validation Criteria for Enhanced Fuzzy Clustering, Pattern Recognition Letters, 29 (2), 97-108.

**MODELING OF A DOOR FRAME PRODUCTION BY FULL TYPE 2 FUZZY APPROACH**

We have studied with a Door Joinery Factory of a Turkish Corp. Group. We investigate the reasons of the Factory's problem and identify the base of the problem as the use of human sources on production time of Door Frame. The systems of the production time of door frame are analyzed with the proposed Full Type 2 fuzzy approach by Türkşen ([2]).

**Key Words:** Door frame, pareto analysis, Full Type 2 fuzzy clustering, cluster validity indices.

## TÜRKİYE'DE BOŞANMA SAYILARI VE SOSYO EKONOMİK GELİŞİMİŞLİĞE GÖRE İLLERİN KÜMELENMESİ

Ayşe ÇAY ATALAY, Ebül Muhsin DOĞAN, Ahmet ATALAY

Atatürk Üniversitesi Narman Meslek Yüksekokulu 25530 Narman/Erzurum Ayatalay@Atauni.Edu.Tr  
Erzurum Teknik Üniversitesi İktisadi İdari Bilimler Fakültesi 25240 ERZURUM muhsindogan@erzurum.edu.tr  
Atatürk Üniversitesi Narman Meslek Yüksekokulu 25530 Narman/ERZURUM ahatalay@atauni.edu.tr

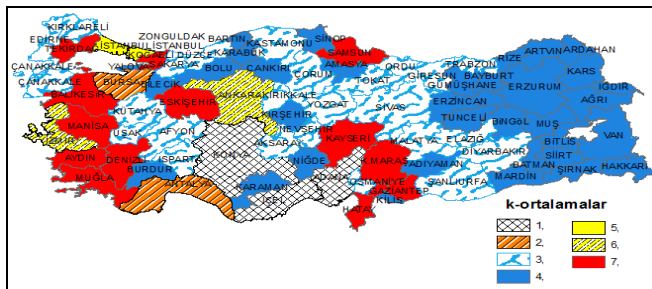
### Genişletilmiş Özet

Bu çalışmada 2011 yılında Türkiye'de illerde gerçekleşen boşanma sayıları ve illerin sosyo ekonomik gelişmişlik sıraları (SEGE) dikkate alınarak illerin kümeleme analizi yapılmıştır. Kümeleme analizi hem klasik k-ortalamlar hem de bulanık c-ortalamlar teknikleri kullanılarak yapılmıştır. Kümeleme analizi ile iller 7 kümeye ayrılmıştır. K-ortalamlar tekniği SPSS programı ile, bulanık c-ortalamlar tekniği ise Matlab programı ile yapılmıştır. Ayrıca her iki yöntem ile elde edilen kümeleme sonuçlarına göre Coğrafi Bilgi Sistemleri(CBS) yazılımı olan ArcInfo yazılımı ile illerin tematik haritaları oluşturulmuştur. Çalışmada kullanılan veriler Türkiye İstatistik Kurumu (TÜİK) den elde edilmiştir.

### 1. 1. K-ortalamlı Kümeleme Yöntemi

Hiyerarşi olmayan kümeleme yöntemlerinden en çok kullanılanı k-ortalamlar tekniği, her iterasyonda yeni bir küme merkezi oluşturularak, bir eleman yeniden hesaplanan yeni merkeze daha yakın ise o kümeye taşınır. Bu yöntemde bireyler, gruplar için kareler toplamını en küçük yapacak şekilde k kümeye bölünür  $x_1, x_2, \dots, x_n$  değişkenlerinin her biri p değişkenli gözlem vektörleri, çok boyutlu X uzayında birer nokta ifade ederken, aynı uzayda  $a_{1n}, a_{2n}, \dots, a_{kn}$  her grup birey için küme merkezleri olarak belirlendiğinde, aşağıdaki formüle göre bireyler en küçük uzaklığı veren (en yakın) kümeye sınıflanmaktadır (Pollard, 1981; Tatlıdil, 1996; Özgür, 2003).

$$W_N = \frac{1}{n} \sum_{i=1}^n \min \|x_i - a_m\|^2 \dots \dots \dots (1)$$



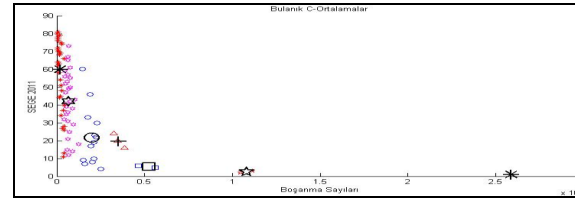
Şekil 1. K-ortalamlar tekniğine göre illerin tematik haritası

### 2. 2. Bulanık C- Ortalamalı Kümeleme (Fuzzy c-Means Clustering) Yöntemi

Bulanık kümeler klasik kümelerden farklı olarak her bir veri noktası birden fazla alt kümeye değişik derecelere ait olabilmektedir. Ancak aynı veri noktasını üst üste düşen değişik kümelerdeki üyelik derecelerinin toplamı 1'e eşit olmalıdır. Bunun anlamı bir i verisinin j kümesine ait olma üyelik derecesi  $u_{ij}$  ise m küme sayısı olmak üzere;

$$\sum_{j=1}^m u_{i,j} = 1 \dots \dots \dots (2)$$

geçerlidir.



Şekil 2. Bulanık c-ortalamlar tekniğine göre illerin dağılımı

Çalışmada iki yöntemle elde edilen kümelerdeki illerde farklılıklar görülmektedir. Bu farklılığın nedeni olarak bulanık c-ortalamlar tekniği k-ortalamlar tekniğine göre başlangıç değerlerinden daha az etkilendiğinden kaynaklanmaktadır. Bulanık c ortalamlar tekniği genellikle daha kararlı sonuçlar ürettiği gözlemlenmiştir.

**Anahtar kelimeler:** kümeleme analizi, k-ortalamlar, bulanık c-ortalamlar

### Kaynaklar

- [1] Tatlıdil, H. (1996), Uygulamalı Çok Değişkenli İstatistiksel Analiz, Cem Ofset Ltd. Şti., Ankara.
- [2] Özdamar, K. (2002), Paket Programlar ile İstatistiksel Veri Analizi(Çok değişkenli Analizler), 4. Baskı, Kaan Kitabevi, Eskişehir.
- [3] Pollard, D. (1981), Strong consistency of k-means clustering, The Annals of statistics. 9 (1), 135-140.
- [4] Özgür, E.. (2003), Çok Değişkenli İstatistiksel analiz Yöntemleri ve Bir Uygulama, Doktora tezi, Sosyal Bilimler Enstitüsü, Gazi Üniversitesi, Ankara.
- [5] Höppner, F., Klawonn, F., Kruse, R., and Runkler, T. (2000), Fuzzy Cluster Analysis, John Wiley&Sons, Chichester.

### ABSTRACT

### CLUSTERING OF PROVINCES IN TURKEY ACCORDING TO NUMBER OF DIVORCE AND SOCIO-ECONOMIC DEVELOPMENT

In this study, the clustering of provinces was carried out taking into account the divorce took place in provinces and socio-economic development of provinces in Turkey in 2011. Clustering analysis was carried out using both the techniques of fuzzy c-means and the classical k-means. The provinces is divided into seven Cluster by cluster analysis. K-means technique was carried out by using SPSS software, fuzzy c-means technique performed with the Matlab. In addition, according to the results based on clustering obtained by the two methods, thematic maps of the provinces were carried out with ArcInfo that is Geographic Information Systems software.

**Keywords :** Cluster analysis, k-means, fuzzy c-means

## BULANIK VERİ ZARFLAMA ANALİZİ YÖNTEMİ İLE ÇORUM İLİNDEKİ LİSELERİN ETKİNLİK ÖLÇÜMÜ

Emre DEMİR, Sevil DEMİR

Hitit Üniversitesi, Osmancık Ömer Derindere MYO

Teknik Programlar Bölümü, 19500, Çorum, Türkiye emredemir@hitit.edu.tr

Üstün Zekâlılar Enstitüsü (ÜZE), Üçtutlar Mah.

Anadolu 2. Sok. No: 10/10A, 19030, Çorum, Türkiye, sewilyalcin@hotmail.com

### 1.Özet

Türk Milli Eğitim Sisteminin önemli bir parçası olan ortaöğretim kurumları **ülkemizin kalkınma ve gelişiminde önemli eğitim kurumlarıdır**. Öğrencilerimizin ortaöğretim kurumlarında daha kaliteli eğitim alabilmesi için öncelikle okullarımızın etkin çalışması gerekmektedir. Bunun sağlanabilmesi için eğitim kurumlarımızın koşullarının devamlı iyileştirilmesi, buna bağlı olarak da okullarımızın mevcut durumunun belirlenmesi gerekmektedir. Kaynakların sınırsız olmaması daha etkin ve verimli bir şekilde kullanılmasının önemini ortaya çıkarmaktadır. Bu kaynakları kullanan karar verme birimlerinin etkinliğinin karşılaştırılmasında kullanılan yöntemlerden biri olan Veri Zarflama Analizi (VZA) ilk kez 1957 yılında Farrell tarafından ortaya atılan Sınır Üretim Fonksiyonu önerisi ile şekillenmiş, VZA yöntemi ilk olarak Charnes, Cooper ve Rhodes tarafından, kamu kuruluşlarının teknik verimliliğini ölçmek ve karşılaştırmak amacıyla geliştirilmiştir. Gerçekleştirilen ilk VZA uygulamasında (Charnes ve diğ., 1978) okulların karşılaştırmalı verimliliklerinin ölçülmesi hedeflenmiştir.

İstisnalar haricinde, tüm girdi ve çıktı verilerinin belirsizlik sebebiyle tam olarak elde edilemediği bir durumda, bu değerlerin yalnızca  $x_{ij}^L > 0$  ve  $y_{rj}^L > 0$  olmak üzere ve  $[x_{ij}^L, x_{ij}^U]$  ve  $[y_{rj}^L, y_{rj}^U]$  şeklindeki aralıklarla gösterilen alt ve üst sınırlar arasında oldukları biliniyor olsun. Etkinlik ölçümü için sınırlandırılmış verilerin kullanıldığı girdiyeye yönelik Aralık Sayılarla Bulanık Veri Zarflama Analizi (BVZA) modeli:

$$\max \theta_{j_0}^U = \frac{\sum_{r=1}^s u_r y_{rj_0}^U}{\sum_{i=1}^m v_i x_{ij_0}^L} \quad (1)$$

$$\theta_j^U = \frac{\sum_{r=1}^s u_r y_{rj}^U}{\sum_{i=1}^m v_i x_{ij}^L} \leq 1, \quad j = 1, \dots, n, \quad u_r \geq \varepsilon > 0, v_i \geq \varepsilon > 0, \quad \forall r, i \quad (2)$$

$$\max \theta_{j_0}^L = \frac{\sum_{r=1}^s u_r y_{rj_0}^L}{\sum_{i=1}^m v_i x_{ij_0}^U} \quad (3)$$

$$\theta_j^U = \frac{\sum_{r=1}^s u_r y_{rj}^U}{\sum_{i=1}^m v_i x_{ij}^L} \leq 1, \quad j = 1, \dots, n, \quad u_r \geq \varepsilon > 0, v_i \geq \varepsilon > 0, \quad \forall r, i \quad (4)$$

biçiminde tanımlanır (Kao ve Liu, 2000, Wang ve diğ., 2005, Güneş, 2006).

Bu çalışmada, girdiler için kullanılan veriler öncelikle Çorum Milli Eğitim Müdürlüğü'nden elde edilmiş daha sonra okullardan alınarak karşılaştırılmıştır. Ancak aynı verilerde farklılıkların olduğu görülmüştür. VZA veriye çok duyarlı bir yöntemdir. Bu yüzden analizde kullanılan verilerde hata bulunabilme düşüncesinden hareketle, Çorum ilindeki liselerin 2012-2013 eğitim-öğretim sürecindeki göreceli etkinlikleri öncelikle Klasik VZA ile hesaplanmıştır. Daha sonra BVZA tekniklerinden biri olan; Aralık Veriler Yöntemi kullanılmıştır. EMS (Version 1.3) paket programı iki farklı durum için çalıştırılmış ve etkinlik değerleri verilmiştir. Etkin olmayan okullar için referans kümeleri belirlenmiş ve örnek potansiyel iyileştirme tablosu oluşturulmuştur. Elde edilen bulgular sonucunda okulların etkinliğini arttırmalarına katkıda bulunulması amacıyla karar verme birimlerine önerilerde bulunulmuştur.

Analizlerde kamu performans karşılaştırılmasında yaygın olarak kullanılan girdiyeye yönelik ölçüğe göre sabit getirili Charnes-Cooper-Rhodes (CCR) Modeli (1978) kullanılmıştır. Etkinlik ölçümü için okullardaki öğrenci sayısı, öğretmen sayısı ve şube sayısından oluşmak üzere üç adet girdi ve yükseköğretime yerleştirme (başarı) oranı, YGS puan ortalamaları, LYS Matematik-Fen, Türkçe-Matematik, Türkçe-Sosyal puanları olmak üzere beş adet çıktı belirlenmiştir.

### KAYNAKLAR

- [1] Charnes, A. Cooper, W. W. And Rhodes, E. (1978), Measuring the Efficiency of Decision Making Units, European Journal of Operational Research, Vol.2, No.6, pp. 429-444.
- [2] Kao, C. and Liu, S. T. (2000), Fuzzy Efficiency Measures in Data Envelopment Analysis, Fuzzy Sets and Systems, Vol.113, pp. 427-437.
- [3] Güneş, T. (2006), Bulanık Veri Zarflama Analizi, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- [4] Wang, Y.M., Greatbanks, R. and Yang, J.B. (2005), Interval Efficiency Assessment Using Data Envelopment Analysis, Fuzzy Sets and Systems, Vol.153, pp. 347-370.

### ABSTRACT

#### EFFECTIVENESS MEASUREMENT OF HIGH SCHOOLS IN ÇORUM BY FUZZY DATA ENVELOPMENT ANALYSIS

In this study, initially, relative efficiency in the education processes was calculated at high schools in Çorum in 2012-2013 year of education by means of traditional Data Envelopment Analysis (DEA). Then technique of interval data, which is one of the Fuzzy Data Envelopment Analysis (FDEA), was used. As a consequence of the data, benchmarks were defined for non-efficient schools and sample potential improvement tables. After that, recommendations were noted for decision making units in order to contribute the efficiency of the intuitions.

**Key Words:** High Schools, Fuzzy Data Envelopment Analysis, İnterval Data, Efficiency, Productivity.

### TÜRKİYE EKONOMİSİNİN BAZI GÖSTERGELERİ ÜZERİNE EKONOMETRİK ÇÖZÜMLEMELER VE ÖNGÖRÜ

Onur TOKA<sup>1</sup>, Cengiz ARIKAN<sup>2(\*)</sup>

<sup>1</sup>Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800, Ankara, TÜRKİYE  
onur.toka@hacettepe.edu.tr

<sup>2</sup>Gümrük ve Ticaret Bakanlığı, 06800 Ankara, TÜRKİYE

C.Arikan@gtb.gov.tr

Ekonominin kuramsal olarak belirlemiş olduğu etkiler, ülkelerde çeşitli farklılıklar gösterebileceği gibi bire bir aynı sonuçları da verebilir. Değişen ekonomik koşullar ve hızlı gelişen teknoloji ekonomiyi de yakından etkilemektedir. Dışarıdan gelen çeşitli etkiler ekonominin kuramsal tutumunda bazı şaşırtıcı sonuçlar verebilmektedir. Örneğin büyümenin istihdam yaratması ve işsizliğin azalması gibi çıkarımların ortaya çıkmaması, balon büyüme, sıcak para gibi deyimlerin doğmasına sebep olmaktadır.

Ülkelerdeki faiz değerleri, yatırımlar, istihdam, işsizlik gibi ülkenin ekonomik derecelendirmesinde önemli değişkenler birbirleriyle uzun dönemli ilişkiler içindedir. Örneğin keynesyen ekonomi politikasına göre para arzı, faizler ve yatırımlar ile uzun dönemli ilişkiler içerisinde olur ve belirlenecek para arzı kısa dönem içerisinde yatırımlar, milli gelirlerde değişiklikler yaratmaktadır. Aynı şekilde para arzındaki değişiklikler yine bu değişkenlerdeki duruma göre uzun bir dönem içerisinde şekillenmektedir.

Bu çalışmanın amacı Türkiye'nin 1998 yılından 2012 yılı ikinci yarısına kadar bazı faiz değerleri, GSYİH ve bir ölçümü deflator ile para arzı arasında uzun dönemli bir tahmin modeli elde edilmeye çalışmak ve nedensellik ile ilgili sonuçların yorumlanabilmesidir. Bu amaçla ilk olarak kullanılan serilerin mevsimsellik durumları kontrol edilmiş ve ilgili değişkenlerin mevsimsellik düzeltmeleri Tramo Seats teknikleri ile yapılmıştır. Daha sonra değişkenlerin birim kök sınamaları yapılmıştır. Ardından ise klasik ekonometrik yöntemler olarak tabir edilen Eş Bütünleşme Analizi, Granger Nedensellik testi ve VECM teknikleri uygulanarak sonuçlar elde edilmiştir. Kurulan modeller yardımıyla öngörü modelleri elde edilmiş ve sonuçlar yorumlanmıştır.

Elde edilen birincil sonuçlara göre, ele alınan değişkenlerin 5. gecikmeli vektör hata düzeltme modellerinin öngörü yapma imkânı verdiği ve kurulan modellerin üçünün uzun dönemdeki ilişkiyi açıklayabildiği ve dengeye geldiği görülmüştür. Ayrıca değişkenlerin birbirlerinin nedenleri oldukları da ortaya konulmuştur.

#### KAYNAKLAR

[1] Anderson, R.G., Hoffman, D., Rasche, H.R., 2000, A Vector Error-Correction Forecasting Model of the U.S. Economy, The Federal Reserve Bank of St. Louis Working Paper Series, St. Louis.

[2] Dickey, D.A. and Fuller, W.A., 1981, Distribution of the estimators for autoregressive time series with a unit root, Econometrica 49, 1057-72.

[3] Engle, Robert F. and C.W.J. Granger, 1987, Cointegration and error correction: Representation, estimation and testing, Econometrica 55, 251-276

[4] Johansen, S. and Juselius, K, 1990, Maximum Likelihood Estimation and Inference on Cointegration – with Applications to the Demand for Money, Oxford Bulletin of Economics and Statistics 52, 169-210.

[5] Pınarönü, N. 2007, Johansen ve Kısmi Lineer Tekil Değer Eşbütünleşme Testlerinin Bir Uygulama Problemi Üzerinde Karşılaştırılması, Gazi Üniversitesi Fen Bilimleri Enst., Yüksek Lisans Tezi, Ankara.

#### ABSTRACT

APPLYING SOME ECONOMETRIC METHODS ON SOME TURKISH ECONOMIC VARIABLES AND FORECASTING

### BİLDİRİ OTURUMLARI 3

#### SESSION 3

#### Ekonometri



In this study, the relationship between GDP, Interest Rate, Money Supply (M1) and Inflation Rate (CPI) have been investigated whether there is a long run relation or not. The quarterly data have been used from 1998.Q1 to 2012.Q2. To investigate the relations classical econometric methods like Granger Causality Cointegration and VECM are applied. According to results, there is a long run relation between variables.

**Key Words:** Turkish Economy, Macro Economics, Cointegration, VECM

## BORSA ENDEKS GETİRİLERİNDE İKİLİ UZUN HAFIZA ÖZELLİĞİNİN ANALİZİ İÇİN BİR UYGULAMA

Serpil TÜRKYILMAZ<sup>1</sup>, Mesut BALİBEY<sup>2</sup>

<sup>1,2</sup>Bilecik Şeyh Edebali Üniversitesi, Fen-Edebiyat Fakültesi, Matematik Bölümü, Bilecik, Türkiye

[serpil.turkyilmaz@bilecik.edu.tr](mailto:serpil.turkyilmaz@bilecik.edu.tr), [mesut.balibey@bilecik.edu.tr](mailto:mesut.balibey@bilecik.edu.tr)

### Giriş

Uzun Hafıza (Long Memory) dinamikleri bir finansal zaman serisinin ortalaması ve varyansındaki doğrusal olmayan ilişkilerin varlığını gösteren önemli bir işarettir. ARCH ve GARCH türü modellerin Engle(1982) ve Bollerslev(1986) tarafından önerilmesi ile birlikte borsa getirilerinin oynaklıklarının modellenmesi popüler olmuştur. Fakat bu modeller finansal bir zaman serisinin ortalama ve oynaklığındaki uzun hafıza özelliğini dikkate almamaktadır. Getiri ve oynaklığın otokorelasyon fonksiyonlarındaki hiperbolik olarak azalma yani yavaş ortalamaya dönme özelliği uzun hafıza olarak tanımlanır. Özellikle finansal yatırım kararları alınırken ve para politikaları açısından finansal piyasa davranışlarının doğrusal olup olmadığı konusu çok önemlidir. Finansal piyasalardaki fiyat hareketlerinin doğrusal olmayan bir yapı göstermesi yatırım kararları alınırken standart istatistiksel analizlerin hatalı sonuçlar vermesine neden olacaktır. Bu nedenle finansal piyasaların karmaşık yapısı akademisyenlerin, politika yapımcıların ve yatırımcıların piyasalara farklı bir bakış açısıyla yaklaşmalarına neden olmuştur. Doğrusal olmayan fiyat hareketleri Etkin Piyasa Hipotezi' ni test edilebilir hale getirmiştir. Bu çalışmanın konusu (D-8: Türkiye, Pakistan, Bangladeş, Malezya, Endonezya, Mısır, İran ve Nijerya) ülkelerinden bazılarının borsa getirileri ve oynaklıklarındaki ikili uzun hafıza özelliklerinin (2010-2013) dönemi için ARFIMA-FIGARCH türü modeller ile incelenmesini içermektedir. Çalışma verileri (2010-2013) dönemi için söz konusu ülkelerin menkul kıymetler borsası günlük kapanış fiyatlarını kapsamaktadır.

### Sonuçlar

Bu çalışmada gelişme yolundaki ülkeler arasında oluşturulan işbirliği örneklerinden birisi olan Türkiye' nin de içerinde bulunduğu (D-8) ülkeleri Pakistan, İran, Bangladeş, Malezya, Endonezya, Mısır ve Nijerya' dan bazılarının borsa getirilerindeki ikili uzun hafıza özellikleri ARFIMA-FIGARCH türü model türleriyle incelenmiş ve Etkin Piyasa Hipotezi test edilmiştir. D-8 ülkeleri içerisinde finansal sistemleri gelişmiş birlik üyesi ülkelerin diğer ülkelerin finansal sistemlerinin gelişimine katkıda bulunabilmesi oldukça önemlidir. Ülke ekonomilerinin en önemli göstergelerinden biri olan menkul kıymetler borsaları için Etkin Piyasa Hipotezi' nin test edilmesi yatırım politikaları için oldukça önemlidir. Para politikasının ekonomiyi etkileme kanallarından birisi de hisse senedi piyasasıdır. Hisse senedi fiyatları ekonomide net gözlenebilen aktif fiyatları arasındadır ve parasal şoklardan hemen ve doğrudan etkilenmektedir. Hisse senedi piyasasındaki hareketler para politikası kararlarını, yatırım kararlarını önemli ölçüde etkilemektedir. Söz konusu ülkeler için hisse senedi piyasalarındaki ikili uzun hafıza özelliklerinin test edilmesinin yatırımcılar, politika yapımcılar ve akademisyenler için önemli ipuçları sağlayacağı, D-8 ülkeleri içerisinde Türkiye açısından geçmişte kriz yaşayan ancak son dönemlerde hızlı bir ekonomik gelişme gösteren Endonezya, Malezya gibi ülkelerin borsalarının etkinliği test edilerek karşılaştırma imkanı sunacağı beklenmektedir. Bu anlamda ekonomi ve finans literatürüne önemli bir katkı sağlayacağı düşünülmektedir.

### KAYNAKLAR

- [1] Caporale G.M. and L. Gil-Alana (2011), Fractional Integration and Cointegration in US Financial Time Series Data, Economics and Finance Working Paper Series, 11-02, 1-41.
- [2] Christodoulou-Volos C. and F.M. Siokis (2006), Long range dependence in stock market returns, Applied Financial Economics, 16, 1331-1338.
- [3] Kang S.H., Cheong C., and S.M. Yoon, (2010), Long memory volatility in Chinese stock markets, Physica A, 389, 1425-1433.

## ABSTRACT

### AN APPLICATION FOR ANALYSIS OF DUAL LONG MEMORY PROPERTY IN STOCK MARKET RETURNS

The Stock Markets are one of the major indicators of common economic success in countries. It is possible to understand the effects of inflation, growth, investments and crises on economy following changes in the stock market. In this sense, the concept of efficiency is important for the financial markets. Efficient Market Hypothesis advocates that investors do not have an information about future prices of stock markets by means of the historical data. According to this hypothesis, it is not possible to obtain a return above normal since all of information which could affect the price of stock market reflect to the prices before. Furthermore, the direction and intensity of stock prices movements couldn't forecast. The hyperbolic decreasing in autocorrelation functions of the return and it's volatility or the property of mean revision slowly is defined as long memory. The positive autocorrelation between the price movements occurs in the case of the long run dependence property these price movements. In the case of long memory property, the prices of the stock market will have a predictable structure, and trend in the past of prices can use for future price estimations. In this study, it will be tested that how the Efficient Market Hypothesis works in a real market. In addition, the long memory properties in stock market returns and their volatilities of some of D-8 countries (Turkey, Pakistan, Bangladesh, Malaysia, Indonesia, Egypt, Iran, Nigeria) which have a growing importance in the global economy have been analyzed. Furthermore, existence of long memory has been tested both for the returns and volatility of the time series by using different methods such as Rescaled Range Statistics (R/S), Geweke and Porter-Hudak (GPH) Model. The ARFIMA-FIGARCH class models are estimated under the different assumption of distribution such as student-t, skewed student-t and GED distribution. In this study, the long memory property both in mean and volatility of stock market returns have been examined for some of D-8 countries. The findings obtained about efficiency of the stock market returns in these countries will be guiding in terms of investors, policy makers and academics.

**Key Words:** Long Memory, ARFIMA-FIGARCH models, Stock Market, Efficient Market Hypothesis

### TÜRKİYE'NİN GAYRİ SAFİ YURTIÇİ HASILASININ(GSYH) İSTATİSTİKSEL OLARAK MODELLENMESİNDE RİDGE PARAMETRELERİNİN UYGULANMASI

Anđan KARAİBRAHİMOĐLU<sup>1</sup>, Yasın ASAR<sup>2</sup>, Hakan BAŞBOZKURT<sup>3</sup>, Aşır GENÇ<sup>4</sup>

<sup>1</sup>Selçuk Üniversitesi, Fen Fakültesi, İstatistik AD(Doktora öğrencisi) (Necmettin Erbakan Üniversitesi)

Meram Tıp Fakültesi Tıp Eğitimi ve Bilişimi Anabilim Dalı Uzmanı) Konya, Türkiye email: [akara@konya.edu.tr](mailto:akara@konya.edu.tr)

<sup>2</sup> Necmettin Erbakan Üniversitesi Fen Fakültesi Matematik Bölümü Araştırma Görevlisi, Konya, Türkiye email: [vasar@konya.edu.tr](mailto:vasar@konya.edu.tr)

<sup>3</sup>Selçuk Üniversitesi, Fen Fakültesi, İstatistik AD (Doktora öğrencisi) Konya, Türkiye email: [hakan.basbozkurt@gmail.com](mailto:hakan.basbozkurt@gmail.com)

<sup>4</sup> Selçuk Üniversitesi, Fen Fakültesi, İstatistik AD Fen Bilimleri Enstitüsü Müdürü ve İstatistik AD Başkanı Konya, Türkiye email: [agenc@selcuk.edu.tr](mailto:agenc@selcuk.edu.tr)

Çoklu bağlantı, özellikle ekonometrik zaman serisi verilerinde sıkça karşılaştığımız bir sorundur. Çözümü için çok sayıda yöntem vardır ve ridge regresyonu bunlar arasında en yaygın kullanılan araçlardan birisidir. Ridge regresyonunda yanlış tahmin edici olan  $k$  parametresi analizin en önemli parçasıdır. Literatürde,  $k$  parametresi tahmin edicisi öneren çok sayıda yayın mevcuttur.

Bu çalışmada dünyadaki gelişme ve trendlere paralel olarak ve Türkiye'nin ekonomisini yakından ilgilendiren 1968 ve 2010 yılları arasındaki Harcama Yöntemiyle Gayri Safi Yurtiçi Hasıla(GSYH)-1987&1998 fiyatlarıyla değerlerini modelledik. Dış ticaret ve üretim değerlerine ait parametreleri kullanarak çoklu doğrusal regresyon ile GSYH'yi açıklamaya çalıştık. Fakat modelimizde çoklu bağlantı sorunu ortaya çıktı. Bu sorun çözülmezse uygun sonuçlar ve tahminler elde edilemez. Bu durumda genellikle,  $Y = X\beta + \varepsilon$  genel regresyon

denkleminin çözümü için gerekli olan  $X'X$  matrisinin bozuk olmasına bağlı olarak  $\beta$  yansız tahmin edicisi hesaplanamaz. Çalışmamızın amacı ridge regresyonu kullanarak bu sorunu gidermektir. Bu nedenle, daha önce önerilmiş olan bazı  $k$  parametrelerini karşılaştırarak en küçük hata kareler ortalamasına sahip olan parametreyi modele uygulayacağız.

**Anahtar kelimeler:** Çoklu bağlantı, Yansız tahmin edici, Ridge regresyonu, Ridge parametresi, simülasyon

### APPLICATION OF RIDGE PARAMETERS IN STATISTICAL MODELLING OF GROSS DOMESTIC PRODUCT (GDP) OF TURKEY

Multicollinearity is a commonly faced problem especially in econometric time series data. There are various methods for solving this problem and ridge regression is one of the mostly used tools among these methods. In ridge regression, estimating the biased parameter  $k$  is the most important part of the analysis. There are lots of papers proposing estimators of  $k$  parameter in the literature.

In this study, we modeled Gross Domestic Products (GDP) of Turkey by cost components (at 1987 & 1998 prices) between the years 1968 and 2010, closely concerning the economy of Turkey in parallel to the growth and trends in the world. We tried to explain GDP using some parameters in the axes of foreign trade and production by multiple linear regressions. However, there was a multicollinearity problem in the model. If this problem cannot be solved, then it cannot be achieved to compatible results. In this case, the unbiased estimator of  $\beta$ , generally, cannot be computed because of ill-conditioned matrix  $X'X$  in general multiple regression model  $Y = X\beta + \varepsilon$ . The purpose of our study is to overcome this problem using ridge regression. For this reason, we will compare some of the proposed estimators of  $k$  on this model and give the best one having the least mean squared error.

**Key words:** Multicollinearity, Unbiased estimator, Ridge regression, Ridge parameter, simulation

## TURİST TALEBİNE EKONOMETRİK BİR YAKLAŞIM: TÜRKİYE ÖRNEĞİ

Mehmet YILMAZ ve Buse BÜYÜM

Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü 06100 Tandoğan-Ankara, TÜRKİYE

### KAYNAKLAR

- [1] Ünsal, E. (2005), Makroekonomi, İmaj Yayıncılık, Ankara
- [2] Pamukoğlu, G. (1990), Import-Substitution Industrialization in Turkey, Massachusetts Institute of Technology
- [3] Alkhamisi M. A., Shukur G. (2008), "Developing Ridge Parameters for SUR Model" Communications in Statistics - Theory and Methods, 37:4, 544-564
- [4] Hoerl Arthur E., Kennard Robert W. (1970), "Ridge Regression: Biased Estimation for Nonorthogonal Problems", Technometrics, Vol. 12, No. 1, pp. 55-67
- [5] Dorugade, A.V. (2013), "New ridge parameters for ridge regression", Journal of the Association of Arab Universities for Basic and Applied Sciences <http://dx.doi.org/10.1016/j.jaubas.2013.03.005>

### Turizmin Tanımı, Dünyadaki Gelişimi ve Önemi

Turizm, yerli veya yabancı tüketicilerin bir ülke veya bölgeyi ziyareti sırasında duyacağı ihtiyaçları tatmin edecek hizmetlerin bütünüdür. Turizm dünyanın en hızlı büyüyen sektörlerinden biridir. Dünya turizm hareketleri, özellikle II. Dünya Savaşı'ndan sonra büyük bir ivme kazanmıştır. 1980-2000 yılları arasında, dünya turizm faaliyetleri toplam uluslararası ticaret hacmi içinde %15 oranında yüksek bir paya sahip olmuştur. 1990'lı yılların sonuna gelindiğinde hizmetler sektörü içinde turizm birinci sıraya yükselmiştir. 2000 yılında 683 milyon kişi turizm hareketine katılmış, turizm gelirleri ise 475 milyar dolar olarak gerçekleşmiştir. 2010 yılına gelindiğinde turist sayısı yaklaşık %38 artış ile 940 milyona, turizm gelirleri ise yaklaşık 2 kat artarak 919 milyar dolara ulaşmıştır. Bu nedenle turizm gelirleri, bir ülkenin dış ticaret gelirleri açısından büyük önem taşır ve istihdam ile gelir düzeyine uyurucu etkilerde bulunarak o ülkenin kalkınmasında doğrudan etki yaratır.

### Türkiye' de Turizmin Gelişimi

Türkiye'deki turizmin gelişim süreci, özellikle 1980 yılından sonra büyük bir gelişme göstermiş ve 1982 yılında çıkarılan "Turizmi Teşvik Kanunu" bu gelişim için bir dönüm noktası olmuştur. Bu teşvik tedbirlerinin sağlanması ile 1980-2000 yılları arasında işletme belgeli tesislere ait yatak sayısı 56.044 den 325.168 e yükselmiş ve bu süreçte, gelen turist sayısı da yaklaşık 8 kat artarak 1.288.000 den 10.412.000 e çıkmıştır. 2012 yılı sonunda ise işletme belgeli tesislere ait yatak sayısı 715.692 ve gelen turist sayısı da 31.782.832 ye çıkmıştır.

### Turist Sayısını Etkileyen Faktörler

Turizm sektöründeki gelişmeler birçok ülke için oldukça önemli olduğundan bu sektördeki yatırımların planlanması ve turistik işletmelerin kendilerini bir sonraki yıla daha iyi hazırlayabilmeleri için gelebilecek turist sayılarının öngörülebilmesi hayati bir önem taşır. Ülke bazında bir öngörü modeli belirleyebilmek için ülkenin coğrafi konumu, sosyal, politik ve teknolojik durumu, ülkenin mevsim yelpazesinin genişliği ve bu mevsimlerdeki kaynakların etkin kullanımı, kültürel ve tarihi zenginlikleri gibi faktörleri göz önüne almak gerekir. Bu faktörlere ek olarak, ülkenin reklam tanıtım ağı, diğer ülkelere olan uzaklığı, nisbi döviz kurları, (Bahar ve Kozak 2005) turizm teşviklerinin varlığı, turizm şirketlerinin sayısı ve işletme belgeli turistik tesis sayısı ile yatak kapasiteleri de ülkeye gelecek turist sayısı için önemli unsurlardandır.

### Literatürde Kullanılan Model Önerileri

Yaklaşık 30 yıldan beri turizm talebini modellemek ve öngörmek için araştırmacılar, parametrik ve parametrik olmayan yöntemler ile regresyon modelleri önerisinde bulunmuşlardır. 2000 yılından günümüze indekslerde taranan 130 u aşkın çalışmanın turizm talebi ve bu talebin öngörülmesi ile ilgili olduğu görülmüştür. Bu çalışmalarda genellikle lineer ve lineer olmayan dinamik modeller ( Otoregresif (AR), Entegre edilmiş Otoregresif Hareketli Ortalama (ARIMA) ve Mevsimsel Otoregresif (SAR) ) kullanılmıştır (Song ve Li, 2008). Turistik zenginlikleri olan ülkelerde turizm gelirleri önemli yer tuttuğundan turist sayısının ve turizm gelirlerinin ülke bazında modellenmesi önemlidir. Örneğin, Song et al. (2010) çalışmalarında Hong-Kong' a gelen turist sayısını modellemek için gelen turistlerin kişi başı milli gelirleri ve harcadıkları para vs. gibi değişkenlerle hem statik hem de dinamik model önerisinde bulunmuşlardır. Chen (2011) turist talebini öngörmek için lineer ve lineer olmayan modelleri birleştirmiştir.

### Öngörü Modelinde Kullanılacak Materyal ve Yöntem

Bu çalışmada, 1966-2012 yılları arasındaki Türkiye' ye gelen turist sayıları ve işletme belgeli tesislere ait yatak sayıları baz alınarak turist sayısının öngörülmesi amaçlanmıştır. Öngörü döneminde işletme belgeli tesislere ait yatak sayıları da bilinmediğinden bu sayıları da öngörecektir zaman serisi ve üstel düzeltirme modeli önerilerinde bulunulmuştur. Temel iktisadi yasaya göre, hizmetin belirli bir eşik noktası vardır ve bu noktadan sonra hizmet ne kadar artarsa artsın, gelen turist sayısı bir doyum noktasına ulaşacaktır. Bu yasayı sinayacak modeller literatürde genellikle ters model olarak adlandırılırlar. Yapılan çalışmada, öncelikle ters modeller ele alınmış daha sonra bunlara ek olarak karma modeller önerilmiştir. Bu modeller içinde ekonometrik sorunları olan modeller incelenmiştir. Son olarak, ex-post öngörü ile RMSE değerleri hesaplanıp öngörülecek model belirlenerek, 2013-2020 yılları arasında Türkiye'ye gelebilecek turist sayıları öngörülmüştür.

## KAYNAKLAR

- [1] Chen K.Y. (2011), Combining Linear and Nonlinear Model In Forecasting Tourism Demand, Expert Systems with Applications, 38, 10368–10376.
- [2] Song H. and Witt S.F. (2000), Tourism Demand Modelling and Forecasting: Modern Econometric Approaches , Oxford: Pergamon.
- [3] Song H. and Li G. (2008), Tourism Demand Modelling and Forecasting, A Review of Recent Research Tourism Management, 29 (2). 203 - 220. ISSN 0261-5177.
- [4] Song H. (2010), Tourism Demand Modelling and Forecasting: How Should Demand Be Measured?, Tourism Economics, 16 (1), 63–81.

## ABSTRACT

### AN ECONOMETRIC APPROACH TO TOURISM DEMAND: EVIDENCE FROM TURKEY

The world countries concentrate on planning tourism policy because the tourism sector is developing rapidly day by day. Therefore, forecasting is an essential analytical tool in tourism policy. This study focuses on modelling tourism demand with using number of tourists who come to Turkey in 1966-2012 and number of beds which belong to accommodation facilities with tourism operation licensed. In this sense, forecasting of international tourism arrivals to Turkey for 2013-2020 is purposed. Primarily inverse models and then linear and nonlinear dynamic models and their mixtures are discussed. In order to improve forecast accuracy, error terms of suggested models are examined in terms of model assumptions.

**Key Words:** Tourism demand; tourism demand modelling; forecasting; inverse models; time series models, exponential smoothing models

## TÜRKİYE'DE EKONOMİK DEĞİŞKENLERİN TÜKETİCİ GÜVEN ENDEKSİ ÜZERİNDEKİ ETKİSİ

Feyza GÜNAY\*, Abdullah YALÇINKAYA, Mustafa GAZİOĞULLARI

Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06100, Ankara, TÜRKİYE

[fgunay@ankara.edu.tr](mailto:fgunay@ankara.edu.tr) , [ayalcinkaya@ankara.edu.tr](mailto:ayalcinkaya@ankara.edu.tr)

### Tüketici Güven Endeksi ve Ekonomik Değişkenlerin Bazıları

Tüketici Güven Endeksi (TGE); ülke ekonomisinin anlık durumu hakkında tüketicilerin görüşlerini ve gelecek ile ilgili beklentilerini gösteren ulusal bir ekonomik göstergedir. Türkiye'de TGE, Türkiye İstatistik Kurumu (TÜİK) tarafından aylık olarak yapılan Tüketici Eğilim Anketleri sonucunda elde edilmektedir. Endeksin 100'den büyük olması; tüketici güveninde iyimser durum, 100'den küçük olması; tüketici güveninde kötümser durum, 100 olması ise; tüketici güveninde ne iyimser ne de kötümser durum olduğunu göstermektedir.

Ülke ekonomisinin tüketicilerin görüş ve gelecek beklentilerinde etkili olduğu bilindiğinden TGE'nin de bazı ekonomik değişkenlerden etkileneceği aşikardır. Bu ilişkiyi incelemek için ele alılabilecek ekonomik değişkenler sırasıyla; DOLAR, EURO, İMKB, TUF, UFE, NÜFUS, GSYİH, vs. olarak verilebilir. Bu değişkenlerin bazıları aşağıdaki gibi tanımlanabilir.

Tüketici Fiyatları Endeksi (TÜFE); bireylerin ortalama tüketim kalıplarını yansıtan ürün ve hizmetlerden oluşan sepetin aylık dönemler itibarıyla fiyat değişimini temsil eder. Sepette yer alan mal ve hizmetlerin, miktar ve kalite değişimleri göz önüne alınarak endeksin sadece fiyat hareketlerini yansıtmaması sağlanır. Üretici Fiyatları Endeksi (ÜFE); genel olarak yurt içi üretimden satışa konu olan seçilmiş ürünlerin, üretici fiyatlarındaki değişiminin bir göstergesidir.

Gayri Safi Yurt İçi Hâsıla (GSYİH); bir ülkenin ekonomik büyüklüğünün birkaç ölçütünden biridir. GSYİH, Gayri Safi Milli Hasıladan farklı olarak, bir ülke sınırları içerisinde belli bir zaman içinde, üretilen tüm nihai mal ve hizmetlerin para birimi cinsinden değeridir. Bu tanımda belli bir zaman, bir ay, üç ay ya da bir yıl olabilir. GSYİH genellikle bir yıl için ele alınır. Nihai mal ve hizmetler ise, üretilen toplam mal ve hizmetlerden üretim için kullanılan ara mallar düşüldükten sonra geriye kalan değerdir. GSYİH verilerini bulmak için şu formül uygulanır.  $GSYİH = \text{tüketim} + \text{yatırım} + \text{devlet harcamaları} + (\text{ihracat} - \text{ithalat})$

İstanbul Menkul Kıymetler Borsa Endeksi (İMKB); Belli bir günde İstanbul Menkul Kıymetler borsasındaki menkul değerlerin fiyatlarındaki değişim göstergesidir.

### Amaç ve Yöntem

Bu çalışmada amaç, Türkiye İstatistik kurumunun anket ile elde ettiği Tüketici Güven Endeksini farklı istatistiksel yöntemlerle elde edebilmektir. Bunun için Tüketici Güven Endeksini açıklayabilecek Ekonomik değişkenlerle en uygun modelin elde edilmesi problemi ele alınmıştır.

Bu çalışmada, 2004-2012 yılları arasında Türkiye İstatistik Kurumu tarafından aylık olarak yapılan anketler sonucunda hesaplanan TGE verilerinin Türkiye'deki ekonomik değişkenlerden nasıl etkilendiğini araştırmak için aylık Dolar, Euro, İmkb, Tufe, Ufe, Gsyih ve Nüfus verileri toplanmıştır. Tüketici Güven Endeksi bağımlı değişken olarak ele alınmış ve TGE'yi açıklamak için en uygun model çoklu regresyon yöntemi ile oluşturulmaya çalışılmıştır. Ayrıca bu süreçte çoklu doğrusallık sorunu olmamasına, hataların standart sapmasının en küçük ve modelin açıklama yüzdesinin en büyük olmasına dikkat edilmiştir. En uygun model belirlendikten sonra Ardışık Bağımlılık ve Değişen Varyans sorunlarının varlığı tespit edilerek varsa düzeltici önlemlerle yok edilmiştir. İleriye dönük öngörülerde bulunmak için kestirim denklemleri belirlenmiştir. Analizler çeşitli bilgisayar programları aracılığıyla yapılmış ve sonuçları değerlendirilmiştir.

## KAYNAKLAR

- [1] Gujarati D. N. (1978), Basic Econometric, New York, McGraw-Hill.
- [2] Türkiye İstatistik Kurumu (TÜİK), <http://www.tuik.gov.tr/>
- [3] Türkiye Cumhuriyeti Merkez Bankası (TCMB), <http://www.tcmb.gov.tr/>

## ABSTRACT

**THE EFFECT OF ECONOMIC VARIABLES IN TURKEY ON CONSUMER CONFIDENCE INDEX**

Consumer Confidence Index (TGE) that is calculated every month with some survey about consumer's confidence by Turkish Statistical Institute (TUIK) is affected from economical variables in Turkey. So some variables are taken like Dollar, Euro, İstanbul Stock Exchange (İMKB), Consumer Price Index (TUFİE), Producer Price Index (UFE),Gross Domestic Product (GSYİH), Population. Thus TGE can be predicted using multiple regression model without necessity of costly survey studies. Fit and high R-Squared value prediction equations are found without multicollinearity, heteroscedasticity, autocorrelation problems.

**Key Words:** Consumer Confidence Index, Multiple Regression, Prediction Equation, Economic Variables

**BİLDİRİ OTURUMLARI 3**

*SESSION 3*

*Statistics Theory 2*

**ROBUST ESTIMATORS FOR THE SHAPE PARAMETERS OF BURR XII DISTRIBUTION**

Fatma Zehra DOĞRU<sup>1\*</sup>, Olcay ARSLAN<sup>1</sup>

Ankara University, Faculty of Science, Department of Statistics, 06100 Ankara/Turkey<sup>1</sup>

fzdogru@ankara.edu.tr, oarslan@ankara.edu.tr

**Introduction**

In general, the parameters of a given distribution are estimated by classical methods like maximum likelihood and least squares. These estimators are very sensitive to outliers. So the robust estimators are suggested as an alternative to these estimators. We consider the Burr XII distribution which was first introduced by Burr (1942). This distribution is widely used in areas such as business, engineering, reliability and hydrology as a failure model. Its cumulative density function (cdf) and probability density function (pdf) are given as follows:

$$F(x) = 1 - \frac{1}{(1+x^c)^{k+1}}, x \geq 0, c > 0, k > 0 \quad (1)$$

and

$$f(x) = ck \frac{x^{c-1}}{(1+x^c)^{k+1}}, x \geq 0, c > 0, k > 0, \quad (2)$$

where  $c$  and  $k$  are shape parameters.

The shape parameters of Burr XII distribution were estimated by using several different methods. Hossain and Nath (1997) compared maximum likelihood, least square and maximum product of spacing estimation methods in presence and absence of outliers. Robust regression method to estimate the parameters of Burr XII distribution for complete and multiply-censored data with outliers has been given by Wang and Cheng (2010). Also, B-robust estimation method for the shape parameters of this distribution has been studied by Doğru and Arslan (2013).

In this study we propose robust estimation methods to estimate the shape parameters of the Burr XII distribution and compare the proposed estimators with the estimators given in literature. The least squares estimation method which is obtained by minimizing the following objective function with respect to parameters:

$$\min_{c,k} \sum_{i=1}^n \epsilon^2 = \sum_{i=1}^n (y_i - \log k - \log \log(1+x_i^c))^2. \quad (3)$$

Similar to the least squares estimation method, robust estimation method for the parameters of Burr XII distribution will be obtained by minimizing the following objective function:

$$\min_{c,k} \sum_{i=1}^n \rho(y_i - \log k - \log \log(1+x_i^c)). \quad (4)$$

We will use two  $\rho$  functions like bisquare and Huber's functions:

$$\rho(x) = \begin{cases} 1 - (1 - (x/b)^2)^3, & |x| \leq b \\ 1, & |x| > b \end{cases} \quad (5)$$

and

$$\rho(x) = \begin{cases} x^2, & |x| \leq b \\ 2b|x| - b^2, & |x| > b \end{cases} \quad (6)$$

We present a small simulation study to show the performance of these estimation methods in absence and presence of outliers. Also we use a real data example to demonstrate the performance of the estimators based on robust estimators.

**REFERENCES**

- [1] Burr, I.W. (1942), Cumulative frequency functions, *Annals of Mathematical Statistics*, 13, 215-232.
- [2] Doğru, F.Z. and Arslan, O. (2013), B-Robust estimation method for the parameters of Burr XII distribution, 14th International Symposium on Econometrics Operation Research and Statistics-(ISEOS), Sarejevo.
- [3] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.
- [4] Hossain, A.M. and Nath, S.K. (1977), Estimation of parameters in the presence of outliers for a Burr XII distribution, *Commun. Statist. Theory Methods* 26, 813-827.
- [5] Wang, F.K. and Cheng, Y.F. (2010), Robust regression for estimating the Burr XII parameters with outliers, *Journal of Applied Statistics*, 37, 5, 807-819.

**Key Words:** Burr XII distribution, B-robust, robust estimators, least squares

ON MARGINAL DISTRIBUTION OF CONDITIONAL LARGEST ORDER STATISTICS FROM BIVARIATE SAMPLE

Gülder KEMALBAY

Yıldız Technical University, Faculty of Art&Science, Department of Statistics, 34220, Istanbul, TURKEY, kemalbay@yildiz.edu.tr

Introduction

Suppose that  $X_1, \dots, X_n$  be a random sample from a population with distribution function  $F_X(x)$ . The corresponding order statistics of this sample are obtained by arranging  $X_i$ 's in nondecreasing order. In literature, a lot of attention has been devoted to the study of order statistics. For a comprehensive study for theory and applications, we refer to [1] and [4], among others. Although the order statistics attracted the attention of several authors, many of results dealing with order statistics have been derived for the univariate sample case.

In this study, a bivariate random sample  $(X_1, Y_1), \dots, (X_n, Y_n)$  with joint distribution function  $F(x, y)$  is considered. Let us assume that  $X_1, \dots, X_n$  and  $Y_1, \dots, Y_n$  be independent and identically distributed random variables with distribution functions  $F_X(x)$  and  $F_Y(y)$ , respectively. Additionally,  $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ ,  $Y_{1:n} \leq Y_{2:n} \leq \dots \leq Y_{n:n}$  be the corresponding marginal order statistics. Let  $X_{r:n}$  and  $Y_{s:n}$  be the  $r^{th}$  and  $s^{th}$  order statistics and their marginal distribution functions are obtained by univariate binomial distribution as follows:

$$F_{X_{r:n}}(x) \equiv P\{X_{r:n} \leq x\} = \sum_{i=r}^n \binom{n}{i} F_X(x)^i [1 - F_X(x)]^{n-i}, \quad (1)$$

$$F_{Y_{s:n}}(y) \equiv P\{Y_{s:n} \leq y\} = \sum_{j=s}^n \binom{n}{j} F_Y(y)^j [1 - F_Y(y)]^{n-j}. \quad (2)$$

The joint distribution of  $X_{r:n}$  and  $Y_{s:n}$  can be derived from bivariate binomial distribution if one considers the fourfold model with  $A = \{X_i \leq x\}$  and  $B = \{Y_i \leq y\}$ . If  $\xi$  and  $\eta$  denote the number of occurrences of events  $A$  and  $B$  in  $n$  independent trials of the fourfold experiment, respectively, then the joint distribution of bivariate order statistics is given by

$$\begin{aligned} F_{X_{r:n}, Y_{s:n}}(x, y) &\equiv P\{X_{r:n} \leq x, Y_{s:n} \leq y\} \\ &= \sum_{i=r}^n \sum_{j=s}^n P\{\xi = i, \eta = j\} = \sum_{i=r}^n \sum_{j=s}^n \sum_{k=a}^b C(n; i, j, k) \\ &\times P(AB)^k P(AB^c)^{i-k} P(A^c B)^{j-k} P(A^c B^c)^{n-i-j+k} \end{aligned} \quad (3)$$

where  $C(n; i, j, k) = n! / k!(i-k)!(j-k)!(n-i-j+k)!$ ;  $a = \max(0, i+j-n)$ ;  $b = \min(i, j)$  and probabilities are given as follows:  $P(AB) = F(x, y)$ ;  $P(AB^c) = F_X(x) - F(x, y)$ ;  $P(A^c B) = F_Y(y) - F(x, y)$ ;  $P(A^c B^c) = 1 - F_X(x) - F_Y(y) + F(x, y)$ .

For more details, one can see [4].

Recently, by using some modifications of bivariate binomial distribution, the distribution function of bivariate order statistics  $(X_{r:n}, Y_{s:n})$  under condition that a certain number of observations are truncated, i.e. they fall in the set  $\mathbf{B}_{uv} = \{(t, s) \in \square^2 : t \leq u, s \leq v\}$ , have been obtained by [3]. The conditional distribution of bivariate order statistics is denoted by  $F_{r,s:n}(x, y|u, v)$ . Furthermore, by using the properties of extreme order statistics  $(X_{n:n}, Y_{n:n})$ , a different approach is presented for obtaining the conditional distribution function of bivariate largest order statistics, which is denoted by  $F_{n,n:n}(x, y|u, v)$ .

In this study, the marginal distribution functions of conditional bivariate extreme order statistics  $(X_{n:n}, Y_{n:n})$  are derived by taking the limit of conditional joint distribution function  $F_{n,n:n}(x, y|u, v)$  as  $x \rightarrow \infty$  and  $y \rightarrow \infty$ , respectively. In the case of underlying distribution is Farlie-Gumbel-Morgenstern (FGM), the corresponding dependence analysis of conditional largest order statistics involving association parameter  $\alpha$  is studied. Moreover, some numerical and graphical results are illustrated. These obtained results can find application for statistical modelling of extreme values in risk management, finance and insurance as well as reliability theory.

REFERENCES

- [1] Arnold B.C., Balakrishnan N. and Nagaraja H.N. (2008), A First Course In Order Statistics, USA, Siam.
- [2] Bairamov I. (2013), Reliability and Mean Residual Life of Complex Systems with Two Dependent Components per Element, IEEE Transactions on Reliability, 62(1), 276-285.
- [3] Bairamov I. and Kemalbay G. (2013), Some Novel Discrete Distributions Under Fourfold Sampling Schemes and Conditional Bivariate Order Statistics, Journal of Computational and Applied Mathematics, 248, 1-14.
- [4] David H.A. (1981), Order Statistics, Second Edition, New York, John Wiley & Sons.

ÖZET

İKİ DEĞİŞKENLİ ÖRNEĞE AİT KOŞULLU MAKSİMUM SIRA İSTATİSTİKLERİNİN MARJİNAL DAĞILIMLARI

Bu çalışmada, koşullu iki değişkenli maksimum sıra istatistiklerinin dağılımı incelenmiş ve marjinal dağılımları elde edilmiştir. Farlie-Gumbel-Morgenstern (FGM) dağılımı varsayımı altında bazı sayısal ve grafiksel sonuçlar sunulmuştur.

**Anahtar Kelimeler:** Koşullu Sıra İstatistikleri, Ekstrem Değerler, Marjinal Dağılımlar, Bağımlılık.

NEGATIVE CUMULATIVE DEPENDENCE OF RANDOM VARIABLES AND APPLICATIONS

Ümit IŞLAK, Deniz Topuz\*

University of Southern California, College Of Letters, Arts and Sciences, Department of Mathematics, 3620 S Vermont Ave, Los Angeles, CA, 90089 , umitislak@hotmail.com

Bahçeşehir Üniversitesi, Fen Edebiyat Fakültesi, Matematik-Bilgisayar Bölümü, 34510, İstanbul, TÜRKİYE, denizz.t@gmail.com

In many statistical problems, the standard assumption that the underlying random variables are independent is not plausible. Increases in some variables are often related to decreases or increases in other variables. To understand this phenomenon partially, the concept of negative association of random variables was introduced in [4] and simple properties of this family were analyzed. Formally, a finite collection of random variables  $\{X_1, X_2, \dots, X_n\}$  is said to be negatively associated (NA) if for any disjoint subsets  $A_1, A_2 \subset \{1, \dots, n\}$ ,

$$E[f(X_i; i \in A_1)g(X_j; j \in A_2)] \leq E[f(X_i; i \in A_1)]E[g(X_j; j \in A_2)]$$

whenever  $f$  and  $g$  are coordinate-wise nondecreasing functions for which the expectations exist. The purpose of this study is to analyze another sort of dependence, negative cumulative dependence, that generalize association of random variables.

Let  $X_1, X_2, \dots, X_n$  be random variables and  $S_k = \sum_{i=1}^k X_i$  for  $k = 1, 2, \dots, n$ .  $X_1, X_2, \dots, X_n$  are said to be negatively cumulative dependent (NCD) if

$$E[f(S_{k-1})g(X_k)] \leq E[f(S_{k-1})]E[g(X_k)] \text{ for } k = 2, \dots, n$$

whenever  $f$  and  $g$  are nondecreasing functions for which the expectations exist. NCD random variables were first introduced [3] where they have observations about Poisson approximation under negative dependence assumptions. Here we will obtain some properties of NCD variables and consider new applications. We start with the following proposition which summarizes the relation between NCD and NA.

**Proposition 0.1. i)** NCD and NA are equivalent for two random variables  $X_1$  and  $X_2$ .

**ii)** If  $X_1, X_2, \dots, X_n$  are NA, then they are also NCD.

**iii)** There exists  $X_1, X_2, \dots, X_n$  that are NCD, but not NA.

Sketch of proof: Proofs of i. and ii. easily follows from the definitions. We give a counter example for iii. Let  $X_1, X_2, X_3$  be uniformly distributed over  $\{(0, c, c), (c, 0, 0), (c, c, 0), (c, 0, c), (c, c, c)\}$  for some  $c > 0$ . It can be checked that  $X_1, X_2, X_3$  are NCD, but  $\text{Cov}(X_2, X_3) = c^2/25 > 0$  so that they are not NA.

Thus, NCD forms a strictly larger class than NA. Proofs regarding applications of NCD usually have an inductive nature as can be felt from the very definition of this family. Thus, the observation made in item i. of Proposition 0.1 is important in many problems for the basis step. Also note that item ii. in Proposition 0.1, in particular implies that NCD also generalizes independence. One other useful observation about NCD random variables is stated in the following lemma.

**Lemma 0.2.** If  $X_1, X_2, \dots, X_n$  are NCD, then for any  $\theta > 0$ ,

$$E[e^{\theta \sum_{k=1}^n X_k}] \leq \prod_{k=1}^n E[e^{\theta X_k}].$$

Again the proof of Lemma 0.2 is inductive and we skip the details. This lemma is very useful as it gives an upper bound on the moment generating function of  $S_n = \sum_{i=1}^n X_i$ . In particular, for  $\theta > 0$ , letting  $t \geq 0$  and  $\mu = E[S_n]$ , an application of Markov's inequality gives

$$(0.1) \quad P(S_n - \mu \geq t) = P(e^{\theta S_n} \geq e^{\theta(\mu+t)}) \leq \frac{E[e^{\theta S_n}]}{e^{\theta(\mu+t)}} \leq \frac{\prod_{k=1}^n E[e^{\theta X_k}]}{e^{\theta(\mu+t)}}$$

Using (0.1) and standard estimation techniques, one can now obtain concentration inequalities (such as Bernstein) for NCD random variables. NCD random variables also satisfy the following convex stochastic ordering result.

**Theorem 0.3** Let  $\{X_1, X_2, \dots, X_n\}$  be a sequence of NCD random variables and also let  $\{X'_1, X'_2, \dots, X'_n\}$  be a sequence of independent random variables so that  $X_i$  and  $X'_i$  have the same distribution for each  $i$ . Then

$$E[f(\sum_{i=1}^n X_i)] \leq E[f(\sum_{i=1}^n X'_i)]$$

for any convex function on  $\mathbb{R}$ , whenever the expectations on the right hand side exists.

Proof of this result is lengthy, uses induction, and properties of convex functions and NCD random variables. Instead of including its proofs, we mention two special cases : (i)  $f(x) = e^{\theta x}$  for some  $\theta > 0$  and (ii)  $f(x) = x^r$  for some  $r > 0$ . The first special case allows one to estimate the moment generating function of  $S_n$  using the moment generating function of sum of independent copies of  $X_i$ 's as we discussed above. The second one enables one to obtain useful moment inequalities for  $S_n$ . In particular, with the notations of Theorem 0.3, one has  $E[(\sum_{i=1}^m X_i)^m] \leq E[(\sum_{i=1}^m X'_i)^m]$  for which the right hand side can be manipulated using elementary properties of independent random variables.

Finally we note that one other use of NCD random variables is in distributional approximations. In particular, it is possible to extend "various results from Stein's method" for Poisson and normal approximation to the case of NCD variables. Here we will not be discussing this aspect, but we remind that one might need exchangeability besides NCD in some of these problems. Although this is restrictive, it is still important as many applications involve exchangeable random variables.

REFERENCES

[1] Dubhashi, D.P. and Ranjan, Balls and bins: a study in negative dependence, Random Structures Algorithms 13, no. 2, 99-124, 1998  
 [2] Pemantle, Robin, Towards a theory of negative dependence, J.Math. Phys.41,no. 3, 1371-1390,2000  
 [3] Boutsikas, Michael V. And Koutras, Markos V., A bound for the distribution of the sum of discrete associated or negatively associated random variables, Ann. Appl.Probab.10, no.4, 1137-1150, 2000  
 [4] Joag-Dev, K. And Proschan, Negative association of random variables, with applications, Ann. Statist. 11, no. 1,286-295, 1983  
 [5] Papadatos, N and Papathanasiou, V., Poisson approximation for a sum of dependent indicators: an alternative approach, Adv. in Appl.Probab. 34, no. 3, 609-625,2002



AN IMPROVED NEW CLASS OF EXPONENTIAL RATIO ESTIMATORS FOR POPULATION  
MEDIAN IN SIMPLE RANDOM SAMPLING

Sibel ALADAG\*, Hulya CINGI

\*General Director of Service Provision, Republic of Turkey Social Security Institution, Bakanlıklar, Ankara, Turkey, e-mail: saladag@sgk.gov.tr  
University of Hacettepe, Faculty of Science, Department of Statistics, Beytepe, Ankara, Turkey, e-mail: hcingi@hacettepe.edu.tr

1. Introduction

Median is often regarded as more appropriate measure of location than mean when variables have a highly skewed distribution, such as income, expenditure, production are studied in survey sampling. In literature, there have been many studies for estimating the population mean and population total but relatively less effort has been devoted to the development of efficient methods for estimating the population median.

In simple random sampling, Gross (1980), defined sample median and obtained  $\tilde{M}_Y$  is consistent and asymptotically normal with mean  $M_Y$  and variance

$$V(\tilde{M}_Y) = \lambda M_Y^2 C_{M_Y}^2 \quad (1)$$

where  $\lambda = (1 - f)/4n$  and  $f = n/N$ .

Kuk and Mak (1989), suggested a ratio estimator and obtained the MSE equation

$$\tilde{M}_{YR} = M_Y \frac{M_X}{\tilde{M}_X} \quad (2)$$

$$MSE(\tilde{M}_{YR}) = \lambda M_Y^2 \{C_{M_Y}^2 + C_{M_X}^2 - 2C_{M_X} C_{M_Y} \rho_{M_X M_Y}\} \quad (3)$$

where  $\rho_{M_X M_Y}$  is the correlation coefficient between sampling distribution of  $\tilde{M}_X$  and  $\tilde{M}_Y$  which is defined as  $\rho_{M_X M_Y} = 4P_{11} - 1$  and  $P_{11}$  is the proportion of units in the population with  $X \leq M_X$  and  $Y \leq M_Y$ . Kuk and Mak (1989) defined a matrix of proportions ( $P_{ij}$ ) seen in Table 1.

Table 1: Matrix of Proportions

	$Y \leq M_Y$	$Y > M_Y$	Total
$X \leq M_X$	$P_{11}$	$P_{21}$	$P_{.1}$
$X > M_X$	$P_{12}$	$P_{22}$	$P_{.2}$
Total	$P_{.1}$	$P_{.2}$	$1$

Aladag and Cingi (2012), make the first contribution in using exponential estimator for estimating the population median. Aladag and Cingi (2012) propose following estimators and obtain the MSE equations given in the below:

$$t_1 = \tilde{M}_Y \exp\left(\frac{M_X - \tilde{M}_X}{M_X + \tilde{M}_X}\right) \quad (4)$$

$$MSE(t_1) \cong \lambda M_Y^2 \left( C_{M_Y}^2 + \frac{C_{M_X}^2}{4} - C_{M_Y} C_{M_X} \rho_{M_X M_Y} \right) \quad (5)$$

$$t_2 = \tilde{M}_Y \exp\left(\frac{(\eta M_X + \zeta) - (\eta \tilde{M}_X + \zeta)}{(\eta M_X + \zeta) + (\eta \tilde{M}_X + \zeta)}\right) \quad (6)$$

$$MSE(t_2) \cong \lambda M_Y^2 \left( C_{M_Y}^2 + \theta^2 C_{M_X}^2 - 2\theta C_{M_Y} C_{M_X} \rho_{M_X M_Y} \right) \quad (7)$$

2. Suggested Estimators

Following Singh et. al. (2009), we suggest new exponential ratio estimators for population median. We define an improved new class of exponential ratio estimators for population median and derive the minimum mean square error (MSE) equation of the proposed estimator. We compare MSE equations and find theoretical conditions which make each proposed estimator more efficient than the others.

3. Conclusion

We define an improved new class of exponential ratio estimators for population median and develop some exponential ratio estimators for population median using some known value of the population parameters. We find theoretical conditions which make each proposed estimator more efficient than the others. We also support these conditions by using numerical example.

REFERENCES

- [1] Aladag, S., Cingi H., (2012), A New Class of Exponential Ratio Estimators for Population Median in Simple Random Sampling, 8<sup>th</sup> International Symposium of Statistics, 11-13 October, Eskisehir, Turkey.
- [2] Gross, T. S., (1980). Median estimation in sample surveys, Proc. Surv. Res. Meth. Sect. Amer. Statist. Ass. 181-184.
- [3] Kuk, A. Y. C., Mak, T. K., (1989). Median estimation in the presence of auxiliary information, Journal of the Royal Statistical Society Series, B, 51, 261-269.
- [4] Singh, R., Chauhan, P., Sawan, N., Smarandache, F., (2009), Improvement in Estimating the Population Mean Using Exponential in Simple Random Sampling, Bulletin of Statistics & Economics, 3 (A09), 13-19.

ABSTRACT

Singh et. al. (2009) suggested a ratio-type exponential estimator for estimating the finite population mean. Following Singh et. al. (2009), we define an improved new class of exponential ratio estimators for population median and derive the minimum mean square error (MSE) equations of the proposed estimators for constrained and unconstrained choice of  $\alpha_1$  and  $\alpha_2$ . We compare MSE equations and find theoretical conditions which make each proposed estimator more efficient than the others given in literature. These conditions are also supported by using numerical examples.

**Key words:** Auxiliary information, exponential estimator, median estimation, simple random sampling.

**ROBUST ESTIMATORS FOR THE DISTRIBUTED LAG MODEL BASED ON t AND SKEW t DISTRIBUTIONS**

Yetkin TUAÇ, Olcay ARSLAN, Mehmet Niyazi ÇANKAYA  
Döğol cad. Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü, 06100 Tandoğan/ANKARA  
ytuac@ankara.edu.tr, oarslan@ankara.edu.tr, mncankaya@ankara.edu.tr

**Introduction**

All distributed lag equations state that a dependent variable Y is determined by a weighted sum of past values of an independent variable X:

$$y_t = \sum_{i=0}^p \beta_i x_{t-i} + u_t \quad t=1, 2, \dots, n. \quad (1)$$

When  $u_t$  are  $IN(0, \sigma_u^2)$   $\beta$  vector can be estimated by using Maximum Likelihood Estimation method, this also gives the LS estimates.  $t_v(0, \sigma_u^2, v)$  can be assumed for estimating  $\beta$  vector as an heavier tailed alternative to the normal distribution. If the data have skewness skew t-distribution can be used for modelling errors. In this study we will model  $u_t$  with the Jones and Faddy skew t-distribution  $t_{a,b}(0, \sigma_u^2, a, b)$ . We will model  $u_t$  with t and skew t-distributions to estimate  $\tilde{\beta}$ . The pdf of Jones and Faddy skew t-distribution  $(t_{a,b}(0, \sigma_u^2, a, b))$  is

$$f_{JFst}(u_t) = \frac{1}{B(a, b) 2^{a+b-1} \sqrt{a+b}} \left[ 1 + \frac{u_t}{\sqrt{a+b+u_t^2}} \right]^{a+1/2} \left[ 1 - \frac{u_t}{\sqrt{a+b+u_t^2}} \right]^{b+1/2}. \quad (2)$$

Secondly there is a special form of the distributed lag model which is Almon Model defined as below. Basicly the procedure is based on the assumption that the  $\beta_i$  in (1) lie on a low degree polynomial. We will assume that it is a quadratic as defined below

$$\beta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2. \quad (3)$$

Using this, equation (1) can be expressed as

$$y_t = \sum_{i=0}^p (\alpha_0 + \alpha_1 i + \alpha_2 i^2) x_{t-i} + u_t \quad (4)$$

$$= \alpha_0 z_{0t} + \alpha_1 z_{1t} + \alpha_2 z_{2t} + u_t, \quad (5)$$

where

$$z_{jt} = \sum_{i=0}^p i^j x_{t-i}. \quad (6)$$

Then,  $\beta$  parameter can be also estimated under the same assumptions used on the error of distributed lag model given in the first part above.

**REFERENCES**

- [1] Almon S. (1965), The Distributed Lag Between Capital Appropriations and Expenditures, *Econometrica*, Vol. 33, No. 1 pp. 178-196.
- [2] Arslan O. (2011), A Review on the Univariate Skew t- Distributions, *Far East Journal of Theoretical Statistics* Vol.34 No. 1 pp. 17-34.
- [3] Çabuk H.A. and Akdeniz F. (2012), An Application of Generalized Maximum Entropy and Some Biased Estimation Methods For The Almon Distributed Lag Model: Bootstrap Efficiency, *Advances and Applications in Statistics*, Vol. 31, Issue 2, pp. 103-124.
- [4] Jones M.C. and Faddy M.J (2003), A Skew Extension of The t-distribution, with Applications, *J. Roy. Statist. Soc. Ser. B* 65, 159-174.
- [5] Maddala G.S (1974), Ridge Estimators for Distributed Lag Models, Working Paper No.69.

**Key Words:** Distributed lag models, Almon Model, Skew t-distributions.

THE IMPACT OF REPLACEMENT MIGRATION POLICY ON DEMOGRAPHIC INDICATORS  
AND LABOR FORCE IN TURKEY

Osman Nuri ERDEM Doc. Dr. A. Sevtap SELCUK-KESTEL

T.C. Kalkınma Bakanlığı  
Necatibey Cad. No.110/A  
06100 Yücetepe ANKARA  
onerdem@kalkinma.gov.tr

Orta Dogu Teknik Üniversitesi  
Uygulamalı Matematik Enstitüsü  
06800 ANKARA  
kestel@metu.edu.tr

BİLDİRİ OTURUMLARI 3

SESSION 3

*Applied Statistics 1*

**Extended Summary**

Population decline and population aging are two particular demographic trends seen over developed countries because of rising life expectancy followed by falling fertility rates, and by this demographic transition process, will be seen in Turkey and other countries as well. By 2040, under the base scenario of no fertility and no migration policy, the demographic window of opportunity which refers to the time period in which the proportion of the working age population is particularly prominent will be closed in Turkey. Under the same scenario, by 2050, Turkey's population will stop to increase and start to shrink with an older age structure. In this context, replacement migration is needed to compensate declines in the size of population, population of the working age and to delay the aging of the population.

In this study, it is examined that to what extent can migration policies be a solution to the future aging problem of Turkey, in terms of the age structure of the population, labor force participation and employment levels, comparing to the effects of possible fertility increasing policies. The aging and projected decline of populations of developed countries have lead to attempts to discover the extent to which increasing migration could change the trends under five different scenarios with regard to the international migration streams (UN, 2000). Lutz and Scherbov (2003) combine 7 different fertility assumptions with 4 different migration assumptions at the level of EU-15 by 2050. Since the studies on the effects of the combined migration and fertility policies on the population age structure, labor force and employment levels by using a population projection model with up-to-date data for Turkey are scarce, the main motivation of this study is to find out relevant policy options for the solution of future aging problem of Turkey with the mentioned tools and data.

Base year population size and age structure, life expectancy, fertility and net migration rate data for Turkey and other countries, which are retrieved from TURKSTAT, OECD, World Bank and CIA Factbook statistics, are used for taking base year values and benchmark values for the target years of the projections. Set of scenarios is built by combining different fertility levels and migration assumptions with the associated labor force and employment levels and these scenarios are compared according to child population, working age population, old age population, labor force and employment levels resulted by projections. Population projections are made by the cohort component method with single ages by using MS Excel VBA. Labor force and employment are calculated by applying assumed labor force participation and unemployment rates to the working age population. The time period covered in the projections is approximately 90 years starting from 2012. The following table shows the earlier results of the study on open years of demographic opportunity window, population increase rate and age structure of the population, where "Base" refers to the base scenario, "Base TFR2.5" assumes the high fertility increase with no migration scenario and "2018 25 2037 500 TFR2" is one of the reasonable representatives of the thousands of migration scenarios with a moderate fertility assumption. The table illustrates the impact of high fertility rate and migration with moderate fertility rate on the working age and old age population by the end of the century.

Table. Comparison of Scenarios by Population Parameters (2050 and 2100)

Scenarios	DOW*	2050			2100		
		PIR**	15-64%	65+%	PIR**	15-64%	65+%
Base	25	0.0%	63.1%	20.8%	0.4%	56.2%	28.8%
Base TFR2.5	27	0.6%	60.4%	18.5%	0.6%	58.8%	19.3%
2018 25 2037 500 TFR2	29	0.9%	63.5%	17.6%	0.5%	60.1%	22.0%

\*Open years of the demographic opportunity window.

\*\*Population increase rate.

The analyses concludes that the higher fertility policies alone may delay population decline but can protect against population aging to a very limited extent and after so many years. But with proper migration policies, it is possible to inject people directly to working age levels in the population pyramid while strengthening the base of the pyramid with fertility increasing policies. Additionally, the migration policies properly combined with fertility policies are much more effective than implementing solely fertility increasing policies as, higher fertility deteriorates women labor, by the means of sustaining working age population, labor force and employment at higher levels and slowing the population aging process in Turkey.

#### REFERENCES

- [1] Lutz, W. and Scherbov, S. (2003). Can immigration compensate for Europe's low fertility? Vienna: Vienna Institute of Demography (European Demographic Research Papers 1 (2003)).
- [2] UN. 2000. Replacement Migration: Is it a Solution to Declining and Aging Populations? New York: United Nations. Population Division, ESA/P/WP.160.

#### ABSTRACT

##### THE IMPACT OF REPLACEMENT MIGRATION POLICY ON DEMOGRAPHIC INDICATORS AND LABOR FORCE IN TURKEY

This paper investigates the mutual effects of migration and fertility assumptions in population projections of Turkey and offers relevant policies (combining fertility and migration policies) which sustain working age population, labor force and employment at higher levels and slow down the population aging process for Turkey. Set of scenarios is built by the population projection model which is made by cohort component method using MS Excel VBA. Labor force and employment levels are calculated by applying assumed labor force participation and unemployment rates to the working age population and the projection period of the model covers approximately 90 years.

**Key Words:** Replacement migration, demographic policy, fertility rate, aging, labor force, employment, population projection.

#### INTRINSIC FORECASTING IN REVERSE LOGISTICS: THE CASE OF ELECTRICAL AND ELECTRONIC EQUIPMENT

Nezir AYDIN<sup>a</sup>, Berk AYVAZ<sup>b</sup>, Erkan İŞIKLI<sup>c\*</sup>

Yıldız Technical University<sup>a</sup>, Istanbul Trade University<sup>b</sup>, Istanbul Technical University<sup>c</sup>  
neaydin@yildiz.edu.tr<sup>a</sup>, bayvaz@ticaret.edu.tr<sup>b</sup>, isiklie@itu.edu.tr<sup>c</sup>

#### Introduction

Due to increasing environmental consciousness, practices focusing on effective use and re-use of limited resources and currently available (already manufactured) products have drawn too much attention. Therefore, businesses nowadays are somewhat obliged to implement reverse logistics processes under some political, economic, and environmental restraints. Reverse logistics, which is a significant part of today's supply chain operations, primarily focuses on the collection and recovery of waste equipment or end-of-life products [1]. As its operational aspects (demand forecasting, inventory management, etc.) involve too much uncertainty, a reverse logistics network can be very complex. Deterministic models almost always fail to deal with the inherent uncertainty of such networks, thus stochastic programming techniques are considered. However, stochastic programming assumes that probability distributions of the random variables (parameters) in hand are known, which is often not the case. One way to tackle this problem is to work scenario-based. Another way is to estimate the unknown parameters with the help of statistical techniques. In this study, using monthly data provided by an İzmit-based third-party reverse logistics company, we intend to forecast the recycling and collection size of waste electrical and electronic equipment (e-waste) in Turkey to feed a two-stage stochastic program with capacity constraints. The size of waste collected in a year is one of the most critical parameters in reverse logistics network design as facility capacities, costs and many other factors are directly related to this parameter.

#### Models, Results, and Discussion

Intrinsic forecasting methods such as moving averages, seasonal trend lowess, Holt-Winters smoothing, AR, MA, and ARIMA were considered to achieve robust medium-term estimates of – logarithm of – e-waste size. Due to limited space, only the estimation results of four models that were found more satisfactory were summarized and illustrated. We should also note that we preferred using the natural logarithm of the e-waste size instead of the original time-series in the interest of avoiding heteroskedasticity and reducing the effect of a suspected outlier (observed in April 2009 probably due to large unanticipated shocks). Table 1 shows some criteria used to evaluate the performances of the models in hand. Model 1 is a simple linear regression model (LR) with quadratic trend, Model 2 is a multiple LR with quadratic trend and seasonal dummies, Model 3 is a Holt-Winters smoothing (HW) with no seasonality assumed, Model 4 is an HW with additive seasonality (periodicity assumed to be 12). Even though only the results of LRs with quadratic trend were summarized, we also checked for linear and exponential trend; however, these models were not found significant. Logistic trend and other S-shaped trends were not considered as they are most commonly used to model product life cycles where sales increase gradually and finally decrease after reaching the peak [2]. The time series does not tend to behave similarly in expected lapses of time, thus AR- or MA-type models are not necessary as their main purpose is to model cycles. Seasonality was detected, but the small sample size limited us on making strong inferences about it. There is a smooth (nonlinear) upward trend. Model 2 and 4 appear to be the best ones based on almost all statistics summarized in Table 1. However, Model 2 performed better in validation data (refer to the yellow shaded area in Figure 1 for out-of-sample forecasts).

Table 1. Model Diagnostics (36-month test data)

	R <sup>2</sup>	RMSE	AIC	BIC	Durbin-Watson
Model 1	0.5717	0.7349	82.85	87.60	1.665
Model 2	0.8154	0.5909	74.55	96.72	1.640
Model 3	N/A	0.8189	88.75	90.34	1.606
Model 4	N/A	0.5843	64.54	66.12	1.790

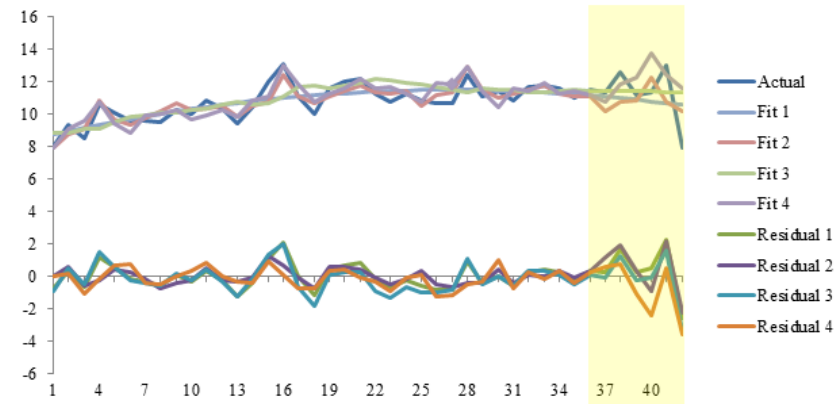


Figure 1. Model Fits and Residuals

## REFERENCES

- [1] Waters, D. (2007). Global Logistics: New Directions in Supply Chain Management, MPG Books Ltd, Great Britain.
- [2] Diebold, F.X. (2007). Elements of Forecasting, Thomson South-Western, USA.

## ABSTRACT

### INTRINSIC FORECASTING IN REVERSE LOGISTICS

Reverse logistics networks involve too much uncertainty on the amount of recycled products. In this study, we explore ways to model this uncertainty in the first step of constructing a stochastic optimization program. Four time-series models were built and tested using monthly data provided by a third-party reverse logistics company and they were compared and validated based on some goodness-of-fit criteria and residual plots.

**Key Words:** Reverse Logistics, E-Waste, Time Series Forecasting

## DETERMINING THE EFFECT OF SOCIO-ECONOMIC CHARACTERISTICS ON THE NET MIGRATION RATE TOWARDS THE PROVINCES IN TURKEY USING BAYESIAN MODEL AVERAGING

Hamza ERDOĞDU

Afyon Kocatepe Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümüans Kampüsü, Afyonkarahisar  
Email:hamzaerdogdu@aku.edu.tr

M.Ali CENGİZ

Ondokuz Mayıs Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü, Samsun  
Email:macengiz@omu.edu.tr

Süleyman DÜNDAR

Afyon Kocatepe Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü, Afyonkarahisar  
Email:dundar@aku.edu.tr;dundar69@hotmail.com

The purpose of this study is to analyze the socio-economic factors that drive national migration flows towards the provinces in Turkey. For this purpose we follow panel-based Bayesian Model Averaging (BMA) method which deals rigorously with model uncertainty especially if having a large number of potential regressors and relatively limited number of observations which our case has. The analysis suggests an alternative way to perform variable selection and provides a minimal list of variables when estimating the effect of other socio-economic characteristics on the net migration rate for a specific province in Turkey. In the context analyzed, the main variables which drive net migration rate with a high posterior probability are the rate of college graduates to the population 15 years old and over, unemployment rate and GSM subscriber per head. Health as well as other socio-economic factors seem to be of minor importance.

**Key words :** Bayesian Model Averaging, net migration rate, provinces of Turkey

### Sources :

- Bijak, J. (2005) 'Bayesian Methods in International Migration Forecasting', CEFMR Working Paper, 6/2005
- Bijak, J. (2006) 'Bayesian Model Averaging in Forecasting International Migration', European Population Conference 2006. Liverpool, United Kingdom, 21-24 June 2006
- Hoeting, J., Madigan, D., Raftery, A., & Volinsky, C. (1999), 'Bayesian model averaging: A tutorial', Statistical science 14(4), 382-401
- Raftery, A. E. (1995), 'Bayesian model selection in social research', Sociological Methodology 25, 111-163

## ROBUST QUADRATIC HEDGING PROBLEM: IN INCOMPLETE MARKETS

Gültaç EROĞLU İNAN<sup>a\*</sup>, Ayşen APAYDIN<sup>b</sup>

<sup>a</sup>Ankara University, Faculty of Science, Department of Statistics, 06100-Tandoğan, Ankara, Türkiye.  
geroglu@science.ankara.edu.tr

<sup>b</sup>Ankara University, Faculty of Science, Department of Statistics, 06100-Tandoğan, Ankara, Türkiye.  
aapaydin@ankara.edu.tr

### 1. Introduction

An option contract is an agreement granting the purchaser (owner) of the option the right to buy or sell a underlying asset at a predetermined price at a predetermined date in the future. There are two types of options: calls and puts. A call is the right to purchase the underlying asset. A put is the right to sell that asset. There must be a buyer and a seller for each type of option. A put is not the opposite side of a call. Puts and calls are different contracts. The item described in the option is said to underlying asset. The price at which the item can be bought or sold under the option is called the strike price or exercise price. The period of time during which an option can be exercised is the life of the option. At the end of the its life (on or following the last day of trading) the option become worthless. Every option is defined by these three specifications: underlying asset, the exercise price and the life or the expiration date.

### 2. Robust Quadratic Hedging Problem in Incomplete Markets

An option, allows the buyer to generate unlimited profit with limited risk. The risk or loss of the buyer can be unlimited. This loss is defined as the payoff option and it is required to be provided at the end of the period. In an incomplete market, it is impossible to eliminate the risk of an option. In incomplete markets, quadratic risk minimization is often used to determine the optimal hedging strategy. In this study, at first, in an incomplete market, for the optimal solution of the quadratic hedging problem, at one period situation, Föllmer-Schweizer (1989) linear regression approach, was introduced.

While optimization algorithms and software today allow a user to handle a wide variety of very complex optimization problems, the optimal solutions produced by optimization solvers can be very sensitive to small fluctuations in the problem inputs. Since real world data are rarely certain or accurate, a number of optimization methods have been suggested for treating parameter uncertainty. Robust optimization methods have emerged as a computationally attractive alternative to stochastic and dynamic programming methods. Robust optimization requires problems to remain feasible for any values of the uncertain parameters within pre-specified uncertainty sets [1]. Robust optimization method, used to overcome the uncertainty, is used for the hedging problems. In this study, Pinar (2006) robust quadratic hedging problem was discussed.

### KAYNAKLAR

[1] Fabozzi, F.J., Kolm, P.N., Pachamanova, D.A. and Focardi, S.M. 2007. Robust Portfolio Optimization And Management. John Wiley.

[2] Föllmer, B.Y. and Schweizer, M. 1989. Hedging By Sequential Regression : An Introduction To The Mathematics of Option Trading , ASTIN Bulletin, Vol .18, No. 2, pp. 147-160.

[3] Pinar, M.C. 2006. On Robust Quadratic Hedging of Contingent Claims in Incomplete Markets under Ambiguous Uncertainty. Presented at the First Conference on Advanced Mathematical Methods in Finance.

### ÖZET

#### DAYANIKLI KARESEL RİSKTEN KORUNMA PROBLEMİ : NATAMAM PİYASALARDA

Opsiyon Sözleşmesi, opsiyon sahibine bir finansal varlığı, önceden belirlenen gelecekte bir zamanda, önceden belirlenen bir fiyattan alım veya satma hakkı tanıyan bir anlaşmadır. Opsiyon sözleşmesi satın almanın hak sahibine sağladığı en büyük fayda, sınırlı bir riskle, sınırsız kar sağlama olanağı elde etmektir. Sözleşmeyi satan tarafın riski veya kaybı ise sınırsız olabilmektedir. Bu kayıp, opsiyonun bedeli olarak tanımlanmaktadır ve dönem sonunda sağlanmak istenmektedir. Bir natamam piyasada, opsiyonun riskini ortadan kaldırmak mümkün olmamaktadır. Bir natamam piyasada, en iyi riskten korunma stratejisini oluşturmak için genellikle karesel risk minimizasyonu kullanılmaktadır. Bu çalışmada, öncelikle natamam piyasada karesel riskten korunma probleminin optimal çözümü için, tek periyotlu durumda Föllmer-Schweizer (1989) doğrusal regresyon yaklaşımı tanıtıldı. Belirsizlik ile mücadele için kullanılan dayanıklı optimizasyon yöntemi, riskten korunma

problemlerinde de kullanılabilmektedir. Bu çalışmada, Pinar (2006) dayanıklı karesel riskten korunma problemini ele almıştır.

**Anahtar Kelimeler:** Karesel riskten korunma problemi, belirsizlik, dayanıklı optimizasyon, dayanıklı karesel riskten korunma problemi, natamam piyasa.

## A SURVEY ON NONRANDOM PATTERNS OF FUZZY CONTROL CHARTS

Nilüfer PEKİN ALAKOÇ<sup>a\*</sup>, Ayşen APAYDIN<sup>b</sup>

<sup>a</sup>Atılım University, Faculty of Engineering, Department of Industrial Engineering, 06836-İncek, Ankara, Türkiye.  
[npalakoc@atilim.edu.tr](mailto:npalakoc@atilim.edu.tr)

<sup>b</sup>Ankara University, Faculty of Science, Department of Statistics, 06100-Tandoğan, Ankara, Türkiye.  
[apaydin@science.ankara.edu.tr](mailto:apaydin@science.ankara.edu.tr)

### Fuzzy theory and fuzzy control charts

Fuzzy set theory is used to represent the uncertainty. Since it is introduced in 1965 [1], fuzzy set theory is spread to a wide variety of research areas from various point of views. Over the years many applications and implementations are performed in many fields such as medical, social and natural sciences and engineering.

Quality control charts is one of the most important and commonly used tool of statistical process control. The major objective of using a control chart is to eliminate the unexpected sources of variability of the quality characteristic. A control chart performs this goal by monitoring the process over the time. If all the points of the chart are plotted within the limits with a random pattern, then the process is assumed to be "in control" otherwise the process is "out of control". The situations that are defined for the nonrandom patterns of the control charts are sometimes called the zone rules or the sensitizing rules [2]. These rules are widely used with quality control charts for all types of processes.

### The objective

Uncertainty exists in almost all real world systems including the statistical process control problems. Therefore, fuzzy rules are suggested for fuzzy control charts. In this paper, a review of the studies on nonrandom patterns of fuzzy control charts is presented. Since the first fuzzy control chart is introduced, various approaches are proposed for monitoring fuzzy control charts. Almost all these fuzzy control charts necessitate to be controlled for nonrandom patterns. The studies that describes the types of out of control conditions are summarized with the detecting methods and the contributions to the literature are discussed.

### REFERENCES

- [1] Zadeh L.A. (1965), Fuzzy sets, Information and Control, 8, 338-353.
- [2] Montgomery D.C. (1996), Introduction to statistical quality control, John Wiley & Sons, Inc., USA.
- [3] Gwee B.H., Lim M.H. and Soong B.H. (1993), Self-Adjusting Diagnostic System for the Manufacture of Crystal Resonators, Proceedings of IEEE Industry Application Society Annual Meeting, IAS'93, Toronto, Canada, 3, 2014-2020.
- [4] Gülbay M. and Kahraman C. (2006), Development of fuzzy process control charts and fuzzy unnatural pattern analyses, Computational Statistics and Data Analysis, 51, 434-451.
- [5] Fazel Zarandi M.H., Alaedini A. and Türksen I.B. (2008), A hybrid fuzzy adaptive sampling - Run rules for Shewhart control charts, Information Sciences, 178, 4, 1152-1170.

### ÖZET

#### BULANIK KALİTE KONTROL GRAFİKLERİNİN RASGELE OLMAYAN ÖRÜNTÜLERİ ÜZERİNE BİR İNCELEME

Günümüzün karmaşık sistemleri kesin olmayan kararlar, belirsizlikler içerir. Bu nedenle bulanık kalite kontrol grafikleri geliştirilmiştir. Kalite kontrol grafikleri endüstride kabul edilmiş ve en yaygın olarak kullanılan istatistiksel süreç kontrol araçlarından biridir. Bir sürecin kontrol altında olarak tanımlanabilmesi için kontrol grafiğindeki noktaların kontrol sınırlarının içinde ve rasgele olarak dağılması gerekir. Rasgele olmayan örüntüleri tanımlayan kurallar kalite kontrol grafiklerinde sıklıkla kullanılır. Bulanık teori ile geliştirilmiş kontrol grafiklerinde de bu durumların belirlenmesine ihtiyaç duyulur. Bu çalışmada, bulanık kontrol grafiklerinde

rasgele olmayan durumları tanımlayan araştırmalar üzerinde durulmuştur. İncelenen çalışmaların literatüre katkıları hakkında kısa bilgiler verilmiştir.

**Anahtar Kelimeler:** Bulanık kalite kontrol grafikleri, Bulanık küme teorisi, Bulanık sayılar.

A FUZZY TIME SERIES APPROACH BASED ON DIFFERENTIAL EVOLUTION ALGORITHM  
USING DE/BEST/1 MUTATION STRATEGY

Eren BAS<sup>1,\*</sup>, Vedide Rezan USLU<sup>2</sup>, Ufuk YOLCU<sup>3</sup>, Erol EGRIOGLU<sup>2</sup>

<sup>1,\*</sup> Giresun University, Faculty of Arts and Science, Department of Statistics, Giresun 28100, Turkey, e-mail:eren.bas@giresun.edu.tr

<sup>2</sup> Ondokuz Mayıs University, Faculty of Arts and Science, Department of Statistics, Samsun 55139, Turkey, e-mail:rezzanu@omu.edu.tr, erole@omu.edu.tr

<sup>3</sup> Ankara University, Science Faculty, Department of Statistics, 06100 Ankara, e-mail:varyansx@hotmail.com

**Introduction**

The fuzzy set theory was firstly introduced by Zadeh [1] and this concept has found many application areas since then. For the first time, Song and Chissom [2] introduced fuzzy time series. Fuzzy time series approaches consist of three steps. These are called as the fuzzification, the determination of fuzzy relationships and the defuzzification steps. All these steps of fuzzy time series are very important on the forecasting performance of the model. There are many studies that contribute for each step in the literature. Song and Chissom [2] and Chen [3] determined fix interval lengths, arbitrarily whereas Huang [4] used methods based on average and distribution. In addition, Cheng et al. [5] used fuzzy C-means (FCM) clustering techniques. Artificial intelligence optimization algorithms have been used frequently in almost all areas in recent years and also they have been used in different stages of fuzzy time series approaches. Genetic algorithm, particle swarm optimization and differential evaluation algorithm (DEA) are the most popular algorithms among these artificial intelligence optimization algorithms. In this study, we proposed an approach to determine the sub-intervals by using differential evaluation algorithm in fuzzification step and also we used the DE/Best/1 Mutation Strategy in the stage of mutation which is the important stage of DEA. It is aim to obtain more consistent results and so superior forecasting performance by using DE/Best/1 mutation strategy in the mutation stage of DEA in fuzzification stage. By using DE/best/1 mutation strategy, it is aim that the system is rescued from randomness and the best chromosome is constantly kept in the system by updating it, at the each iteration. Thus, the solutions shaped around the best chromosome and the forecasting performance has been improved by avoiding randomness. To evaluate the performance of the proposed method, the proposed method was applied to the time series data of student enrollments at University of Alabama between the years 1971 and 1992 which is frequently used in the literature.

**Application and Conclusion**

In the application, the parameters of DEA (the number of chromosomes (cn), crossover rate (cor) and the number of intervals (m)) were determined as follows and 1600 different solutions with the following properties were obtained.

- The number of chromosomes (cn) was tested between 10 and 100 with increment 10 as totally 10 alternatives.
  - Crossover rate (cor) was tested between 0.1 and 1 with increment 0.1 as totally 10 alternatives.
  - The number of intervals (m) was tested between 5 and 20 with increment 1 as totally 16 alternatives.
- Then the parameters (m, cn and cor) with the smallest the root of the mean squared error (RMSE) value were taken as the best solution among these solutions. At the end of the process, we conclude that the best result was obtained in the case where cn=80, cor=1, m=14. Table 1 presents RMSE values obtained from the proposed method and some other methods proposed in literature, comparatively.

**BİLDİRİ OTURUMLARI 3**

*SESSION 3*

*Time Series*



Table 1. A comparative presentation of enrollments' RMSE values by various methods

Methods	RMSE
Song and Chissom [2]	642
Chen [3]	638
Huang [4]	353
Cheng et al. [5]	478
<b>The Proposed Method</b>	<b>270</b>

When the Table 1 was inspected, it is obviously seen that the proposed method provides the superior forecasting performance.

#### REFERENCES

- [1] Zadeh L.A. (1965), Fuzzy Sets: Inform and Control, 8, pp. 338-353
- [2] Song Q. and Chissom B.S. (1993a), Fuzzy time series and its models: Fuzzy Sets and Systems, 54, pp. 269-277.
- [3] Chen S. M. (1996), Forecasting enrollments based on fuzzy time-series, Fuzzy Sets and Systems, 81, pp. 311-319.
- [4] Huang K. (2001), Effective length of intervals to improve forecasting in fuzzy time-series, Fuzzy Sets and Systems, 123, pp. 387-394.
- [5] Cheng C.H., Cheng G.W. and Wang J.W. (2008), Multi-attribute fuzzy time series method based on fuzzy clustering, Expert Systems with Applications, 34, pp. 1235-1242.

#### ÖZET

#### DE/BEST/1 MUTASYON STRAGESİNİ KULLANAN DİFERANSİYEL GELİŞİM ALGORİTMASINA DAYALI BİR BULANIK ZAMAN SERİSİ YAKLAŞIMI

Bulanık zaman serisi yaklaşımları, klasik zaman serilerinin gözlemleri belirsizlik içerdiği durumlarda sıklıkla kullanılmaktadır. Bulanık zaman serileri genellikle üç aşamadan oluşur. Bu aşamalar sırasıyla, bulanıklaştırma, bulanık ilişkilerin belirlenmesi ve durulaştırma aşamalarıdır. Bu çalışmada, bulanıklaştırma aşamasında alt aralıkların diferansiyel gelişim algoritması ile belirlendiği ve diferansiyel gelişim algoritmasının mutasyon aşamasında DE/Best/1 mutasyon stratejisinin kullanıldığı bir bulanık zaman serisi yaklaşımı önerilmiştir. Önerilen yaklaşım bir gerçek hayat zaman serisine uygulanmış ve elde edilen sonuçlar literatürdeki diğer yöntemlerle karşılaştırılarak, önerilen yöntemin üstün öngörü performansı ortaya konulmuştur.

**Anahtar Kelimeler:** Öngörü, diferansiyel gelişim algoritması, bulanık zaman serileri.

#### A NEW ADAPTIVE NETWORK FUZZY INFERENCE SYSTEM FOR TIME SERIES FORECASTING

Erol EGRIOGLU<sup>1</sup>, Çağdas Hakan ALADAG<sup>2</sup>, Ufuk YOLCU<sup>3</sup>

<sup>1</sup> Ondokuz Mayıs University, Department of Statistics, e-mail:erole@omu.edu.tr

<sup>2</sup> Hacettepe University, Department of Statistics, e-mail:aladag@hacettepe.edu.tr

<sup>3</sup> Ankara University, Department of Statistics, e-mail:varyansx@hotmail.com

#### Introduction

Recent years, forecasting methods which are based on fuzzy set theory have been frequently used in the literature. Adaptive network fuzzy inference systems (ANFIS), fuzzy regression and fuzzy time series approaches were used as fuzzy forecasting techniques. Fuzzy time series approaches were originally proposed for forecasting. ANFIS and fuzzy regression methods were proposed for prediction or regression problems. Time series problem is one of the regression problems. All regression methods can be used for forecasting. Redesigning of the regression methods for forecasting problems can be useful. ANFIS was used for forecasting problems, but it was not originally proposed for forecasting problems. ANFIS was firstly proposed in Jang [1]. It is a rule based fuzzy inference system. There are two critical processes in the ANFIS. These are determination of the rules and estimating the parameters of input and output membership functions. Determination of the rules can be achieved by using subtractive clustering method which was proposed in Chiu [2]. The back propagation learning algorithm is applied to estimate parameters of input and output membership functions. In the literature, Wei [3], Chang et al. [4] and Li et al. [5] applied ANFIS method to obtain forecasts of the time series. In this study, ANFIS is redesigned for forecasting problem and as a result of this, new ANFIS is proposed. In the proposed method, fuzzy c-means and particle swarm optimization techniques are employed. The membership values are obtained from fuzzy c-means and the rule base are constituted from time series by using fuzzy c-means. In the proposed method, it is no need to use membership function for inputs. The parameters of output memberships are estimated by using particle swarm optimization technique. In the next section, some application results of the proposed method are summarized.

#### Summarized Results of the Application

In the application, proposed method was applied three time series. The first time series is Australian beer consumption data. The time series are quarterly observed in 1956 and 1994. Time series have 148 observations. The first 132 observations and last 16 observations were used as training and test set, respectively. The forecasts of time series were obtained by using Winters' multiplicative exponential smoothing method (WMES), Seasonal autoregressive integrated moving average model (SARIMA), Radial basis neural network, feed forward neural network and proposed method. The obtained results are summarized in Table 1. In table 1, root of mean square error (RMSE) and mean absolute percentage error (MAPE) performance criteria are given. It is clearly shown that the proposed method has the best forecasting performance for the test set. Moreover, the proposed method was applied two stock exchange data. The first stock exchange data is Index 100 for the stocks and bonds exchange market of Istanbul which is daily observed in periods 3 October 2008 to 31 December 2008. The second stock exchange data is Index 100 for the stocks and bonds exchange market of Istanbul which is daily observed in periods 1 October 2010 to 23 December 2010. The results of the application for these two data sets will be given in the full paper.

Table 1. The Application Results for Australian Beer Consumption Data Set.

	WMES	SARIMA	RBF	FFANN	ANFIS	Proposed Method
RMSE	53,3295	47,0367	41,7000	24,1052	25,0500	21,3728
MAPE	0,1072	0,0949	0,0686	0,0476	0,0467	0,0400

## REFERENCES

- [1] Jang J.S., (1993) ANFIS: Adaptive-Network-based Fuzzy Inference Systems, IEEE Transactions on Systems Man, and Cybernetics, 23(3), 665–685.
- [2] Chiu S.L., (1994) Fuzzy model identification based on cluster estimation, Journal of Intelligent and Fuzzy Systems, 2, 267–278.
- [3] Wei L.Y., (2013), A GA-weighted ANFIS model based on multiple stock market volatility causality for TAIEX forecasting, Applied Soft Computing, 13, 911–920.
- [4] Chang J.R., Wei L.Y., Cheng C.H., (2011), A hybrid ANFIS model based on AR and volatility for TAIEX forecasting, Applied Soft Computing, 11, 1388–1395.
- [5] Li K., Su H., Chu J., (2011), Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: A comparative study, Energy and Buildings, 43, 2893–2899.

## ÖZET

### ZAMAN SERİSİ ÖNGÖRÜSÜ İÇİN YENİ BİR UYARLAMALI AĞ BULANIK ÇIKARIM SİSTEMİ

Bulanık çıkarım sistemleri birçok gerçek hayat probleminin çözümü için kullanılmaktadır. Bulanık çıkarım sistemlerinden önemli bir tanesi yapay sinir ağlarının öğrenme algoritması olan geri yayılım öğrenme algoritmasını kullanan ve Sugeno'nun çıkarım sisteminin ağ yapısında verildiği uyarlamalı ağ bulanık çıkarım sistemidir (ANFIS). ANFIS orijinal olarak tahmin ve regresyon problemleri için önerilmiştir. Literatürde ANFIS zaman serisi tahmini içinde kullanılmasına rağmen, bu çalışmalarda zaman serileri, diğer zaman serilerinin eş anlamlı değişkenleri ile açıklanmıştır. Oysa zaman serilerinin öngörüsünde, zaman serisinin kendi gecikmeli değişkenlerinin kullanılmasının başarılı öngörü sonuçlarına neden olduğu bilinmektedir. Diğer zaman serileri yerine, zaman serisinin gecikmeli değişkenleri ile çözümleme yapıldığı durumda ANFIS'in yeniden gözden geçirilmesi gerekmektedir. Bu çalışmada ANFIS öngörü problemi için yeniden tasarlanarak, zaman serisi öngörü problemi parçacık sürü optimizasyonu ve bulanık C ortalamaların kullanıldığı yeni bir ANFIS yöntemi önerilmiştir.

**Anahtar Kelimeler:** Bulanık Çıkarım Sistemleri, Öngörü, Bulanık C Ortalamalar Kümeleme yöntemi, Parçacık Sürü Optimizasyonu.

## A HYBRID FUZZY TIME SERIES APPROACH FOR FORECASTING EXCHANGE MARKET OF ISTANBUL

Ozge CAGCAG YOLCU, Faruk ALPASLAN

Ondokuz Mayıs University, Faculty of Arts and Sciences, Department of Statistics, 55139, Samsun, TURKEY,  
E-mail: ozgecagcag@yahoo.com (O.CAGCAG YOLCU), falpas@omu.edu.tr (F. ALPASLAN)

## Introduction

Fuzzy time series forecasting models which were first introduced by Song and Chissom [1] do not require assumptions that stochastic models do. Fuzzy time series forecasting models consist of three steps as fuzzification, the identification of fuzzy relation and defuzzification which play an active role on forecasting performance. In the fuzzification, Song and Chissom [1] and Chen [2] determined fixedly interval lengths arbitrarily whereas Huarng [3] used average and distribution based and also some other researchers used optimization based methods. Approaches not requiring partition of universe of discourse in fuzzification step and using various fuzzy clustering techniques were proposed by some researchers in the literature.

Different approaches were also proposed for the identification of fuzzy relation is the step in which the appropriate model is determined. Aladag et al. [4] and some other researchers proposed approaches using feed-forward artificial neural networks (ANN) in the identification of fuzzy relations. In all of these approaches, only the fuzzy set having the highest membership value is considered. This leads to loss of information and negatively affects the performance of the method. In order to overcome this problem, Yolcu et al. [5] used forecasting models which consider membership values and used fuzzy C-means clustering technique instead of determining the membership values subjectively. The use of ANN in identification of fuzzy relations has many advantages and disadvantages as well. Determination of unit number in hidden layer (architecture structure) and excessive number of parameters to be used during the analysis are the most prominent ones. Nevertheless, as the system output of these approaches consists of fuzzy set number or membership values, fuzzification step is necessary. This may be a factor that increases the model error. An approach not requiring defuzzification step would eliminate forecasting error that may occur in this step and improve the performance of the method. The hybrid approach which is introduced uses Gustafson-Kessel fuzzy clustering technique in fuzzification step and membership values are obtained more systematically. The use of artificial neural network with single multiplicative neuron model (SMNM-ANN) in identification of fuzzy relations eliminates architecture selection problem and the need for defuzzification step by constituting the target values from observations of the real time series. The training of SMNM-ANN is carried out with particle swarm optimization.

## Application and Conclusion

The hybrid method was applied to different time series namely; exchange market of Istanbul (IEX) in 2011 and 2012 years. In the analysis of IEX, we used observations of last two and three month as the out-of-sample observations (test data). Prediction errors for the optimal results obtained from the hybrid method as well as prediction error of other fuzzy time series methods are presented in Table 1.

Table 1. Performance evaluation of methods.

Methods	2011		2012		2012		2012	
	Test Data 1	Test Data 2	Test Data 1	Test Data 2	Test Data 1	Test Data 2	Test Data 1	Test Data 2
Song and Chissom [1]	925.47	1.35	1022.58	1.48	3305.17	3.47	6451.82	7.60
Chen [2]	925.47	1.35	969.74	1.44	683.14	0.80	694.07	0.80
Huarng [3] <sup>1</sup>	1114.43	1.82	1118.63	1.71	669.32	0.78	631.88	0.74
Huarng [3] <sup>2</sup>	985.07	1.53	1053.30	1.50	693.73	0.75	601.59	0.71
Aladag et al. [4]	929.77	1.31	977.45	1.38	815.56	0.89	963.23	1.07
Yolcu et al. [5]	898.62	1.27	960.05	1.37	921.46	0.98	1621.42	1.75

The Hybrid Method 812.02 1.12 886.01 1.22 542.84 0.61 536.93 0.57

**RMSE:** Root mean square error **MAPE:** Mean absolute percentage error <sup>1</sup> Average based <sup>2</sup> Distribution based

When the Table 1 is analyzed, it can be clearly say that forecasting performance of the hybrid method for this data is better than those exist in the literature.

#### REFERENCES

- [1] Song Q. and Chissom B.S. (1993), Fuzzy time series and its models: Fuzzy Sets and Systems, 54, pp. 269-277.
- [2] Chen S. M. (1996), Forecasting enrollments based on fuzzy time-series, Fuzzy Sets and Systems, 81, pp. 311-319.
- [3] Huang K. (2001), Effective length of intervals to improve forecasting in fuzzy time-series, Fuzzy Sets and Systems, 123, pp. 387-394.
- [4] Aladag C.H., Basaran M.A., Egrioglu E., Yolcu U. and Uslu V.R. (2009), Forecasting in high order fuzzy time series by using neural networks to define fuzzy relations, Expert Systems with Applications, 36, pp. 4228-4231.
- [5] Yolcu U., Aladag, C.H., Egrioglu E. and Uslu V.R. (2013), Time series forecasting with a novel fuzzy time series approach: an example for İstanbul stock market, Journal of Computational and Statistics Simulation, 83(4), pp. 597-610.

#### ÖZET

#### İSTANBUL BORSA VERİSİ ÖNGÖRÜSÜ İÇİN MELEZ BİR BULANIK ZAMAN SERİSİ YAKLAŞIMI

Literatürde olasılıksal ve olasılıksal olmayan başlıkları altında değerlendirebileceğimiz birçok zaman serisi öngörü yöntemi mevcuttur. Olasılıksal yöntemlerin içerdiği varsayımları içermeyen bulanık zaman serisi yöntemleri bulanıklaştırma, bulanık ilişkilerin belirlenmesi ve berraklaştırma aşamalarından oluşur. Literatürde bu üç aşama üzerine de farklı yaklaşımlar ortaya konmuştur. Bunun yanında halen, bulanık ilişkilerin belirlenmesinde yapay sinir ağlarının kullanımında mimari seçim problemi, üyeliklerin dikkate alınmaması ya da üyeliklerin uygun belirlenmemesi gibi sorunlar mevcuttur. Bu çalışmada, İstanbul borsa verisini öngörmek amacıyla melez bir bulanık zaman serisi yöntemi tanıtılmıştır. Bu yöntemde belirtilen tüm problemlerin yanı sıra, berraklaştırma aşaması bulanık ilişki belirleme aşaması ile birleştirilerek, bu aşamada oluşabilecek tahmin hatası ortadan kaldırılmıştır.

**Anahtar Kelimeler:** Öngörü; Bulanık Zaman Serileri; Bulanık Kümeleme; Üyelik Değeri

#### AN IMPROVED HIGH ORDER FUZZY TIME SERIES APPROACH

Ufuk YOLCU<sup>1\*</sup>, Ozge CAGCAG YOLCU<sup>2</sup>, Erol EGRIOGLU<sup>2</sup>, Cagdas Hakan ALADAG<sup>3</sup>  
<sup>1\*</sup>Ankara University, Faculty of Sciences, Department of Statistics, 06100, Ankara, TURKEY,  
E-mail: varyansx@hotmail.com

<sup>2</sup>Ondokuz Mayıs University, Faculty of Arts and Sciences, Department of Statistics, 55139, Samsun, TURKEY,

E-mail: ozgecagcag@yahoo.com; erole@omu.edu.tr

<sup>3</sup>Hacettepe University, Faculty of Sciences, Department of Statistics, 06800, Ankara, TURKEY,  
E-mail: chaladag@gmail.com

#### Introduction

Fuzzy time series can be evaluated under two main headings as time-variant and time-invariant. Song and Chissom [1] proposed an algorithm for the analysis of time-invariant time series which is almost the subject of all studies in literature. Fuzzy time series forecasting models consist of three steps as fuzzification, identification of fuzzy relation and defuzzification and each has positive and negative impact on the forecasting performance. In various studies, in the partition of universe of discourse, some researcher determined equal interval lengths arbitrarily, whereas others used average and distribution based and optimization based methods. Separately to determine the dynamic length of interval, Hsu et al. [2] utilized particle swarm optimization while Lee et al. [3], [4] used genetic algorithm. In addition, approaches using fuzzy clustering techniques in fuzzification step were proposed. As the identification of fuzzy relation is the step in which the appropriate model is determined, it can be considered to be the step playing the most active role on forecasting performance. Especially in recent years, feed forward artificial neural networks (FFANN) utilized in this step. In all of these approaches, memberships are ignored in the determination of fuzzy relations that represent the internal relation of fuzzy time series. In order to overcome this problem, Yolcu et al. [5] and some other researchers proposed various approaches which consider all membership values in the determination of relations with FFANN. Although these methods use approaches that consider membership values, they involve only first order fuzzy time series forecasting models. It is obvious that first order fuzzy time series forecasting models would be insufficient to forecast some time series. Having increased number of inputs of ANN which are used in determination of fuzzified relations is the main problem that may arise during transformation of involved only first order fuzzy time series forecasting models into a high-order approach. In such a case, training of the network would become difficult and take longer time and forecasting performance would be influenced negatively.

In this study, we introduced a high-order fuzzy time series forecasting model (HO-FTS) which considers membership values in identification of fuzzy relations and defuzzification steps. The introduced high-order method used fuzzy C-means in fuzzification step and ANN with multiple-input and multiple-output in identification of fuzzy relations. Furthermore, the problem related to the number of input of ANN that is likely to arise in a high-order model was eliminated by implementing operation of intersection to the membership values.

#### Application and Conclusion

The introduced high-order fuzzy time series forecasting model was applied to two different time series namely; exchange market of Istanbul (IEX) and Taiwan Futures Exchange (TAIFEX). The errors of prediction obtained from all methods are compared with each other and the all obtained results are summarized in Table 1.

Table 1. The obtained prediction errors from the methods.

	IEX Data			TAIFEX Data			
	[1]	[5]	HO-FTS	[2]	[3]	[4]	HO-FTS
<b>RMSE</b>	1161.21	929.82	611.46	93.49	102.96	80.02	67.14
<b>MAPE</b>	1.39%	1.13%	0.90%	1.09%	1.14%	0.87%	0.66%
<b>MdAPE</b>	1.30%	0.76%	0.99%	0.77%	0.78%	0.61%	0.48%
<b>DA</b>	50.00%	50.00%	83.33%	53.33%	80.00%	73.33%	80.00%

**RMSE:** Root mean square error

**MAPE:** Mean absolute percentage error

**MdAPE:** Median absolute percentage error

**DA:** Direction accuracy

#### KAYNAKLAR

[1] Song Q. and Chissom B.S. (1993), Forecasting enrollments with fuzzy time series - Part I, Fuzzy Sets and Systems, 54, pp.1-10.

[2] Lee L.W., Wang L.H. and Chen S.M., (2007), Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms, Expert Systems with Applications, 33, pp. 539-550.

[3] Lee L.W., Wang L.H. and Chen S.M., (2008), Temperature prediction and TAIFEX forecasting based on high-order fuzzy logical relationships and genetic simulated annealing techniques, Expert Systems with Applications, 34, pp. 328-336.

[4] Hsu L.Y., Horng S.J., Kao T.W., Chen Y.H., Run R.S., Chen R.J., Lai J.L. and Kuo I.H., (2010), Temperature prediction and TAIFEX forecasting based on fuzzy relationships and MTPSO techniques, Expert Systems with Applications, 37, pp. 2756-2770.

[5] Yolcu U., Aladag, C.H., Egrioglu E. and Uslu V.R. (2013), Time series forecasting with a novel fuzzy time series approach: an example for İstanbul stock market, Journal of Computational and Statistics Simulation, 83(4), pp. 597-610

#### ÖZET

##### GELİŞTİRİLMİŞ YÜKSEK DERECELİ BİR BULANIK ZAMAN SERİSİ YAKLAŞIMI

Bulanık zaman serisi yöntemleri, bulanıklaştırma, bulanık ilişkilerin belirlenmesi ve berraklaştırma aşamalarından oluşur. Literatürdeki çoğu çalışmada bulanık ilişkiler belirlenirken üyelik değerleri ya ihmal edilmekte ya da sübjektif yargılarla belirlenmektedir. Bu sorunları içermeyen bazı yaklaşımlar ortaya konmasına rağmen bu yaklaşımlar yalnızca birinci dereceden öngörü modellerini içermektedir. Birinci dereceden modeller ise birçok zaman serisi öngörü probleminde yetersiz kalabilmektedir. Üyelik değerlerini dikkate alan ve bulanık ilişki belirlemede yapay sinir ağları (YSA) kullanan bu tip modelleri yüksek dereceli bir modele dönüştürmek ise aşırı girdi ve parametre sayısı gibi problemlere yol açacaktır. Bu çalışmada, bu sorunları barındırmayan yüksek dereceli bir bulanık zaman serisi yöntemi tanıtılmıştır. Yüksek dereceli bir modelde oluşabilecek aşırı girdi sayısı problemi ise gecikmeli zaman serilerinin her bir gözlemine ait üyelik değerlerine uygulanan "kesişim" işlemi ile giderilmiştir. Yöntemin performansı iki farklı uygulama ile değerlendirilmiştir.

**Anahtar Kelimeler:** Bulanık Zaman Serileri; Öngörü; Üyelik Değeri; Kesişim İşlemi

#### DIFFERENT PERFORMANCE MEASURES FOR FORECASTING TOURISM DEMAND OF TURKEY

Bulent ALPTEKİN, Mert Can ALAYBEYOĞLU, Cagatay BAL, Cagdas Hakan ALADAG

Hacettepe University, Faculty of Science, Department of Statistics, Ankara, Turkey

E-mails: alptekinblnt@gmail.com, alaybeyoglucan@gmail.com, nemesis06bal@gmail.com, aladag@hacettepe.edu.tr

#### Abstract

There is a need to forecast tourism demand accurately so that directors and investors can make operational, tactical and strategic decisions, examples of which are scheduling and staffing, preparing tour brochures and hotel investments, respectively needs [1]. Similarly, government organizations need accurate tourism demand forecasts to plan required tourism infrastructures, such as accommodation site planning and transportation development, among other needs [5].

International tourist receipts can be a good indicator of the role of tourism in an economy in term of both Gross Domestic Product and foreign exchange generation. Policy makers may subsequently A high international tourist arrivals level may be used in advertising campaigns and also in political discussions to legitimize and emphasize the success of a country in the international community. Similarly, sizeable be convinced to assist tourism development and further increase profitability from tourism activities [3].

Since forecasting tourism demand accurately is a vital issue, many forecasting techniques have been proposed to analyze this kind of time series. Artificial neural networks (ANN) are one of these methods [1]. Although ANN has proved its ability in various time series applications, there are still some problems with its usage. One of these problems is to determine the best architecture which produces the best forecasts [2]. Lots of performance measures such as root mean square error (RMSE), mean absolute percentage error (MAPE), median absolute percentage error (MdAPE), Bayesian information criterion, Akaike information criterion, direction accuracy, and modified direction accuracy can be used for determining the best architecture in ANN.

Every performance measure evaluates forecasting error from different aspects. For this reason Egrioglu et al. [4] proposed weighted information criterion (WIC) which is composed of different performance measures so that it can evaluate the error from different aspects. The numbers of international tourist arrivals to Antalya and İstanbul were employed. These monthly time series include observations between January 2004 and May 2013. Also, the total number of foreign tourist arrivals to Turkey was forecasted in the implementation. This monthly time series has observations from January 2004 to April 2013.

These time series were analyzed with feed forward neural network models. The number of units of input and hidden layers change from 1 to 12 and 144 architectures were examined for each time series. Thus, 432 architectures were totally examined for three time series. In addition to WIC, RMSE, MAPE and MdAPE performance measures were also exploited to determine the best feed forward neural network architecture. These criteria are calculated over the test set. The best architectures obtained according to the performance measures for each time series are given in Table 1. In this table, for example, 12-3-1 represents the architecture which has 12 and 3 neurons in input and hidden layers, respectively. In this study, all obtained results were compared to each other and discussed. Also, all obtained results were examined graphically and presented.

Performance Measures	ANTALYA	ISTANBUL	TURKEY
WIC	12-1-1	12-1-1	12-1-1
RMSE	12-1-1	12-3-1	12-1-1
MAPE	1-11-1	7-10-1	2-10-1
MdAPE	1-11-1	7-10-1	2-10-1

Table 1. The obtained best architectures for each time series

#### KAYNAKLAR

- [1] Aladag C. H. and Egrioglu E.(2010),Forecasting tourism demand of Turkey by using artificial neural networks,The Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Proceedings of the ICCS-X, The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, 1, 384–391.
- [2] Aladag C.H. (2011), A new architecture selection method based on tabu search for artificial neural networks, Expert Systems with Applications, 38, 3287–3293.
- [3] Chaitip P., Chaiboonsri C. and MukhjangR.(2008), Time Series Models for Forecasting International Visitor Arrivals to Thailand, Proceedings of International Conference on Applied Economics, 159-163.
- [4] Egrioglu E., Aladag C.H. and Gunay S. (2008), A new model selection strategy in artificial neural network, Applied Mathematics and Computation 195, 591-597.
- [5] Palmer A., Montano J.J. and Sese A. (2006), Designing an artificial neural network for forecasting tourism time series, Tourism Management 27, 781–790.

#### ÖZET

#### TÜRKİYE TURİZM VERİSİ ÖNGÖRÜSÜ İÇİN FARKLI PERFORMANS ÖLÇÜTLERİ

Gelen turist sayısı öngörü probleminin öneminden dolayı, literatürde bu problemin çözümü için birçok farklı yöntem önerilmiştir. Yapay Sinir Ağları (YSA) yöntemi belirtilen öngörü probleminin çözümünde kullanılan etkin yaklaşımlardan biridir. YSA birçok farklı öngörü probleminde başarısını kanıtlamış olmasına rağmen, halen kullanımında bazı problemler bulunmaktadır. En iyi öngörüler verecek mimarinin belirlenmesi belirtilen problemlerden biridir. Farklı performans ölçütleri hatayı farklı bir açıdan değerlendirmektedir.Bu nedenle, Eğrioglu vd. (2008) tarafından farklı performans ölçütlerini birleştiren ve hatayı aynı anda farklı açılardan ölçebilen Ağırlıklı Bilgi Kriteri (ABK) önermişlerdir. Yapılan çalışmada Antalya'ya, İstanbul'a ve Türkiye'ye gelen yabancı turist sayıları YSA ile öngörülmüştür. En iyi mimarinin belirlenmesinde çeşitli performans ölçütleri kullanılmıştır. Elde edilen tüm sonuçlar birbirleriyle karşılaştırılarak yorumlanmıştır.

**Anahtar kelimeler:** Öngörü,Performans ölçütü, Türkiye turizm verisi, Yapay sinir ağları, Zaman serileri

## BİLDİRİ OTURUMLARI 4

### SESSION 4

### Uygulamalı İstatistik 3

SIRALI VE ÇOK TERİMLİ LOJİT MODELLERİN UYUM İYİLİĞİ GÖSTERGELERİ AÇISINDAN  
KARŞILAŞTIRILMASI

Erkan ARI\*

Cumhuriyet Üniversitesi Fen Fak. İstatistik Böl., 58140, Sivas, Türkiye, eari@cumhuriyet.edu.tr

Zeki YILDIZ

Eskişehir Osmangazi Üniversitesi Fen Fak. İstatistik Böl., 26480, Eskişehir, Türkiye, zylidiz@ogu.edu.tr

Özet

Paralel doğrular varsayımı kimi zaman sağlanmakta kimi zaman ise sağlanamamaktadır. Bu varsayım sağlandığı durumda PO model kullanılırken, sağlanmadığı durumda ise NPO ve PPO modelleri kullanılabilir. Bu varsayım sağlanmadığı durumda kullanılacak bir model de Çok Terimli (Multinomial) Lojit modeldir. Bu çalışmada ele alınan modeller uyum iyiliği kriterleri bakımından karşılaştırılmış ve NPO ve PPO modelleri kullanmanın bağımlı değişkenin yapısındaki tüm bilgiyi kullanması ve sıralı yapının göz ardı edilmemesi açısından çok terimli lojit modele göre avantaj sağladığı sonucuna ulaşılmıştır.

1. Orantısız Olmayan Oran Model (Non-Proportional Odds Model) (NPO)

Fu (1998) tarafından önerilen "Orantısız Olmayan Oran Model" de lojit oluşturma esnasında yığılımlı lojitler kullanılmakta ve paralel doğrular varsayımı sağlanmamaktadır. Dolayısıyla bu modelde bağımsız değişkenlerin bağımlı değişken odds'una etkisi eşit değildir ve bağımlı değişken kategorisi  $m = 1, 2, \dots, M - 1$  ile gösterildiğinde  $\beta$  katsayıları bağımlı değişkenin her bir kategorisi için farklıdır (Fu, 1998). Orantısız Olmayan Oran modelde diğer adıyla Genelleştirilmiş Sıralı Lojit modelde, j'inci kategori için yığılımlı olasılık Orantısız Oran modelinde olduğu gibi Eşitlik 1. ile ifade edilir (Fullerton and Xu, 2012).

$$P(Y \leq y_m | x) = \frac{\exp(\alpha_m - x' \beta_m)}{1 + \exp(\alpha_m - x' \beta_m)} \quad m = 1, 2, \dots, M - 1 \quad [1.]$$

2. Kısmi Orantısız Oran Model (Partial Proportional Odds Model) (PPO)

Kısmi Orantısız Oran Modeli, paralel doğrular varsayımının kimi değişkenler için sağlandığı kimisi için ise sağlanmadığı durumda kullanılan ve güçlü varsayımı oldukça rahatlatan bir modeldir. Model hem orantısız hem de orantısız olmayan modellerin özelliklerini taşımaktadır (Peterson and Harrell, 1990).

3. Çok Terimli (Multinomial) Lojistik Regresyon Analiz

Çok terimli lojistik regresyon modelinde bağımlı değişkenin j'inci kategoriye düşme olasılığı  $\pi_j = P(Y = j)$  ile Eşitlik 2.'deki gibi ifade edilmektedir (Liao, 1994).

$$\pi_j = \frac{\exp(\sum_{k=1}^K \beta_{jk} x_k)}{1 + \sum_{l=1}^{j-1} \exp(\sum_{k=1}^K \beta_{lk} x_k)} \quad j = 1, 2, \dots, J - 1 \quad [2.]$$

Çok Terimli Lojit ve Sıralı Lojit Modeller uyum iyiliği açısından karşılaştırıldığında Çizelge 1.'deki sonuçlara ulaşılmıştır.

Çizelge 1. NPO, PPO, Çok Terimli Lojit Modellerinin Uyum İyiliği Göstergeleri Bakımından Karşılaştırılması

Uyum İyiliği Göstergesi	Model		
	NPO	PPO	Çok terimli
Mac Fadden R <sup>2</sup>	0,0832	0,0785	<b>0,0746</b>
Sapma Ölçüsü	370,2256	372,1018	<b>373,6878</b>
AIC	390,2256	388,1018	<b>393,6879</b>
BIC	424,3421	415,395	<b>427,8044</b>

Çizelge 1.'deki uyum iyiliği göstergelerine (Mac Fadden R<sup>2</sup>, Sapma Ölçüsü, AIC, BIC) göre Sıralı Lojit Modeller (NPO, PPO) veriye daha iyi uyum sağlamaktadır. Ancak burada uyum iyiliği kriterlerine bakılarak Paralel Doğrular Varsayımının sağlanmadığı durumlarda Çok Terimli (Multinomial) Lojit Modelin, Sıralı Lojit Modellerden daha iyi bir seçenek olup olmadığına dair kesin bir sonuç elde edilemez. Çünkü Çok Terimli Lojit model ile Sıralı Lojit Modellerinin odds kestirimi karşılaştırmaları birbirlerinden farklıdır. Sıralı Lojit Modellerde oluşturulan lojitlerde referans ve karşılaştırılan kategoriler yığılımlı olarak ele alınmaktadır. Dolayısıyla orantısız, orantısız olmayan ve kısmi orantısız oran modellerinin odds oranları yığılımlıdır. Çok Terimli Lojit modelinde ise sadece seçilen referans kategorisine göre diğer kategorilerin karşılaştırılması nominal olarak yapılmakta ve bağımlı değişkenin sıralı yapısı göz ardı edilmektedir. Dolayısıyla Paralel Doğrular Varsayımının sağlanmadığı durumlarda NPO ve PPO modelleri kullanmak bağımlı değişkenin yapısındaki tüm bilgiyi kullanması ve sıralı yapının göz ardı edilmemesi açısından avantaj sağlamaktadır.

4. KAYNAKLAR

- [1] Fu, Vincent, K., 1998. Estimating Generalized Ordered Logit Models, STATA Technical Bulletin, 44 p.
- [2] Fullerton, Andrew, S. and Xu, J., 2012. The Proportional Odds with Partial Proportionality Constraints Model for Ordinal Response Variables, Social Science Research, 41,182-198 p.
- [3] Liao, T.F., 1994. Interpreting Probability Models., Logit, Probit and Other Generalized Linear Models, Quantitative Applications in the Social Sciences, Sage Publications 88 p.
- [4] Peterson, B., Harrell, J.R., 1990. Partial Proportional Odds Models for Ordinal Response Variables, Applied Statistics, Vol 39, No. 2, 205-217 p.

ABSTRACT

The Comparison of Ordinal and Multinomial Logit Models In Terms of Indicators of Goodness of Fit

Paralel line assumption is sometimes ensured, sometimes not. When this assumption is ensured, PO model is used, otherwise NPO and PPO models can be used. When this assumption is not ensured, multinomial logit model is one of the models that can be used. The models dealt in this study is compared in terms of goodness of fit and it is concluded that using NPO and PPO models provide an advantage over multinomial logit model in terms of using all the data in the structure of dependent variable and not ignoring the ordinal structure.

**Key Words:** Non-Proportional Odds Model, Partial Proportional Odds Model, Multinomial Logit Model, Goodness of Fit.

**R×C BOYUTLU ÇAPRAZ TABLOLARDAKİ ÇOKLU KARŞILAŞTIRMA YÖNTEMLERİNİN  
İNCELENMESİ: BİR UYGULAMA**

Sengül CANGÜR\*, Handan ANKARALI, Özge PASİN  
Düzce Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim A.D., 81620, Düzce, Türkiye  
E-mail:sengulcangur@duzce.edu.tr

İkiden fazla grupta, nominal düzeyli bir sonuç değişkenine ait hesaplanan oranların karşılaştırılması için oluşturulan R×C çapraz tablosu uygun Ki-kare testleri (Pearson, Fisher-Freeman-Halton, Likelihood Ratio) ile değerlendirilmektedir. Test sonucunda, gruplar arasında anlamlı fark saptanmış ise hangi grup çifti/çiftlerinin farklı oranlara sahip olduğunu belirlemek için, varyans analizi modellerinde kullanılan çoklu karşılaştırma yöntemlerine benzer yöntemlerden yararlanılmaktadır. Literatürde yer alan ve oranlar arası farkı incelemek amacıyla kullanılan çoklu karşılaştırma yöntemleri, Bonferroni, Bonferroni-Holm, Simes-Hochberg, Hommel, Benjamini-Hochberg, Tukey ve Marascuilo yöntemleridir. Karşılaştırılacak oran sayısı, örneklem büyüklüğü ve gerçekte doğru olan sıfır hipotez sayısı gibi etkenlere bağlı olarak ortaya çıkan Family-Wise Hata oranı, False Discovery oranı ve testin gücü açısından söz konusu yaklaşımlar arasında farklılıklar oluşmaktadır. Çok boyutlu çapraz tablolardaki oranlar arası fark için kullanılan Bonferroni yöntemi, SPSS (ver.20) paket programında, diğer yöntemler ise Minitab, SAS, Matlab, Dataplot ve R program/programlama dillerinde yazılan bir makro aracılığı ile uygulanabilmektedir. Bu çalışmada, çapraz tablolarda ki-kare istatistiği anlamlı bulunduğu zaman bir başka ifadeyle ikiden fazla oran arasında farklılığın olmadığına dair kurulan sıfır hipotezi ret edildiğinde, literatürde adları sık geçmeyen ve yaygın kullanılmayan çoklu karşılaştırma yöntemlerinin teorik özellikleriyle birlikte tanımlanması ve sonuçlarının karşılaştırmalı olarak bir örnek üzerinde yorumlanması amaçlanmıştır. Bu amaçla fen lisesi 9, 10 ve 11. sınıfta okuyan toplam 261 öğrenciye uygulanan dikkat eksikliği (DE) tam ölçeğinden elde edilen veriler kullanılmıştır. Sonuçlar değerlendirildiğinde; 11. sınıfta öğrenim gören ve yüksek düzeyde dikkat eksikliği riski taşıyan öğrencilerin oranı (%39.2), 9. ve 10. sınıftakilerden (%13.3, %16.9) anlamlı seviyede yüksek bulunmuştur. 11. sınıfta öğrenim gören ve orta düzeyde dikkat eksikliği riski taşıyanların oranı (%54.1) ise hem 10. (Marascuilo dışında) hem de 9. sınıftakilerden anlamlı seviyede düşük bulunmuştur (%71.9, %80.6). Ayrıca Tukey, Bonferroni, Hommel ve Benjamini-Hochberg yöntemlerine göre hesaplanan 1. Tip hata yapma olasılık değerleri, başlangıçta belirlenen %5 değerine yakın olduğunda, ilk 3 yöntem ile benzer sonuçlar bulunmuştur, buna karşın Benjamini-Hochberg yöntemiyle daha küçük yanlışma olasılıklarının ise aynı olduğu belirlenmiştir. Buna ilaveten Marascuilo yönteminden elde edilen sonuç anlamlı bulunmamıştır. Ayrıca örnekteki koşulların çoğunluğunda, Benjamini-Hochberg yöntemiyle hesaplanan yanlışma olasılıklarının, diğer yöntemlerden elde edilenlere göre daha küçük bulunması literatür bilgisini desteklemektedir. Örnek sonuçlarına göre adı geçen yöntemler içerisinde en güçlü yöntem, Benjamini-Hochberg'tir ve bu bilginin de literatürle uyumlu olduğu görülmüştür. Sonuç olarak, ikiden fazla oran karşılaştırmasında bu yöntemlerin başarılı bir şekilde kullanılabilceği söylenebilir.

**KAYNAKLAR**

- [1] Benjamini Y. and Hochberg Y. (1995), Controlling the false discovery rate: a practical and powerful approach to multiple testing, Journal of the Royal Statistical Society, Series B 57(1): 289-300.
- [2] Hochberg Y. (1988), A sharper Bonferroni procedure for multiple tests of significance, Biometrika, 75(4): 800-2.
- [3] Holm S. (1979), A simple sequentially rejective multiple test procedure, Scandinavian Journal of Statistics, 6(2): 65-70.
- [4] Hommel G. (1988), A stagewise rejective multiple test procedure based on a modified Bonferroni test, Biometrika, 75(2): 383-6.
- [5] Zar J. H. (1999), Biostatistical analysis, 4th edit (Ryu T., ed.), Upper Saddle River, NJ, Prentice-Hall.

**ABSTRACT**

**INVESTIGATION OF MULTIPLE COMPARISON PROCEDURES IN R×C CROSS-TABLES: AN APPLICATION**

Multiple comparison procedures (MCPs), such as Bonferroni, Bonferroni-Holm, Simes-Hochberg, Hommel, Benjamini-Hochberg, Tukey, and Marascuilo can be used to investigate the difference between proportions in R×C cross-tables. The study aims to compare MCPs, when the Chi-square test-statistics is significant for cross-table. For this purpose, the relevant algorithms were applied to data about attention-deficit-disorder of high-school students. According to results, Benjamini-Hochberg method's the highest power. Marascuilo method has the least significant-differences while the others provide an equal number of significant-differences. However, all methods are different from each other for Type I error level. These methods can be used successfully for comparisons of more than two proportions.

**Key Words:** Proportions, Multiple Comparison Procedures, Bonferroni, Tukey, Benjamini-Hochberg, Marascuilo, Attention Deficit Disorder

## AZ GELİŞMİŞ ÜLKELERİN SAĞLIK VE ASKERİ HARCAMALARA İLİŞKİN DEĞİŞKENLER BAKIMINDAN İNCELENMESİ

Fatih ÇEMREK\*, Ayşe İŞİ\*\*, Hakkı POLAT\*

\*Eskişehir Osmangazi Üniversitesi Meşelik Kampüsü İstatistik Bölümü /Eskişehir

\*\*Gazi Üniversitesi Gazi Meslek Yüksekokulu Çubuk Yerleşkesi Çubuk/ANKARA

### 1. Giriş

Birleşmiş Milletlerin düzenli olarak yayınladığı, dünyadaki çoğu insanın en azından beş dakika bile olsa durup düşünmesine biraz da üzülmesine, ama sonra da hayatına kaldığı yerden devam etmesine sebep olan raporlarında, nedenleri, sorunları veya çözümleri tartışılacaktır, dünya üzerinde her yıl yaklaşık 12 milyon insanın açlık ve bunun sebep olduğu sağlık problemlerinden dolayı öldüğü belirtilmektedir. Uzmanlar için daha vahim tarafı olarak, bu şekilde devam edilirse bu sayının her geçen yıl artacağı gerçeğini ifade etmektedir [1].

ABD Kongre üyeleri için hazırlanan silah anlaşmaları raporuna göre 2009'da 58 milyar dolarlık silah anlaşması yapılmış, bu rakam 2012 de 85 milyar dolara çıkmıştır. Aynı yıl BM'nin 30 milyon insanın ihtiyaçlarını karşılamak için insani yardım çağrısı 7 milyar dolardır. Tüm dünya genelinde ise ABD'nin çeşitli ülkelerle sadece bir yıllık silah anlaşmalarının parasal değeri 66 milyar doları bulmaktadır [6].

Buna ek olarak silah ticareti yapılan ülkelerin %80'ine yakının Afrika ve Asya'daki az gelişmiş ve gelişmekte olan ülkeler olması son zamanlarda uzmanları dışarıdan yardımlardan ziyade, bu ülke hükümetlerinin elinde bulunan kaynaklarını silahlara değil de halkın temel ihtiyaçlarının hizmetine nasıl sunacaklarının düşünülmesi gerektiğine itmiştir [2,3,4].

Bu çalışmanın amacı da, dünya üzerinde kişi başına düşen milli hasıla miktarı olarak en düşük düzeye sahip 44 ülkenin, sağlık ve askeri harcama verilerini içeren bazı değişkenlerle incelenmesi, benzerlik ve farklılıklarının istatistiksel yöntemlerle ortaya konmasıdır. Bu amaçla tanımlayıcı istatistiklerin dışında söz konusu 44 ülke için kümeleme analizi uygulanmış ve ele alınan değişkenler bakımından ülkelerin benzerlikleri ortaya konmaya çalışılmıştır.

Çok değişkenli istatistiksel tekniklerden birisi olan kümeleme analizi, grup sayısı bilinmeyen ve gruplandırılmamış verilerin benzerliklerine göre sınıflandırılması amacıyla kullanılmaktadır. Kümeleme analizi, verilerin birimlere veya değişkenlere göre birbirlerine benzerlikleri bakımından ayrı kümelerde toplanmasını sağlayan bir tekniktir. Kümeleme analizi birbirine benzer olan bireylerin aynı gruplarda toplanmasını amaçlaması bakımından diskriminant analizi ile, birbirine benzer değişkenlerin aynı gruplarda toplanmasını amaçlaması nedeniyle de faktör analizi ile benzerlik göstermekte olup veri indirgeme özelliği vardır [5].

### 2. Bulgular

Çizelge 1'de kümeleme analizine tabi tutulan az gelişmiş ülkelerin yaptıkları silah ve sağlık harcamaları görülmektedir. Çizelge incelendiğinde, her ne kadar yapılan silah anlaşmalarının çoğunun resmi olmadığı, el altından yürütüldüğü uzmanlarca ifade edilmiş olsa da bu ülkelerin silah harcamalarının parasal büyüklüklerinin sağlık harcamalarına oranlarının %30'lara yaklaştığı görülmektedir. Çalışmada ayrıca daha ayrıntılı olarak ülkelerin yaptıkları bu harcamaların Gayri Safi Milli Hasıla'daki oranları, kişi başına düşen askeri ve sağlık harcamaları gibi karşılaştırmalı tablolara yer verilmiştir.

Çizelge 1. Bölgelere Göre Sağlık ve Silah Harcamaları

Satır Etiketleri	Sağlık Harcamaları (Milyon Dolar)	Silah Harcamaları (Milyon Dolar)
Afrika	36.727,1	6.838,1
Asya	23.388,8	10.610,8
Orta Amerika	733,6	44,1
<b>Genel Toplam</b>	<b>60.849,5</b>	<b>17.493,0</b>

### 3. Sonuçlar

5 yaş altı ölüm oranları (Binde), Gelişmiş İçme Suyu Kaynaklarının Kullanılabilir Oranı (%), Gelişmiş Sağlık Tesislerinin Kullanımı(%), Yetişkin HIV Oranı (%), Silah Harcamaları (Milyon Dolar) ve Sağlık Harcamaları (Milyon Dolar) söz konusu ülkeler için kümeleme metoduyla analiz edildiğinde, temelde 10 kümenin, 2. iterasyonda 5, 3. iterasyonda da 3 küme yapısının olduğu görülmüştür. Son olarak 4. iterasyonda ise tüm ülkelerin söz konusu değişkenler bakımından 2 ayrı kümeye toplandığı belirlenmiştir.

### KAYNAKÇA

[1] www.fao.org

[2] www.sipri.org

[3] www.unicef.org

[4] www.worldbank.org

[5] Tathdil H, (1996), Uygulamalı Çok Değişkenli İstatistiksel Analiz, Akademi Mat., Ankara.

[6] (<http://workshop.chromeexperiments.com/projects/armsglobe/>)

### ABSTRACT

#### Investigation of Least Developed Countries for Health and Military Expenditure Variables

According to UN reports there are 12 million people die for about starvation in the world every year. Another reports of UN point that 30 million people can recovered from starvation with just by 7 billion aids. Officially there was 85 billion dollars gun trades were made all around the world in 2012.

In this study, was searched about of their expenditure of army and health and some another indicators about health, for the 44 poorest country in the world. And this research show that, that poor countries spend amount of their budget about of 33% of their health expenditure.

**Key Words:** Gun Trades, Health Expenditure, Starvation, Clustering Analysis



## SEÇMENLERİN KARARLARINI ANLAMAK: BİR YAPISAL EŞİTLİK MODELİ ÖNERİSİ

Gamze ÇİNER ve Veysel YILMAZ

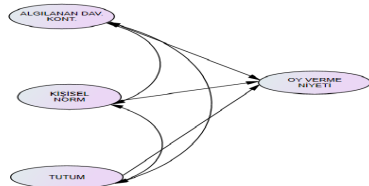
Eskişehir Osmangazi Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü,26000,Eskişehir,  
gamzecerin@gmail.com\* veyselyilmazhoca@gmail.com

### GİRİŞ

Bu çalışmada oy verme davranışının planlanan davranış teorisiyle açıklanması amaçlanmıştır ve yapısal eşitlik modellemesiyle geliştirilmiştir. Planlanmış davranış teorisi, Ajzen tarafından ortaya atılmıştır. Planlanmış davranış teorisi üç faktörün davranışı tetiklediğini öne sürmektedir: kişisel tutum, kişisel norm ve algılanan davranış kontrolü. Teoriye göre; bir davranış olumlu olarak algılandığı takdirde (kişisel tutum), o davranışın gerçekleştirilme olasılığı daha yüksektir. Yine teoriye göre bireyin önemli olarak kabul ettiği kişilerin, herhangi bir davranışa karşı tutumu (kişisel norm) olumlu ise bu tutum bireyi o davranışı gerçekleştirmeye itecektir. Son olarak bireysel algılar kişinin davranış üzerinde kontrol sahibi olduğu (algılanan davranış kontrolü) yönünde ise bu durum davranışın gerçekleştirilmesini daha fazla teşvik edecektir (Nunko ve Ramkissoon, 2010:529). Yani teoriye göre bütün davranışlar belli sebeplere bağlı olarak ortaya çıkar. Davranışların sonuçları önceden hesaplanır, ortaya çıkan sonuçlardan herhangi birine ulaşmak için karar verilir ve karar eyleme dönüştürülür (Özbey, 2010).

Planlanmış davranış teorisine göre üç değişken niyeti etkiler. Bu değişkenler; kişisel tutum, kişisel (subjektif) norm ve algılanan davranış kontrolüdür. Burada kişisel tutum, bireyin kendi inançlarına dayalı olarak bir davranışa karşı oluşturduğu olumlu ya da olumsuz düşünceleri ifade eder. Kişisel norm ise kişinin etrafındaki önemli olduğunu düşündüğü kişilerin fikirlerinin davranışa olan etkisini ifade etmektedir. Algılanan davranış kontrolü ise bir davranışı gerçekleştirmenin zorluğunu ya da kolaylığını ifade etmektedir (Rutherford ve De Vaney, 2009:1-16).

### ARAŞTIRMA MODELİ VE HİPOTEZLER



#### Şekil 1.Araştırma Modeli ( PDT)

Bu faktörlerle davranışa yönelik niyetin açıklanması için aşağıdaki hipotezler kurulmuştur.

Bu hipotezler;

H1: Tutum, davranışa yönelik niyeti önemli ölçüde etkilemektedir.

H2: Algılanan davranış kontrolü, davranışa yönelik niyeti önemli ölçüde etkilemektedir.

H3: Kişisel normlar, davranışa yönelik niyeti önemli ölçüde etkilemektedir.

H4: Algılanan davranış kontrolü ile kişisel norm arasında bir ilişki vardır.

H5: Algılanan davranış kontrolü ile tutum arasında bir ilişki vardır.

H6: Kişisel norm ile tutum arasında bir ilişki vardır.

### Analiz Sonuçları

Elde edilen seçmen örnekleminin %48'i kadınlardan ve %52'si ise erkeklerden oluşmaktadır. Örneklemin %58,9'u üniversite eğitim düzeyinde, %24'ü lise eğitim düzeyinde, %13,6'sı ilköğretim eğitim düzeyinde ve %3,5'u yüksek lisans eğitim düzeyindedir. Kişisel norm; oy verilen partiye yakın çevre desteği, algılanan davranış kontrolü; kişinin verdiği oya tutarlılığı ve tutum; kişinin verdiği oya inancı olarak açıklanır; AKP için, kişinin verdiği oya inancı ve kişinin verdiği oya tutarlılığı tekrar AKP'ye oy verme niyetini düşük oranda açıklarken, yakın çevre desteği, oy verme niyetini yüksek oranda açıklar(0,628). CHP için, yakın çevre desteği tekrar CHP'ye oy verme niyetini düşük oranda açıklarken, kişinin verdiği oya inancı tekrar CHP'ye oy verme niyetini yüksek oranda açıklar(0,56). MHP'de ise, yakın çevre desteği, tekrar MHP'ye oy verme niyetini açıklayamazken kişinin verdiği oya inancı ve kişinin verdiği oya tutarlılığı, tekrar MHP'ye oy verme niyetini çok düşük oranda açıklamaktadır.

### KAYNAKLAR

AJZEN, I.(1991)“The Theory of Planned Behavior, Organizational Behaviorand Human Decisions Processes”

HANSEN, D.,Jensen J.M. (2007) “Understanding voters’ decisions: a theory of planned behavior approach” Innovative Marketing

### ABSTRACT

#### UNDERSTANDING THE VOTERS' DECISIONS: A PROPOSAL OF STRUCTURAL EQUATION MODEL

This study aimed to explain the voting behavior of the theory of planned behavior and structural equation modeling has been developed. At the level of primary education for the AKP and the MHP, CHP voters rate than high. MHP voters to rate the level of secondary education AKP and the CHP was higher than, the highest rate of university and post-graduate education at the level of CHP voters is at.

**Key words:** Factor Analysis, Theory of Planned Behavior, AMOS, Structural Equation Model

### ÖLÜM NEDENLERİNİN MEKÂNSAL ANALİZİ

Adnan KARAIBRAHİMOĞLU<sup>1</sup>, Hakan BAŞBOZKURT<sup>2</sup>, Ayşe BAŞBOZKURT<sup>3</sup>, Yasin ASAR<sup>4</sup>, Aşır GENÇ<sup>5</sup>

<sup>1</sup>Selçuk Üniversitesi, Fen Fakültesi, İstatistik AD(Doktora öğrencisi) (Necmettin Erbakan Üniversitesi Meram Tıp Fakültesi Tıp Eğitimi ve Bilişimi Anabilim Dalı Uzmanı) Konya, Türkiye email: adnankaraibrahim@gmail.com

<sup>2</sup>Selçuk Üniversitesi, Fen Fakültesi, İstatistik AD (Doktora öğrencisi)Konya, Türkiye email: hakan.basbozkurt@gmail.com

<sup>3</sup>Edebiyat Fakültesi, İstanbul Üniversitesi (Doktora Öğrencisi)(Bingöl Üniversitesinde Fen-Edebiyat Fakültesi Coğrafya Bölümü Arş. Gör.), İstanbul, Türkiyeemail: zeynep\_ayse@yahoo.com

<sup>4</sup>Selçuk Üniversitesi, Fen Fakültesi, İstatistik AD (Doktora öğrencisi)(Necmettin Erbakan Üniversitesi Fen Fakültesi Matematik Bölümü Arş. Gör.), Konya, Türkiyeemail: yasinasar@gmail.com

<sup>5</sup>Selçuk Üniversitesi, Fen Fakültesi, İstatistik AD Fen Bilimleri Enstitüsü Müdürü ve İstatistik AD Başkanı Konya, Türkiye email: agenc@selcuk.edu.tr

Küreselleşen dünyamızda enformasyon hızla yayılırken birbiri ile yakın ilişki içerisindeki birimleri etkilemekte, bu etkileşim mekansal yakınlık içerisinde daha fazla olmaktadır. Bu durum, sağlık, sosyal, siyasal ve kültürel boyutlarda mekansal dağılımı etkilemekte ve mekansal etkileşimlerin analiz edilmesini gerektirmektedir. 80'li yıllardan itibaren mekansal analiz gündeme gelmeye başlamış, özellikle son yıllarda uygulamaları ileri boyutlara ulaşmıştır. Mekana bağlı olarak elde edilen verilerde komşuluk ilişkilerinin yönü ve boyutu araştırılmaya başlanmıştır. Mekansal analiz; coğrafi bilgi sistemleri, uzaktan algılama ve istatistik alanındaki hızlı gelişmeler sonucunda çok faydalı bilgiler veren bir analiz yöntemi haline gelmiştir.

Bu çalışmada, 2010 yılı Türkiye'nin İstatistik Birim Bölge Sınıflaması(İBBS-2), 2. Düzeyinde bulunan 26 bölgesine ait ölüm sebeplerinin mekansal ilişkisi incelenecektir. Kurulan modelde ölüm sebeplerini etkileyen demografik ve çevresel faktörlerin, sağlık verileri bağlamında bölge komşulukları üzerindeki etkileri araştırılacaktır. Bulguların ortaya konulmasında mekansal regresyon analizinin yanı sıra coğrafi bilgi sistemleri (CBS) görüntüleme yöntemleri de açıklanarak uygulanacaktır.

Anahtar kelimeler: Ölüm nedenleri, Mekansal Analiz, Mekansal regresyon, CBS, İBBS-2(NUTS-2)

### SPATIAL ANALYSIS OF CAUSES OF DEATH

As the information is spreading out fast in our globing world, it affects the units in a close contact with each other and this interaction can be seen between the nearer ones increasingly. This case affects the spatial distribution in health, social, political and cultural perspectives, and makes necessary to analyze the spatial interactions. Since 80's, the spatial analysis has begun to come up, especially in recent years, its applications have reached to further levels. The direction and size of the correlation of geographical neighborhood have been researching in spatial data. The spatial analysis has become an analysis method providing very useful information as a result of advances in geographical information systems, remote sensing and statistics.

In our study, we will examine the spatial relation of the causes of death belonging to 26 (Nomenclature of Territorial Units of Statistics 2 Level (NUTS-2)) regions of Turkey for 2010 year. In the model, it will be searched that the effects of demographical and environmental factors causing deaths on regional neighborhood in the context of health. We will use spatial regression method in analysis as well as the visualization techniques of geographical information system (GIS) in reveal of findings.

**Key words:** Causes of death, Spatial Analysis, Spatial regression, GIS, NUTS-2

### KAYNAKLAR

- [1] Bell,B.S.R.E. Hoskins, et al. (2006). "Current Practices In Spatial Analysis Of Cancer Data: Mapping Health Statistics To Inform Policymakers And The Public.", International Journal of Health Geographics.
- [2] Çolak, H.E.(2010). Coğrafi Bilgi Sistemleri ile Doğu Karadeniz Bölgesindeki Kanser Vakalarının Konumsal Analizleri. Harita Mühendisliği Ana Bilim Dalı. Trabzon, Karadeniz Teknik Üniversitesi. Doktora.
- [3] Jerret, M..R. Burnett, et al.(2005). "Spatial Analysis of Air Pollution and Mortalityin Los Angeles."Epidemiology 16:727-736
- [4] Semaan,S.,M.Sternberg, et al. (2007). "Social Capital and Rates of Gonorrhoea and Syphilis in the United States: Spatial Regression Analysis of State-LevelAssociations." Social Science & Medicine 64:2324-2341.
- [5] Fischer,M.M. and J. Wang (2011). Spatial Data Analysis: Models, Methods and Techniques, Springer.

**KKA VE K-EN YAKIN KOMŞULUK KULLANILARAK VZA'DA AĞIRLIK KISITLARININ  
BELİRLENMESİ**

Elvan AKTÜRK HAYAT\*, Olcay ALPAY  
Sinop Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü, 57000, Sinop, TÜRKİYE  
elvanakturk@gmail.com, olcayb@sinop.edu.tr

**Veri Zarflama Analizi'nde Değişken Seçimi**

Çoklu girdi ve çıktıya sahip, benzer yapıları olan karar verme birimlerinin (KVB) görelî etkinliklerini değerlendiren Veri Zarflama Analizi (VZA), son yıllarda yöneylem araştırması ve yönetim bilimlerinde sıklıkla kullanılan bir yöntemdir. VZA, Farrell tarafından önerilen sınır (frontier) analizine dayanılarak Charnes, Cooper ve Rhodes tarafından geliştirilmiştir.[1]

VZA, matematiksel programlama tabanlı bir tekniktir ve herhangi bir birimin etkinliği, CCR modeli olarak bilinen temel etkinlik modeli ile ölçülür.[2] Girdi odaklı CCR modeli (1) nolu eşitlikle verilmektedir.

$$z_p = \max \sum_{r=1}^s u_r y_{rp}$$

Kısıtlar:

(1)

$$\sum_{i=1}^m v_i x_{ip} = 1$$
$$\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \quad j = 1, \dots, n$$
$$u_r \geq 0, \quad r = 1, \dots, s$$
$$v_i \geq 0, \quad i = 1, \dots, m$$

Burada, p indisi etkinliği hesaplanacak KVB'yi, x girdileri, y çıktıları, z ilgili birimin etkinlik değerini, n KVB sayısını, m girdi sayısını, s çıktı sayısını göstermektedir. Bu modelin çözümünden elde edilen etkinlik skoru 1 olan KVB'leri, etkin olarak, etkinlik skoru 1'in altında olan her birim ise etkin olmayan olarak değerlendirilir.

Herhangi bir KVB için tahmin edilen etkinlik, modeldeki girdi ve çıktı sayısına bağlıdır. Bir veri setinde girdi ve çıktıların toplam sayısı KVB'lerin toplam sayısına yaklaştığında VZA problemlili olabilmektedir. Uygulamalarda ise genellikle KVB sayısı az, değişken sayısı fazladır. Çok sayıda girdi ve çıktı olması durumunda, önemli olmayan bir değişkeni analizden çıkarmak yerine ağırlık kısıtlamalarına gidilmesi daha uygundur. Bu nedenle değişken seçimi ve ağırlık kısıtlamaları VZA'da en önemli aşamadır. [3]

Sengupta (1990), VZA'da yer alabilecek değişkenlerin belirlenmesinde Kanonik Korelasyon Analizi'nin (KKA) kullanılabilirliğini göstermiştir.[4] VZA için ağırlık kısıtlarının tanımlanmasında KKA'nın kullanılabilirliği Gonçalves vd. tarafından gösterilmiştir.[3]

Çoklu regresyon analizinin genelleştirilmiş şekli olarak kabul edilen KKA, iki ve daha fazla değişken içeren iki değişkenler seti ( $X_1, X_2, \dots, X_p; Y_1, Y_2, \dots, Y_q$ ) arasındaki ilişkiyi doğrusal bileşenler aracılığı ile değerlendiren çok değişkenli bir yöntemdir. KKA'nın temel amacı, iki değişken setinin her biri için maksimum korelasyonlu ve birim varyanslı doğrusal bileşenler elde edip, değişken setleri arasındaki ilişkiyi maksimize edecek en uygun boyutu belirlemektir.[5]

**BİLDİRİ OTURUMLARI 4**

*SESSION 4*

*İstatistik Yöntemler*

Bu çalışmada, VZA'da ağırlık kısıtları için kullanılacak diğer bir yöntem olarak veri madenciliği tekniklerinden K-en yakın komşu algoritması ele alınmıştır. K-en yakın komşu algoritması, sınıflandırma problemini çözen denetimli öğrenme yöntemleri arasında yer alır.

Çalışmada, çok boyutlu veriler söz konusu olduğunda VZA'da kullanılacak değişkenler arasındaki ilişkiye bağlı olarak hesaplanan ağırlıklar matrisi, KKA ve k-en yakın komşu algoritması kullanılarak belirlenmeye çalışılmıştır. Bu iki yöntemden elde edilen ağırlıklar kullanılarak 2 ayrı VZA modeli elde edilmiş ve KVB'lerin etkinlikleri karşılaştırılmıştır.

#### KAYNAKLAR

- [1] Charnes A., Cooper W.W. and Rhodes E. (1978), Measuring the Efficiency of Decision Making Units, European Journal of Operational Research, Vol.2 No.6,429-444.
- [2] Cooper, W.W., Seiford, L.M., and Tone, K. (2000), Data Envelopment Analysis, Boston USA, Kluwer Academic Publishers.
- [3] Gonçalves A.C., Almeida R.M.V.R., Lins M.P.E. and Samanez C.P. (2013), Canonical Correlation Analysis In The Definition of Weight Restrictions for Data Envelopment Analysis, Journal of Applied Statistics, Vol. 40, No.5, 1032-1043.
- [4] Sengupta J.K. (1990), Tests of Efficiency in Data Envelopment Analysis, Computers and Operations Research, Vol.17, Issue 2, 123-132.
- [5] Anderson T.W. (1999), Asymptotic Theory for Canonical Correlation Analysis, J Multivar Analysis, 70(1):1-29.

#### ABSTRACT

#### DETERMINATION OF WEIGHT RESTRICTIONS USING BY CCA AND K-NEAREST NEIGHBOR IN DEA

Data Envelopment Analysis(DEA) is becoming an increasingly popular management tool. The goal of the DEA is to evaluate the relative performance of similar Decision Making Units(DMUs). Variable selection and weight restrictions important issues in DEA. The estimated efficiency for any DMU depends on the inputs and outputs included in the model. In this study, Canonical Correlation Analysis (CCA) and k-nearest neighbor algorithm are used to determine the weight restrictions matrices and compared the efficiencies of DMUs with using DEA.

**Key Words:**Data Envelopment Analysis, Canonical Correlation Analysis, K-nearest neighbor

#### BAĞIMLI İKİ ÖRNEKLEM TESTLERİNDE YENİDEN ÖRNEKLEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Özge KARADAĞ<sup>1</sup>, Hülya OLMUŞ<sup>2</sup>, Serpil Aktaş ALTUNAY<sup>1</sup>,

<sup>1</sup>Hacettepe Üniversitesi, Fen Fakültesi

İstatistik Bölümü, 06800 Çankaya/Beytepe/ANKARA

ozgekaradag@hacettepe.edu.tr, spxl@hacettepe.edu.tr

#### 1. Bağımlı İki Örneklem Karşılaştırmasında Yeniden Örnekleme Yaklaşımı

Örneklem bağımlı olduğu durumlarda yani aynı örneklem üzerinde yapılan iki farklı ölçüm söz konusu olduğunda bu ölçümler arasındaki farkın anlamlı olup olmadığını test etmek amacıyla bağımlı / eşleştirilmiş örneklem t-testi kullanılır. Bağımlı iki örneklemde,  $X_1$  ve  $X_2$  eşleştirilmiş iki örneklem ve  $D$  değişkeni fark vektörü,

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ \vdots & \vdots \\ X_{n1} & X_{n2} \end{bmatrix} \quad D = \begin{bmatrix} X_{11} - X_{12} \\ X_{12} - X_{22} \\ \vdots \\ X_{n1} - X_{n2} \end{bmatrix} \quad (1)$$

tanımlansın ve burada amaç

$$H_0: \mu_D = 0$$

yokluk hipotezini test etmektir.

Eşleştirilmiş iki örneklem arasındaki değerlere ilişkin fark değerlerinin normal dağılım göstermesi gerekir. Ancak farkların dağılımı normal dağılıma uyum göstermiyorsa, eşleştirilmiş t testi uygun olmaz. Bu durumda alternatif çözüm olarak parametrik olmayan yöntemlerden Wilcoxon işaretli sıra testi tercih edilir. Ancak son yıllarda, parametrik olmayan yöntemlere alternatif olarak yeniden örneklem (resampling) yöntemleri geliştirilmiştir. Yeniden örneklem, kitenin dağılımı bilinmediğinde ve diğer yöntemler geçersiz olduğunda oldukça kullanışlı yöntemlerdir. Yeniden örneklem yöntemlerinin temeli, teorik dağılımlara dayanmakta ve aynı örneklem içerisinde yeniden örnekler seçmeye bağlı olmaktadır.

Liteatürde yer alan yeniden örneklem metodlarından biri permütasyon testi diğeri bootstrap yöntemidir. Rasgeleleştirme testi olarak da adlandırılan permütasyon testinin temeli sıralı istatistiklere dayanan bir yeniden örneklem tekniğidir. Diğer bir yeniden örneklem tekniği olan Bootstrap Örneklemesi metodu ise, yoğun matematik formüllerden uzak, sınırlı varsayımlara sahip, anlaşılması ve kullanılması oldukça kolay bir yöntemdir [3]. Bootstrap'ın temeli n hacimli rasgele örneklemde, yerine koyarak seçim yöntemi ile yeniden n hacimli örneklem yaratıp, ilgilenilen istatistiğin defalarca gözlemlenmesine dayanır. Ayrıca bootstrap yönteminde örneklem seçilmesi yerine koyma yöntemine göre, permütasyon testinde ise yerine koymadan yöntemine yapılmaktadır [1, 2].

Yeniden örneklem tekniği ile elde edilen örneklem üzerinden bağımlı örneklem temelinin oluşturulan fark değerleri  $D_i^*$  ile gösterilirse, fark değerlerinin ortalaması ve varyansı aşağıdaki eşitliklerle hesaplanır:

$$\bar{D}_n^* = \frac{\sum_{i=1}^n D_i^*}{n} \quad \sigma_n^{2*} = \frac{\sum_{i=1}^n (D_i^* - \bar{D}_n^*)^2}{n-1} \quad (2)$$

Bu değerler yerine konulduğunda fark değerlerine ait  $T_n$  istatistiği eşitlikte verildiği gibi elde edilir.

$$T_n = \frac{\bar{D}_n^*}{\sigma_n^*/\sqrt{n}} \quad (3)$$

Çalışmanın amacı, bağımlı gruplarda eşleştirilmiş t-testi ve alternatifi olan Wilcoxon işaretli sıra testine alternatif olarak yeniden örneklem yöntemlerinden permütasyon testi ve Bootstrap yöntemi ile bir simülasyon çalışması yapmaktır. Bu nedenle, farklı dağılımlar, farklı örneklem büyüklükleri, farklı korelasyon yapıları düşünülerek, ele alınan yöntemlerin testin gücü bakımından karşılaştırılmıştır. Sonuç olarak, yeniden örneklem yöntemlerinden permütasyon ve bootstrap yönteminin, eşleştirilmiş örneklem t-testine bir alternatif olarak kullanılabilirliğini ortaya koymak ve bazı durumlarda daha iyi sonuç verdiğini göstermek amaçlanmıştır.

#### KAYNAKLAR

- [1] Butar F. and Bandulasiri A. (2009), Comparison of the power of the paired samples using permutation tests, Journal of Mathematical Sciences & Mathematics Education Vol. 4 No. 2.
- [2] Konietschke F. and Pauly M. (2012), Bootstrapping and Permuting paired t- test type statistics, Stat Comput. DOI 10.1007/s11222-012-9370-4.
- [3] Simon, L. and Bruce P (1991), Resampling: A Tool for Everyday Statistical Work, Chance, 4, 22-32.

#### COMPRASION OF RESAMPLING METHODS FOR PAIRED SAMPLE TESTS

#### ABSTRACT

In this study, it is attempted to compare the resampling methods including bootstrap and resampling and classical methods for paired sample test. A simulation study was performed to test the difference between the paired samples for both parametric and nonparametric methods under different conditions by taking different distributions, different population correlations and also different sample size. Results are interpreted based on simulation study with respect to their power of test values.

**Key Words:** permutation test, Bootstrap sampling, paired t test

#### YOL-BAKIM BİRİMLERİNDE YOL-BAKIM ARAÇLARI İLE YOL GÜVENLİĞİNİ ARTTIRMAK İÇİN TEDBİRLERİN PLANLANMASI

Urfat NURİYEV\*, Farnaz GOLGHASEMI SORKHABI  
Ege Üniversitesi, Fen Fakültesi, Matematik Bölümü, Bornova/ İzmir, 35100,Türkiye

urfat.nuriyev@ege.edu.tr

#### Giriş

Yol-bakım araçları ile yol güvenliğini arttırmak için tedbirlerin planlanması metodolojisi aşağıdaki durumlara bağlıdır. Yol ağlarında herhangi bir yolun veya onun bir kısmı için yol güvenliği, gereken güvenilirlik koşulları ve yolun özelliklerinin uygun olup olmamasıyla belirlenir. Metodolojiye göre planlanan periyodun başlangıcından önce belirlenen  $K_k$  kompleks göstericisinin değerinin yolun güvenliği olan ve planlanan seviyelerin uygunluğu ile karakterize olur [1].

Yolun tamir ve yeniden kurulması için ayrılmış tüm maliye kaynakları göz önüne alınarak  $K_k$  değişkeninin o andaki değerini belirledikten sonra yolun güvenliğini arttırmak için uygun  $\Pi_1 - \Pi_{19}$  sade göstericilerine göre hareketin güvenliğini kvalimetrik modeli kurulur (yani, yolun geçiş kısmının genişletilmesine  $\Pi_1$  göstericisinin iyileştirilmesi uygun gelir, köprülerin genişletilmesine  $-\Pi_3$  göstericisi vb.) [2].

Yol güvenliğinin artırılması tedbirlerinin esaslandırılması, belirlenen bütçe kapasitesini aşmadan yol güvenliğinin tasarlanan kompleks göstericisinin maksimal değerini belirlemeye getirilir.

Matematiksel Model

Problemin matematiksel modeli aşağıdaki gibi ifade edilir:

$$\sum_{i=1}^{19} c_i(x_i - \Pi_i) \leq C \quad (1)$$

$$\Pi_i \leq x_i \leq 5, \quad i=1,2,\dots,19, \quad (2)$$

k.a.

$$\text{enb } K_k = \sum_{i=1}^{19} a_i x_i \quad (3)$$

Burada,  $x_i$  ve  $\Pi_i$  - hareketin tehlikesizliğine yol durumlarının etkisini gösteren kvalimetrik modelin basit göstericilerinin giriş ve çıkış değerleridir:

$a_i$  - yukarıdaki göstericilerin değerliliği katsayıları;

$c_i$  - göstericilerin değerini bir puan artırılmasını sağlayan tedbirlerin birim değeri;

$C$  - bir yıl için hareketin güvenliğini arttırmak için tedbirlere ayrılan finans kaynaklarının planlanan kapasitesi.

Problem daha açık bir şekilde yazılabilir:

$$C_1(x_1 - \Pi_1) + C_2(x_2 - \Pi_2) + \dots + C_{19}(x_{19} - \Pi_{19}) \leq C, \quad (4)$$

$$\Pi_i \leq x_i \leq 5, (i=1,2,\dots,9,11,\dots,19), \quad (5)$$

$$C_{10} = 0, \quad (6)$$

$$C_{11k} = 0. \quad (7)$$

k.a.

$$\text{enb } K_k = 0.15x_1 + 0.06x_2 + 0.05x_3 + 0.11x_4 + 0.13x_5 + 0.2x_6 + 0.03x_7 + 0.03x_8 \\ + 0.03x_9 + 0.08K_{11k} + 0.10x_{16} + 0.03x_{17} + 0.01x_{18} + 0.02x_{19} \quad (8)$$

Bu model ise değişkenleri iki taraftan da sınırlı tamsayı tekboyutlu Knapsak problemidir [3].

#### KAYNAKLAR

[1] Sidenko V.M. ve d. ,(1973), Avtomobil'nye dorogi (sovershenstvovanie metodov proektirovaniya i stroitel'stva)) [Highways (Improvement of Design and Construction methods)]. Kiev, Budivel'nik Publ , 278 p.

[2] Sidenko V.M. ,(1981), Rokas S. Yu. Upravlenie kachestvom v dorojnom stroitel'stve. Moskva, Transport, 252 p.

[3] Bakır M. A. ve Altunkaynak B., T., (2003), Tamsayı Programlama, Ankara, Nobel Yayınları, 620 s.

#### ABSTRACT

##### ROAD MAINTENANCE ORGANIZATIONS, BY THE MEANS OF ROAD MAINTENANCE TOOLS AND PLANNING OF MEASURES, IMPROVE THE SAFETY OF THE MOVEMENT.

In this study, the road-maintenance services, tools and road maintenance organizations of the movement to improve the safety measures optimally Planning of an integer programming model is proposed. The variability of the suggested model, in both sides, is presented in a single bounded Knapsack Problem.

**Key Words:** Mathematical modeling, movement safety, Knapsack Problem

#### YAŞAM SÜRDÜRME ANALİZİNDE GAMMA KIRILGANLIK MODELİ

Emel BAŞAR

Gazi Üniversitesi Fen Fakültesi İstatistik Bölümü 06500 Ankara Türkiye  
ebasar@gazi.edu.tr

#### Giriş

Yaşam sürdürme analizinin amacı, önceden tanımlanan herhangi bir olayın meydana gelmesine kadar geçen zaman sürelerinden oluşan ve durdurulmuş (censored) gözlemlerin bulunduğu veriyi çözümlenektir. Yaşam sürdürme zamanı üzerinde etkili olduğu düşünülen eşdeğişkenlerin (covariates) yer aldığı modeller, verinin tanımlanan risk bakımından çözümlenmesinde sıklıkla kullanılmaktadır. Ancak her zaman modelin içinde yaşayan ve eşdeğişkenler tarafından açıklanamayan bir değişkenlik bulunabilmektedir. Açıklanamayan bu değişkenlik risk bakımından önemli olabilmektedir. Eşdeğişkenler tarafından açıklanamayan ve değişimi gözlemlenemeyen heterojenlik (unobserved heterogeneity) modellenmek istenildiği zaman kırılğanlık (frailty) kavramı kullanılmaktadır [1]. Kırılğanlık modellerinde bireyler arasında önemli farklılıkların var olduğu ve temel bir rastgelelik (randomness) bulunduğu varsayılmaktadır.

Yaşam sürdürme analizinde asıl önemli olan ve çözümlenmek istenen kavram hazard hızı olmaktadır. Hazard hızının zaman boyutunda izlediği yol, gözlem altındaki bireylerin karşı karşıya kaldığı riskin yine zaman boyutunda gösterdiği davranışla tanımlanmaktadır. Böylece hazard hızının biçimi gözlemlendiği zaman, bunun bireylere ilişkin ölçülemeyen etkilerden mi kaynaklandığı yoksa bireyler arasındaki farklılığa mı bağlı olduğunu bilmek zor olmaktadır. Kırılğanlığın varlığı, birbirinin içine geçmiş pek çok hatanın ortaya çıkmasına yol açmakta ve yanlış sonuca ulaşılmasına neden olabilmektedir. Bu bakımdan kırılğanlık kavramı yaşam sürdürme analizinde büyük bir öneme sahip olmaktadır.

#### Kırılğanlık Modelleri

Kırılğanlık bireyler arasında önemli farklılıklar olduğunu varsaymaktadır. Ortaya çıkan bu farklılıkların ölçülebilmesi ve risk bakımından değerlendirilmesi için çeşitli modeller kurulabilmektedir. Bir kırılğanlık modelinde esas alınan temel düşünce, bireyler arasındaki gözlemlenemeyen heterojenliği, rastgele etki olarak hazard fonksiyonuna dâhil edebilmektedir. Bu bakımdan kırılğanlık modelleri koşullu modellerdir ve kırılğanlık faktörü belli bir dağılıma sahip olan rastgele bir değişkendir.

Gözlemlenen yaşam sürdürme zamanlarını gösteren rastgele değişken  $T$  ve gözlemlenemeyen rastgele kırılğanlık değişkeni  $Z$ , orantılı hazard varsayımı altında aşağıda verildiği gibi birleştirilebilir.

$$h(t|Z) = Zh_0(t) \quad (1)$$

Burada  $h_0(t)$  temel (baseline) hazard fonksiyonu ve  $Z$  negatif olmayan rastgele değişkendir.  $Z$ 'nin değişkenliği bireyler arasındaki heterojenliğin derecesini belirlemektedir ve kırılğanlık faktörü  $Z$ , temel hazard fonksiyonu üzerinde çarpımsal bir rol oynamaktadır [2].

Koşullu yaşam sürdürme fonksiyonu ise;

$$S(t|Z) = e^{-ZH(t)} \quad (2)$$

olarak ifade edilebilir. Burada  $H(t) = \int_0^t h_0(u)du$ , birikimli temel hazard fonksiyonudur. Marjinal yaşam sürdürme fonksiyonu  $S(t)$  aşağıda verildiği gibi elde edilebilmektedir.

$$S(t) = E[S(t|Z)] = E[e^{-ZH(t)}] \quad (3)$$

(3) eşitliği ile verilen beklenen değer, kırılgnlık dağılımına ilişkin Laplace dönüşümü ile elde edilebilmektedir.

Eşdeğişken bilgisinin modelde yer alması durumunda,  $Y$  eşdeğişken bilgisini göstermek üzere, (1)'de verilen model aşağıdaki biçimde ifade edilebilir;

$$h(t|Z, Y) = Zr(Y)h_0(t) \quad (4)$$

Burada  $r(Y)$  bilinmeyen bir risk fonksiyonudur. Yukarıda verilen modellerde Kırılgnlık faktörünü tanımlamak üzere kullanılan dağılımlar, parametre tahminlerini elde etmek bakımından Laplace dönüşümünün sağlandığı dağılımlar olmaktadır [3].

Bu çalışmada, risk fonksiyonu olarak çeşitli parametrik dağılımlar ve yarı parametrik bir model olan Cox orantılı hazard modeli dikkate alınarak, kırılgnlık faktörünün gamma dağılımına sahip olması durumu incelenmiş ve modellerin gerçek bir örnek üzerinde uygulaması yapılmıştır.

#### KAYNAKLAR

- [1] Vaupel, J. W., Manton, K. G., Stallard, E. (1979), The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16, 439-454.
- [2] Hanagal D.D. (2011), *Modeling Survival Data Using Frailty Models*, Chapman & Hall,
- [3] Aalen, O. O., Borgan, Ø., Gjessing, H., K. (2008), *Survival and Event History Analysis: A Process Point of View*, Springer-Verlag, New York.

#### ABSTRACT

##### GAMMA FRAILITY MODEL IN SURIVAL ANALYSIS

Frailty models are often used to describe the unobserved heterogeneity in survival data. The frailty factor is usually assumed to act multiplicatively on a baseline hazard function. Frailty models are conditional models and the frailty factor is random. Therefore a frailty distribution needs to be specified in the frailty model. In this study gamma frailty model has been considered and the model has been illustrated with real data.

**Key Words:** Survival analysis, proportional hazards model, gamma frailty, parametric models

#### BLOKLANMIŞ $2^{n-k}$ KESİRLİ ÇOK ETKENLİ TASARIMLARDA EN İYİ BLOK SEÇİMİ

Erdoğan KOLAY\*, Nazan DANACIOĞLU  
Sinop Üniversitesi Fen Edebiyat Fakültesi, İstatistik Bölümü, 57000, Sinop, Türkiye

erdinckolay@gmail.com

##### $2^{n-k}$ Kesirli Çok Etkenli Tasarımlar ve Bu Tasarımlarda Bloklaşma

Birden fazla etkenin yanıt değişkeni üzerindeki etkisinin etken düzeylerinin olası tüm kombinasyonlarının denenecek araştırıldığı deneyler çok etkenli (ÇE) deneylerdir. Bu tasarımların en önemli özelliği etkenlerin yanıt değişkeni üzerindeki etkisi hesaplanırken bütün gözlemlerin kullanılması ve etkileşimler hakkında bilgi elde edilebilmesidir (Montgomery, 1984).

ÇE bir tasarımda  $n$  etken varsa ve her bir etken iki düzeyli ise tamamlanmış ÇE tasarım;

$$\underbrace{2 \times 2 \times 2 \dots \times 2}_n = 2^n$$

etken içerir.  $2^n$  tasarımları, etken sayısı arttığında deney maliyetini azaltıklarından tercih edilmektedir. Örneğin 4 etkenli bir tasarımda, etken düzeyleri 2'den 3'e çıkarılırsa tasarımın genişliği  $2^4 = 16$ 'dan  $3^4 = 81$ 'e çıkar ve deneme kombinasyonu sayısı neredeyse 5 kat genişler (Mee, 2009).

Tamamlanmış bir  $2^n$  ÇE deneyinde etken sayısı arttığında ve özellikle denemeler pahalı olduğunda, deneme kombinasyonlarının tamamının denemesi zordur. Bu nedenle deneme sayısının az olduğu ve tamamlanmış bir  $2^n$  ÇE deneyinin bir alt grubunun ya da kesirinin kullanıldığı KÇE (kesirli çok etkenli) (fractional factorial) ya da kesirli tekrarlar (fractional replication) tercih edilir. Bu tasarım sınıfının pratik ve teorik önemi Box, Hunter ve Hunter (1978) tarafından gösterilmiştir (Mukerjee ve Wu, 2006).

Aynı zamanda, KÇE bir tasarımı bloklara ayırmak da gerekli görülebilmektedir. Bloklaşma, deneylerde genellikle sistematik gürültüyü (systematic noise) kontrol etmek için kullanılır. Böyle gürültüler, günden güne değişim ya da gruptan gruba değişimden meydana gelebilir. Bloklaşma olmadan sistematik gürültü, etki tahmini etkisini ve tahminin doğruluğunu etkileyebilir.

Bir  $2^{n-k}$  KÇE tasarımını bloklara ayırırken, optimal blok yapısının nasıl belirleneceği sorunu ortaya çıkmış; son yıllarda, özellikle 2-düzeyle KÇE tasarımları ele alan çalışmalar yapılmıştır.

Sun, Wu ve Chen (1997) (SWC), 2-düzeyle KÇE tasarımlarda kabul edilebilirlik (admissibility) ölçütünü öne sürerek optimal blok yapılarını (optimal blocking schemes) katalog şeklinde sunmuşlardır. Sitter, Chen ve Feder (1997) (SCF), kesirli çözüm (fractional resolution) ve artırılmış kelime uzunluklarını (refinement of word length) önermiş ve kelime uzunluğu yapısını (word length pattern) En Az Sapma (EAS) ölçütüne uyarlayarak optimal blok yapılarını elde etmeye çalışmışlardır. Chen ve Cheng (1999) (CC), kelime uzunluğu yapılarını tahmin kapasitesine (estimation capacity) göre oluşturmuştur. Cheng ve Wu (2002) (CW), tahmin kapasitesi en yüksek olacak şekilde kelime uzunluğu yapısını tekrar düzenleyerek; Xu (2003), en az moment sapma (minimum moment aberration) (EAMS) ölçütünü kullanarak optimal blok yapılarını listelemişlerdir. Xu ve Lau (2006), en az moment sapma ölçütünü bloklaşmış KÇE'ler için genişletmiş ve tasarımları karşılaştırırken bu yöntemi kullanmışlardır.

$2^q$  blokta düzenlenen  $2^{n-k}$  KÇE tasarımlarında en iyi blok yapısının seçimi EAS ve EAMS ölçütlerine göre yapılır. Bu çalışmada bloklaşmış iki düzeyli KÇE tasarımlarda optimal blok yapısının belirlenmesinde kullanılan ölçütler  $2^{7-3}$  ve  $2^{8-2}$  tasarımlarına uygulanarak karşılaştırılmıştır.

#### KAYNAKLAR

[1] Montgomery, D. C. 1984. Design and Analysis of Experiments, Second Edition. John Wiley&Sons, New York, 538p.

[2] Sun, D.X., Wu, C.F.J., Chen, Y. 1997. Optimal blocking schemes for  $2^p$  and  $2^{n-p}$  designs. Technometrics, 39: 382-390.

[3] Xu, H., Lau, S. 2006. Minimum aberration blocking schemes for two and three level fractional factorial designs. Journal of Statistical Planning and Inference, 136:4088 4188.

#### ABSTRACT

#### CHOICE OF OPTIMAL BLOCK SCHEMES IN BLOCKED $2^{n-k}$ FRACTIONAL FACTORIAL DESIGNS

Factorial designs are widely used in many areas of scientific investigation. A full  $2^n$  factorial design requires all combinations of two levels of each of  $n$  variables. Fractional factorial designs permit investigation of the effects of many factors in fewer runs than a factorial design. Blocking is commonly used in design of experiments to reduce systematic variation and at the same time to increase precision of effect estimation. In literature several criteria have been proposed for ranking blocked fractional factorial designs. In this paper minimum aberration criterion and minimum moment aberration criterion were studied to compare blocked fractional factorial designs.  $2^{7-3}$  and  $2^{8-2}$  designs were investigated to choose optimal blocked design.

**Key Words:** Fractional Factorial Designs, Minimum Aberration, Minimum Moment Aberration, Optimal Blocking Schemes.

#### İKİ ORTAK BOZULMA KRİTERLİ SİSTEMLERİN PROFUST GÜVENİLİRLİKLERİ

Sevcan DEMİR ATALAY, Ali MERT

Ege Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 35040.Bornova, İzmir, TÜRKİYE.  
sevcan.demir@ege.edu.tr, ali.mert@ege.edu.tr

#### 1. İki Bozulma Kriterli Sistem

Literatürde sistemler yapılarına göre doğrusal (lineer) veya dögüsel, çalışma prensiplerine göre F veya G sistemleri, değerlendirilme karakteristiklerine göre bulanık veya bulanık olmayan vb. şekilde farklı özelliklerine göre sınıflandırılmaktadır. Bunun yanı sıra seri, paralel, n-den-f-çıkışlı ve ardıl n-den-k-çıkışlı gibi sistem yapıları literatürde yer almaktadır. Chiang ve Niu'nin (1981) bahsettiği n yedek istasyonlu bir telekomünikasyon sistemi ile n pompa istasyonlu petrol boru hattı sistemi ardıl n-den-k-çıkışlı:F sistemine ilişkin iki örnektir. n-den-f-çıkışlı:F(G) sistemi en az f bileşen bozulduğunda (çalıştığında) bozulan (çalışan) sistemdir. Benzer şekilde ardıl n-den-k-çıkışlı:F(G) sistemi en az ardıl k bileşen bozulduğunda (çalıştığında) bozulan (çalışan) sistemdir. Kuo ve Zuo'ya (2003) ait "Optimal Reliability Modelling" adlı çalışma sistemlerle ilgili geniş bir içerik sunmaktadır. Sözü edilen bu iki sistemde de tek kriterli bir sistem tasarımı söz konusudur. Tek bir bozulma kriteri yerine iki ortak bozulma kriterinin göz önüne alındığı güvenilirlik yapıları da tanımlanmaktadır. Bu yapılar sistemin gereksinimlerine göre yukarıda verilen iki kriterin "VE" ve "VEYA" mantıksal operatörleri ile birleştirilmesi yoluyla oluşturulur. Bu çalışmada "VE" operatörü ile birleştirilmiş sistemler ele alınmış olup, n elemanlı bir sistemde en az k ardıl çalışmayan (çalışan) "VE" en az f çalışmayan (çalışan) bileşen olduğunda çalışmayan (çalışan) sistem olarak tanımlanan  $\langle n, f, k \rangle: F(G)$  sistemleri incelenmektedir. Çalışmada doğrusal ve dögüsel yapıda iki ortak bozulma kriterli  $\langle n, f, k \rangle: F(G)$  sistemlerinin Profust güvenilirlikleri irdelenmektedir.  $L_{\langle n, f, k \rangle: F}$  ( $L_{\langle n, f, k \rangle: G}$ ) ile doğrusal iki ortak bozulma kriterli  $\langle n, f, k \rangle: F(G)$  sistemini ve  $C_{\langle n, f, k \rangle: F}$  ( $C_{\langle n, f, k \rangle: G}$ ) ile dögüsel iki ortak bozulma kriterli  $\langle n, f, k \rangle: F(G)$  sistemini sembolize etmek mümkündür.

#### 2. Tekrar İstatistikleri ile Profust Güvenilirlik

Klasik güvenilirlik analizinde sistem "çalışma" ve "çalışmama" olmak üzere iki durumda bulunmakta ve sistemin davranışı olasılık teorisi yardımıyla tam olarak karakterize edilmektedir. Bulanık güvenilirlik sistemlerinden biri olan Profust güvenilirlikte ise sisteme ilişkin olasılık varsayımı korunurken sistemin iki durumlu olması varsayımı bulanık durum varsayımı yoluyla gerçeğe daha uygun modelleme yapmaya çalışılmaktadır. Her iki güvenilirlik kavramında da ulaşılmak istenilen sistemin çalışır durumdan çalışmama duruma geçişinin gerçekleşmemesi olasılığıdır. Profust güvenilirlik aşağıda verilen eşitlik ile elde edilmektedir:

$$R(t_0, t_0 + t) = 1 - \sum_{j=1}^{n-1} \mu_{R_j} (m_{ij}) P\{t_0 + t \text{ anında, sistem } S_j \text{ durumundadır}\} \quad (1)$$

(1) nolu ifade de  $\mu_{R_j} (m_{ij})$ ; sistemin  $S_n$  durumundan  $S_j$  durumuna geçişi olayının, yani sistemin ne ölçüde "çalışır" durumdan "çalışmama" durumuna geçişi olarak ifade edilebileceğini gösteren üyelik fonksiyonudur. Bu çalışmada F sistemleri ve G sistemleri için iki ayrı üyelik fonksiyonu; sistemin durumunun daha hızlı bir şekilde bozulma durumuna gitmeye başladığı bileşen sayısına (l) göre tanımlanmıştır. Tekrar istatistikleri kullanılarak sistemlerin klasik güvenilirliklerini hesaplamak mümkün olduğu gibi Profust güvenilirlikleri de hesaplanabilir. Tekrar istatistikleri ile ilgili geniş bir içerik Balakrishnan ve Koutras (2002) da yer almaktadır. Çalışmada ilgililenen  $\langle n, f, k \rangle: F(G)$  sistemlerinin profust güvenilirliklerini doğrusal ve



döngüsel yapı ayırımı yapılmaksızın tekrar istatistikleri kullanılarak (2) ve (3) eşitlikleri ile hesaplamak mümkün olmaktadır:

$$\hat{R}_{(n,f,k;F)}^0 = 1 - \sum_{j=1}^{n-1} \mu_{p_{ij}} (m_{nj}) P\{L_n^{(0)} \geq k \wedge S_n^{(0)} \geq f\} \quad (2)$$

$$\hat{R}_{(n,f,k;G)}^0 = 1 - \sum_{j=1}^{n-1} \mu_{p_{ij}} (m_{nj}) P\{L_n^{(1)} < k \vee S_n^{(1)} < f\} \quad (3)$$

Bu ifadelerde ki  $L_n^{(1)}$  ve  $L_n^{(0)}$  sırasıyla bir dizideki (doğrusal veya döngüsel) en uzun başarı ve başarısızlık tekrarlarını  $S_n^{(1)}$  ve  $S_n^{(0)}$  ise sırasıyla bir dizideki (doğrusal veya döngüsel) toplam başarı ve başarısızlık sayılarını göstermektedir.

### 3. Sayısal Örnek

Aşağıdaki tabloda belirli parametre değerlerine göre çalışmada ilgilenilen sistemlerin (doğrusal ve döngüsel  $\langle n, f, k \rangle: F(G)$ ) profust güvenilirlik değerleri verilmektedir.

**Çizelge 1.** Belirli parametre (n, p, f, k, l, λ) değerlerine göre sistemlerin güvenilirliği

n	p	f	k	l	λ	$\tilde{R}_{\langle n, f, k \rangle; F}^L$	$\tilde{R}_{\langle n, f, k \rangle; F}^C$	$\tilde{R}_{\langle n, f, k \rangle; G}^L$	$\tilde{R}_{\langle n, f, k \rangle; G}^C$
5	0.8	3	2	3	0.7	0.9620	0.9584	0.9386	0.9530
10	0.3	6	3	4	0.5	0.3679	0.3486	0.3783	0.3795

#### KAYNAKLAR

[1] D.T. Chiang and S. Niu (1981), Reliability of consecutive-k-out-of-n:F system, IEEE Trans. Reliability, 30, 87-89.

[2] N. Balakrishnan and M.V. Koutras (2002), Runs and Scans with Applications, New York:Wiley Series in Probability and Statistics.

[3] W.Kuo and M.J.Zuo (2003), Optimal Reliability Modeling, New York: John Wiley&Sons.

#### ABSTRACT

##### PROFUST RELIABILITIES OF SYSTEMS WITH TWO COMMON FAILURE CRITERIA

In this study, by employing run statistics, we investigated profust reliabilities of linear and circular, F and G systems with two common failure criteria which are connected with logical operator "AND".

**Key Words:** Profust Reliability, Linear  $\langle n, f, k \rangle: F(G)$  System, Circular  $\langle n, f, k \rangle: F(G)$  System

## BİLDİRİ OTURUMLARI 4

### SESSION 4

#### Biyostatistik

## DEVELOPING AN ITEM BANK FOR PROGRESS TEST AND APPLICATION OF A COMPUTERIZED ADAPTIVE TESTING BY SIMULATION IN MEDICAL EDUCATION

Ayşen Melek Ayтуğ Koşan<sup>1</sup>, Nizamettin Koç<sup>2</sup>, Atilla Halil Elhan<sup>3</sup>, Derya Öztuna<sup>3</sup>

### Introduction

Progress test is a form of assessment that measures the ability levels of all the students in a certain program and their progress in time for that educational program at the same time by providing them with the same questions and repeating the process at regular intervals with parallel tests. [1]. Abiding by an indicator chart geared at the cognitive learning objectives anticipated at the end- objectives of curriculum, a question sample which is representative of all the disciplines and content area is used in PT. Due to its contributions to the education, PT is used in medical education worldwide in a widespread manner, yet it has limitations like requiring a lot of effort, human resources and time for its preparation, implementation and evaluation and requiring several questions leading to a fatigue in the students [2]. Implementing PT with Computerized Adaptive Testing (CAT) method that would allow for the evaluate medical knowledge without resulting in boredom and by allowing for savings from time and human resources by using limited number of items will contribute significantly to the elimination of above mentioned limitations.

### Objective

The aim of this study was to develop an item bank for the Progress Tests used in Ankara University School of Medicine (AUSM) and to explore the potential of Computerized Adaptive Testing (CAT) for evaluating the medical knowledge in AUSM students.

### Method

The study group consisted of Year 1-5 students for the academic year 2010-2011. The Progress Test consisted of 200 items and 1206 students participated in the study (89.7% of the total). In the examination of the psychometric characteristics of the item bank, Rasch model for dichotomous items responses was performed. By using the parameters obtained from the Rasch analysis, data was generated to be used in simulation applications.

After testing the assumptions of Rasch model, the final item bank was also examined in terms of content validity. Then, using this item bank, several simulated CAT applications [stopping rules of different standard errors/reliability values] were carried out and their results were checked against simulated applications from 1.000 simulees. In addition, the mean number of items used in each CAT condition was evaluated for student burden.

### Findings:

After Rasch Analysis, an item bank consisting of 103 items and including a unidimension was obtained. Person Separation Index-(PSI) and KR-20 of the item bank was calculated as 0.77. The findings of the research have indicated that the correlation for all simulation conditions ([N(0,1)] and [N(0,3)] for standard values 0.3, 0.4, 0.5, 0.548) between  $\theta$  estimations obtained from paper-and-pencil ( $\theta$ Rasch) and CAT applications ( $\theta$ Rasch) were high (ICC 0.928-0.999).

The average mean of items used in CAT, which was carried out with derived responses for 1000 individuals from the distribution of mean with 0 and variance with 1, varies between 11 and 45 for various levels of standard error. In CAT applications estimation can be made with 14 questions. When compared to paper and pencil tests and CAT resulted in a 56.3-86.4% decrease in the number of questions.

<sup>1</sup>. Ankara Üniversitesi Tıp Fakültesi Tıp Eğitimi ve Bilişimi Anabilim Dalı, Cebeci hastanesi Kampüsü, Dikimevi, Ankara  
E-mail adresi: [aysenayl@yahoo.com](mailto:aysenayl@yahoo.com)

<sup>2</sup>Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme Değerlendirme Anabilim Dalı

<sup>3</sup>Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı

The average mean of items used in CAT, which was carried out with derived responses for 1000 individuals from the distribution of mean with 0 and variance with 3, varies between 12 and 75 for various levels of standard error. When compared to paper and pencil tests CAT resulted in a 27.6-88.3% decrease in the number of questions.

**Result:** This study showed that it is possible to develop a proper item bank for the Progress Test using Rasch Model and burden of this assessment of medical knowledge with respect to the number of items administered can be reduced by the application of a CAT procedure.

### REFERENCES

[1] Freeman A., van der Vleuten, Nouns, Z., Ricketts C., Welbourn, R.B. (2010). Progress Testing Internationally. Medical Teacher. 32:451- 455.

[2] Wrigley,W., van der Vleuten., C.P.M., Freeman, A., Muijtjens, A. (2012). A Systemic Framework for the Progress Test: Strengths, Constraints and Issues: AMEE Guide No. 71. Medical Teacher. 34: 683-697

### TIP EĞİTİMİNDE GELİŞİM SINAVI SORU BANKASI OLUŞTURULMASI VE BENZETİM VERİLERİ İLE BİLGİSAYAR UYARLAMALI TEST UYGULAMASI

Bu çalışmada, Ankara Üniversitesi Tıp Fakültesinde yürütülmekte olan gelişim sınavı için bir soru bankası oluşturulması ve bu soru bankasını kullanılarak Bilgisayar Uyarlamalı Test yöntemi ile Gelişim Sınavının uygulanabilirliğinin ortaya konulması amaçlanmıştır. 1206 AUTF öğrencisi 200 sorudan oluşan ve tıp bilgisi değerlendirilen 200 soruya cevap vermişlerdir. Soru bankasının oluşturulmasında iki kategorili yanıtlar için Rasch modeli kullanılmış, 103 maddeden oluşan tek bir boyut içeren soru bankası elde edilmiştir.

Yapılan benzetim çalışmaları sonucunda, Gelişim Sınavı kağıt kalem testi uygulamaları ve BUT uygulamasının ile elde edilen yetenek kestirimleri arasındaki korelasyonun yüksek olduğu gösterilmiştir. BUT'da kullanılan madde sayısı ortalamasının standart hatanın farklı düzeyleri için 11 ile 45 arasında değiştiği ve, kağıt kalem uygulamalarına göre madde sayısında %56.3-86.4 arasında azalma sağlandığı gösterilmiştir.

**Key Words:** Progress test, Computerized adaptive test, Medical Education, Rasch Models

**SOL ATRİYUM ÇAPININ TEMEL BİLEŞENLER REGRESYONU, KISMİ EN KÜÇÜK KARELER  
REGRESYONU ve YAPAY SINIR AĞLARI İLE TAHMİNLERİNİN KARŞILAŞTIRILMASI**

Fatma KIZILKAYA, Cemil ÇOLAK, Oktay KIZILKAYA

Öğr. Gör. Fatma KIZILKAYA\*, Hakkâri Üniversitesi, Yüksekova Melek Yüksek Okulu, 30000, Hakkâri,  
TÜRKİYE, askin.fatma@hotmail.com

Doç. Dr. Cemil ÇOLAK, İnönü Üniversitesi, Tıp Fakültesi, Biyoistatistik Bölümü, 44100, Malatya, TÜRKİYE,  
cemilcolak@yahoo.com

Arş. Gör. Oktay KIZILKAYA, Hakkâri Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, 30000, Hakkâri,  
TÜRKİYE, o.kizilkaya.4@hotmail.com

**Özet**

**Amaç:** Bu çalışmada sol atriyum çapının, Temel Bileşenler Regresyonu (TBR), Kısmi En Küçük Kareler regresyonu (KEKR) ve Yapay Sinir Ağları (YSA) ile tahmininin yapılması, elde edilen tahminlerin karşılaştırılması amaçlanmaktadır.

**Gereç ve Yöntem:** Fırat Üniversitesi Tıp Fakültesi Kardiyoloji Polikliniğine gelen 127 hipertansif hastanın ekokardiyografi raporları prospektif olarak toplanmıştır. Sol atriyum çapını etkilediği düşünülen değişkenler ele alınmış, bu değişkenler arasında çoklu bağlantı probleminin var olduğu tespit edilmiştir. Bu problemin giderilmesi için TBR, KEKR ile sol atriyum çapının tahmini yapılmıştır. Ayrıca, YSA'da değişik YSA modelleri ve ileri beslemeli geri yayımlı ağ algoritmasının gelişmiş bir türü olan Levenberg-Marquardt eğitim algoritmasından yararlanarak en iyi sonuçlar elde edilmeye çalışılmıştır.

**Bulgular ve Sonuçlar:** Çalışmada Tahminlerin birbirine üstünlüklerini belirlemede açıklayıcılık katsayısı  $R^2$  ve RMSE değerleri kullanılmıştır.

**Tablo 1.** Tahminlerin Karşılaştırılması

Uyum Kriteri	TBR	KEKR	YSA
RMSE	2.825	2.640	2.482
$R^2$	0.30	0.33	0.41

Tahminler değerlendirildiğinde; YSA ile yapılan tahminin TBR ve KEKR'ye göre RMSE değerinin daha düşük ve açıklayıcılık katsayısının daha yüksek olması nedeniyle YSA'nın sol atriyum çapının tahmininde daha iyi sonuç verdiği belirlenmiştir.

**KAYNAKLAR**

- [1]. **Abudu, S.**, 2010. Modeling of Daily Pan Evapoation Using Partial Least Squares Regression.
- [2]. **Albayrak A. S.**, 2005. Çoklu Doğrusal Bağlantı halinde EKK Tekniğinin Alternatifli Yanlı Tahmin Teknikleri ve Bir Uygulama. ZKÜ Sosyal Bilimler Dergisi,1,1.
- [3]. **Collins, B.**, 2010. Partial Least Squares Regression.
- [4]. **Norliza Binti A.**, 2006. Comparing Three Methods of Handling Multicollinearity Using Simulation Approach, Universiti Teknologi Malaysia.
- [5]. **Öztemel, E.**, 2003. Yapay Sinir Ağları, İstanbul: Papatya Yayıncılık.

**COMPARISON OF THE LEFT ATRIUM DIAMETER ESTIMATIONS WITH PRINCIPAL  
COMPONENTS REGRESSION, PARTIAL LEAST SQUARES REGRESSION AND ARTIFICIAL  
NEURAL NETWORKS METHODS**

In this study, it was aimed to estimate of diameter of the left atrium with Principal Component Regression, Partial Least Square Regression and Artifical Neural Networks. Multicollinearity problem that occur in case of failure of assumption of independence between the explanatory variables was examined. Principal components regression and partial least square regression that used to overcome this problem were described. Finally, artificial neural network was examined. In the part of application, echocardiography reports of 127 hypertensive patients who came to Cardiology Polyclinic of Medicine Faculty of Fırat University were collected prospectively. The obtained data were analyzed by all of methods described above and the results were compared.

**Key Words:** The Multicollinearity Problem, Principal Components Regression, Partial Least Squares Regression, Artificial Neural Networks.

## MOLEKÜLER VARYANS ANALİZİ (AMOVA) TEKNİĞİ

Nuri ÇELİK\*

\*Bartın Üniversitesi Fen Fakültesi İstatistik Bölümü Merkez/Bartın

ncelik@bartin.edu.tr

Gül OLGUN KARACAN

Ankara Üniversitesi Fen FakültesiBiyoloji Bölümü Tandoğan/Ankara

golgun@gmail.com

### 1. Moleküler Varyans Analizi Tekniği

Moleküler Varyans Analizi (Analysis of Molecular Variance, AMOVA) türler arası moleküler farklılığı araştırmada kullanılan bir tekniktir. Ayrıca AMOVA, farklı genotiplerden elde edilen belirleyicilerle popülasyon yapısı hakkında bilgi vermektedir. DNA yapısını özetlemekte kullanılan en önemli metot olan electrophoresisler bir bantta istenen genotipin olması durumunda 1 olmaması durumunda 0'lardan oluşan bir veri seti elde etmektedir. Söz konusu vektöre DNA haplotipi adı verilmektedir. Elde edilen vektörleri kullanarak AMOVA toplam farklılığın gruplar içi ve gruplararası şeklinde bileşenlerine ayrılması ile hesaplanmaktadır.

AMOVA, moleküler datadan direk olarak popülasyon ayrılıklarını tahmin eden ve bu farklılığı test eden bir tekniktir. Buna göre  $p_j$  1 ve 0'lardan oluşan Boolean vektörü olmak üzere, iki haplotip arasındaki uzaklık

$p_j - p_k$  şeklinde hesaplanmaktadır. Hesaplanan Euclidean uzaklığının karesi ise

$$\delta_{jk}^2 = (p_j - p_k)'W(p_j - p_k) \quad (1)$$

şeklinde hesaplanmaktadır. Burada W genellikle birim matris şeklinde tanımlanmaktadır.

### 2. Yöntem

Yöntem (1) eşitliği ile verilen uzaklık karesini kullanarak ayrıış kareler toplamını (sum of squares of deviation,  $SSD_{Total}$ ) hesaplamaktadır. Buna göre N birimlik her bir vektör uzunluklarından toplam 2N tane hesaplanacağından,  $SSD_{Total}$ ,

$$SSD_{Total} = \frac{1}{2N} \sum_{j=1}^N \sum_{k=1}^N (p_j - p_k)'W(p_j - p_k) \quad (2)$$

şeklinde tanımlanmaktadır. Buna göre AMOVA modeli için tanımlanan lineer model,

$$p_{jig} = p + a_g + b_{ig} + c_{jig}, \quad j = 1, 2, \dots, N_{ig}; i = 1, 2, \dots, I_g; g = 1, 2, \dots, G \quad (3)$$

şeklinde (Cockerham, 1969, 1973, Weir and Cockerham 1984, Long 1986). Burada, a grupların etkisini, b popülasyon etkisini, c ise popülasyon içindeki birey etkisini göstermektedir. p her lineer modelde tanımlanan genel ortalama ve g grup sayısını, i popülasyon sayısını, j ise i. popülasyon içindeki birey sayısını göstermektedir. Buna göre AMOVA tablosu,

Çizelge 1: AMOVA Tablosu

Değişim Kaynağı	Serbestlik Derecesi	Kareler Ortalaması
Bölgeler Arası	$G - 1$	$MSD / (AG)$
Her Bir Bölge İçindeki Popülasyonlar Arası	$\sum_{g=1}^G I_g - G$	$MSD / (AP / WG)$
Her Bir Popülasyon İçindeki Bireyler Arası	$N - \sum_{g=1}^G I_g$	$MSD / (WP)$
Toplam	$N - 1$	

AMOVA tablosu için kullanılan test istatistiği ise  $\Phi$  test istatistiği varyans bileşenleri kullanılarak hesaplanmaktadır.

Bununla beraber, uygulamada Türkiye içinde kayalık faresi apodemus mystacinus olarak bilinen bir fare türünün ayrılışları incelenmiştir. Gerçek veri olarak popülasyonların mtDNA sitokrom b bölgelerinin dizileri kullanılmıştır.

### KAYNAKLAR

[1] Cockerham, C.C. (1973), Analysis of gene frequencies, Genetics, 74, 679-700

[2] Cockerham, C.C. (1969), Variance of gene frequencies, Evolution, 23, 72-84

[3] Excoffer, L., Smouse, P. and Quattro, J. (1992), Analysis of molecular variance inferred for metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data, Genetics, 131, 479-491,

[4] Wright, S., (1951), The genetical structure of populations, Ann. Eugen., 15, 323-354.

[5] Olgun Karacan G (2013), Türkiye'deki Apodemus mystacinus (Mammalia: Rodentia)'un mtDNA (Sitokrom b ve Kontrol Bölgesi) RFLP ve DNA Dizi Analizi, Ankara.

### ABSTRACT

Analysis of Molecular Variance (AMOVA) is a method for studying molecular variation within species. Also, It is used to construct the population structure with the data from genotypes. The data consists of 1s and 0s, while 1s represent the presence of a marker and 0s represents the absence of the marker in data. So, AMOVA uses this data to calculate the sum of squares of total deviation and its components. In this work we use the series of rock mouse's (apodemus mystacinus) mtDNA.in Turkey.

**Key words:** AMOVA, DNA, Biostatistics,

### DSRCT: REKTUM KANSERİ TEDAVİSİNDE KARAR DESTEĞİ

Ashı SUNER<sup>1\*</sup>, Oğuz DİCLE<sup>2,3</sup>, Selman SÖKMEN<sup>4</sup>, Gökhan KARAKÜLAH<sup>5</sup>,  
Can Cengiz ÇELİKOĞLU<sup>1</sup>

<sup>1</sup>Dokuz Eylül Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 35160, İzmir, TÜRKİYE, asli.suner@deu.edu.tr,  
cengiz.celikoglu@deu.edu.tr

<sup>2</sup>Dokuz Eylül Üniversitesi, Sağlık Bilimleri Enstitüsü, Medikal İnfomatik Anabilim Dalı, 35340 İzmir,  
TÜRKİYE

<sup>3</sup>Dokuz Eylül Üniversitesi, Tıp Fakültesi, Radyoloji Anabilim Dalı, 35340, İzmir, TÜRKİYE,  
oguz.dicle@deu.edu.tr

<sup>4</sup>Dokuz Eylül Üniversitesi, Tıp Fakültesi, Genel Cerrahi Anabilim Dalı, 35340, İzmir, TÜRKİYE,  
selman.sokmen@deu.edu.tr

<sup>5</sup>Dokuz Eylül Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyomekanik Anabilim Dalı, 35340, İzmir, TÜRKİYE,  
gokhan.karakulah@deu.edu.tr

#### [1] Çalışmanın Amacı

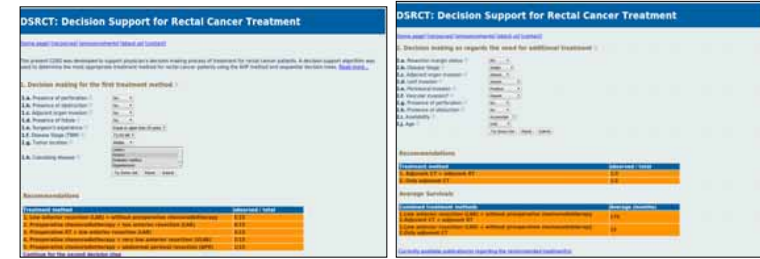
Rektum kanseri, tüm dünyada kadınlarda ikinci, erkeklerde ise üçüncü sırada en yaygın olarak görülen kanser türüdür. Rektum kanseri tedavisine ilişkin karar verme sürecinde, hastalığa ait çok sayıda kriter rol oynamakta ve farklı pek çok tedavi seçeneği bulunmaktadır [1, 2]. Bu çalışmada, rektum kanseri hastalarına ilişkin en uygun tedavi yönteminin belirlenmesine yardımcı ve potansiyel kullanıcılara, rektum kanserine ilişkin tedavi planlamasında karar desteği sağlamaya yönelik, gerçek hasta verilerine dayalı ve bilgisayar tabanlı bir aracın geliştirilmesi amaçlanmıştır.

#### [2] Yöntem ve Gereçler

Yöneylem araştırmasının çok kriterli karar verme tekniklerinden biri olan analitik hiyerarşi süreci (AHP) yöntemi ve karar ağacı yönteminin bir arada kullanılmasıyla bir karar modeli geliştirilmiştir [3]. Hasta için tedavi kararı verilirken AHP yöntemi ile elde edilen öncelikler kullanılarak en yüksek önceliğe sahip kriterden en düşük önceliğe sahip kriterlere doğru değerlendirme yapılmıştır. Uygulanacak tedavinin belirlenmesinde kriter öncelikleri kullanılarak karar ağacı oluşturulmuştur. Karar ağacı yardımıyla oluşturulan model, gerçek hasta verilerini, uzman görüşlerini ve literatür desteğini kullanan bir web ara yüzü haline dönüştürülmüştür. Hasta verileri ve uzman görüşleri Dokuz Eylül Üniversitesi, Tıp Fakültesi, Genel Cerrahi Anabilim Dalı'ndan sağlanırken; literatür desteği MeSH terimleri yardımıyla PubMed veritabanından elde edilmiştir. 1988–2010 yılları arasında verilmiş tedavi kararlarının yer aldığı 388 hasta verisi, Ubuntu Linux server sürüm 12.04.1 üzerine Mysql sürüm 5.5.31 kurularak oluşturulan hasta kayıt veri tabanına eklenmiştir. Potansiyel kullanıcıların bu veri tabanına erişimini sağlamak için HTML5 standartlarına uygun ve PHP sürüm 5.3.10 ile desteklenmiş kullanıcı ara yüzleri tasarlanmıştır. Oluşturulan karar desteği ile karar vericiye belirli kriter düzeylerindeki hastalar için uygulanabilecek tedavi yöntemleri önerilmekte, bu tedavi yöntemlerine ilişkin ortalama sağkalım süreleri belirtilmektedir. Ayrıca önerilen tedavi yöntemlerine ilişkin literatür bilgisi de kullanıcıya sunulmaktadır. Böylece belirli bir hasta için birden çok tedavi yöntemi önerildiğinde, karar verici ortalama sağkalım sürelerini ve literatür desteğini göz önünde bulundurarak en uygun tedavi kararını verebilmektedir.

#### [3] Karar Süreci

Uzmanlar karar verme sürecinin iki ardışık karar basamağından oluştuğunu belirtmişlerdir. İlk basamak olan uygulanacak tedavi tipi belirlendikten sonra bu basamak esas alınarak, ikinci basamakta eğer gerekiyorsa ek tedavi kararı verilmektedir. Kriter önceliklerinin belirlenmesi için uygulanan AHP analizi sonucunda her iki karar basamağı için matrisler tutarlı bulunmuştur ( $CR_1=0,03<0,1$  ve  $CR_2=0,01<0,1$ ). İlk karar basamağı için en önemli kriter 0.331 öncelik değeriyle “perforasyon varlığı” olurken, ikinci karar basamağı için “rezeksiyon marjın durumu” en yüksek öncelik değeriyle (0.214) sahip kriterdir. Analitik hiyerarşi süreci ve karar ağaçları yöntemlerinin birleştirilmesi ile oluşturulan karar modelinin gerçek hasta verileri ile yapılan uygulaması olan DSRCT karar destek aracının ilk ve ikinci basamaklarına ilişkin ekran görüntüleri Şekil 1. 'de görülmektedir.



Şekil 1. DSRCT karar destek aracının karar basamaklarına ilişkin ekran görüntüleri.

#### [4] Sonuçlar ve Tartışma

Bu çalışmada, uzman görüşü yanında, hasta verilerine ve hastalığa ilişkin kriterlere dayalı olarak karar desteği sağlayabilen bir karar destek sistemi geliştirilmiştir. Sistemin tutarlılığı uzmanlar tarafından değerlendirilmiş ve her iki basamağın tutarlılığına ilişkin sonuçlar, model için yeterli ve kabul edilebilir sınırlar içerisinde bulunmuştur. Bu karar destek sisteminin rektum kanserinin tedavi planlamasında potansiyel kullanıcılara yardımcı olacağı ve uzmanlık öğrencilerinin eğitim sürecine destek sağlayacağı düşünülmektedir.

#### KAYNAKLAR

- [1] Ferlay J., Shin H. R., Bray F., Forman D., Mathers C. and Parkin D. M. (2008). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. International Journal of Cancer, 127, 2893-917.
- [2] Stocchi L., Nelson H., Sargent D. J., O'Connell M. J., Tepper J. E., Krook J. E., Beart R. Jr. and North Central Cancer Treatment Group. (2001). Impact of surgical and pathologic variables in rectal cancer: a United States community and cooperative group report. Journal of Clinical Oncology, 19, 3895-902.
- [3] Suner A., Çelikoğlu C. C., Dicle O. and Sökmen S. (2012). Sequential decision tree using the analytic hierarchy process for decision support in rectal cancer. Artificial Intelligence In Medicine 56(1): 59-68.

#### ABSTRACT

##### DSRCT: DECISION SUPPORT FOR RECTAL CANCER TREATMENT

In the treatment of the rectal cancer which is the second most common type of cancer in women and the third in men worldwide; there are various criteria and treatment options that effect the decision of which treatment should be applied. It was aimed to create a system to support the physician's decisions in order to determine the most appropriate treatment method. The decision model that built by combining analytic hierarchy process and decision tree was presented to the user by using a web-based application. This study supports the decision making process using patient data as well as expert knowledge.

**Key Words:** Analytic hierarchy process (AHP), decision trees, rectal cancer treatment, clinical decision support systems (CDSS), decision making, expert opinion.

**BAĞIMLI DURDURMA VARSAYIMI ALTINDA YİNELEMELİ OLAYLARIN HOMOJEN  
POISSON SÜRECİ İLE MODELLENMESİ**

Hande KONŞUK\*, Ayten YİĞİTER  
Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800 Beytepe, Ankara, TÜRKİYE

hkonsuk@hacettepe.edu.tr, yigiter@hacettepe.edu.tr

**Giriş**

Son yıllarda sağkalm analizinde daha fazla bilgi verebilecek çoklu olay verileri sıklıkla çalışılan bir konudur. Çoklu olay verisi iki kategoriye ayrılır. Bunlar yinelemeli olaylar ve yarışan risklerdir. Yinelemeli olaylarda, birey ya da birim başarısızlık olayını birden fazla sayıda deneyimlerken, yarışan risklerde ise başarısızlığa neden olan farklı tipteki olaylardan yalnız bir tanesini deneyimler.

Yinelemeli olay verileri çoğu kez uzun dönem izlenen çalışmalarla toplanmıştır ve çalışmadaki başarısızlık zamanı, bireyin ölümü, çalışmanın sona ermesi ya da izlenme döneminde bireyin çalışmadan çekilmesi sebepleriyle durdurmaya maruz kalabilir. Eğer durdurma, çalışmanın sona ermesinden kaynaklıyorsa, durdurma zamanı bağımsız ya da bilgi içermeyen şekilde kabul edilir. Birçok uygulamada yinelemeli olaylar, bireyin çalışmadan çekilmesi ya da bireyin ölümü ile sonlandırılır. Bu şekilde meydana gelen durdurma zamanı bilgi içerir ve yinelemeli olay zamanları ile ilişkilidir. Bu ilişkiyi göz ardı ederek modelleme yapmak yanlış sonuçlar elde edilmesine neden olmaktadır.

Yinelemeli olayların analizi için bağımlı ve bağımsız durdurma varsayımı altında birçok model önerilmiştir. Bağımsız durdurma varsayımı altında önerilen modellerden en önemli üç tanesi şunlardır: Prentice ve arkadaşları (1981), yinelemeli olayları tabakalara ayırarak olaylar arasında geçen zamanı, yoğunluk fonksiyonları kullanarak modellemişlerdir. Wei ve arkadaşları (1989), her bir yinelemeye dek geçen zamanı marjinal modelleri kullanarak modellemişlerdir. Andersen ve Gill (1982), yinelemeli olayları bağımsız artış varsayımı altında yoğunluk/oran fonksiyonları kullanarak modellemişlerdir. Bağımlı durdurma varsayımı altında yapılan bazı çalışmalar şunlardır: Li ve Lagakos, durdurma değişkeninin ölüm olayı olduğunu düşünerek Wei, Lin ve Weissfeld'in önermiş olduğu yöntemi bağımlı durdurma için uyarlamışlardır. Cook ve Lawless, yinelemeli olaylar ve durdurma olayı arasındaki bağımlılık yapısını bileşik modeller kullanarak incelemişlerdir.

Literatürde bağımlı ve bağımsız durdurma varsayımı altında yinelemeli olaylar, sayma süreçleri kullanılarak da modellenmektedir. Bunlardan en yaygın kullanılanı homojen Poisson sürecidir.

Liu, Wolfe ve Huang, bağımlı durdurma içeren yinelemeli olayları homojen Poisson süreci kullanarak modellemiş; yinelemeli olaylar ile durdurma olayı arasındaki bağıntıyı kurmak için iki model önermişlerdir. Durdurma olayına ilişkin yoğunluk fonksiyonunu,

$$P(dN^D(t) = 1 | F_r) = Y_i(t) d\Lambda_i(t) \equiv Y_i(t) \lambda_i(t) \quad (1)$$

ve durdurma olayının varlığı altında yinelemeli olay sürecine ilişkin yoğunluk fonksiyonunu,

$$P(dN^R(t) = 1 | F_r, D \geq t) = Y_i(t) dR_i(t) \equiv Y_i(t) r_i(t) dt \quad (2)$$

biçiminde tanımlamışlardır. Liu ve arkadaşları, yinelemeli olay süreci ile bağımlı durdurma sürecini modellerken bilgi içermeyen durdurma sürecinin bilgisini modele katmamışlardır.

Bu çalışmada, bağımlı durdurma varsayımı altında yinelemeli olay süreci, bileşik modeller kullanılarak modellenmiştir. Bağımlı durdurma olarak ölüm olayı dikkate alınmıştır. Yinelemeli olay ve bağımlı durdurma süreçleri, Poisson süreçleri kullanılarak modellenmiştir. Liu ve arkadaşlarının modellerine ek olarak bağımsız durdurma bilgisi de modele katılmıştır. Parametreler, en çok olasılık tahmin yöntemi kullanılarak elde edilmiştir. Farklı senaryolar altında yapılan simülasyon çalışmasının sonuçları verilmiştir.

**KAYNAKLAR**

[1] Prentice P.L., Williams B.J. and Peterson A.V. (1981), On the regression analysis of multivariate failure time data. *Biometrika*, 68(2): p. 373-379.

[2] Wei L.J., Lin D.Y., and Weissfeld L. (1989), Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of American Statistical Associations*, 84(408): p. 1065-1073.

[3] Andersen P.K. and Gill R.D. (1982), Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, 10(4): p. 1100-1120.

[4] Liu L., Wolfe R.A. and Huang X. (2004), Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60(3): p. 747-756.

[5] Zeng D. and Lin D.Y. (2009), Semiparametric Transformation Models with Random Effects, *Biometrics*, 65, 746-752.

**ABSTRACT**

**MODELLING RECURRENT EVENT DATA WITH HOMOGENEOUS POISSON PROCESS UNDER DEPENDENT CENSORING**

There has been much work done on the analysis of recurrent event data. There are three major methods which are commonly used in the literature. These methods are gap time models proposed by Prentice, Williams and Peterson (1981), marginal hazard models proposed by Wei, Lin and Weissfeld (1989) and intensity/rate functions proposed by Andersen and Gill (1982). In these studies, censoring is assumed independent from recurrent event process. However, in many studies, recurrent event data with a terminal event such as death are often encountered. In such cases independent censoring assumption is violated therefore, recurrent event data is modeled in the presence of dependent censoring otherwise biased estimates are obtained. In this study, we modeled recurrent event data with homogeneous Poisson process under dependent censoring assumption. We defined three intensity functions for death process, recurrent event process given death and recurrent event process under independent censoring. The parameters are estimated using maximum likelihood method. Simulation studies have been conducted for different scenarios.

**Key Words:** Recurrent event data, Homogeneous Poisson process, dependent censoring

## SAĞLIK YÖNETİMİ ENFORMASYON SİSTEMLERİNİN ERKEN TEŞHİS VE TANI İÇİN STOKASTİK MODELLER İLE GELİŞTİRİLMESİ

Adem DOĞANER, Sinan ÇALIK  
Fırat Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 23119, Elazığ/TÜRKİYE  
adoganer@firat.edu.tr scalik@firat.edu.tr

### Sağlık Yönetimi Enformasyon Sistemleri

Bilişim sistemleri, gelişen bilgisayar teknolojileri vasıtasıyla hizmet sektörlerinde büyük bir ihtiyacı karşılayarak ve birçok işlem konusunda zaman kazanımı ve pratiklik sağlamaktadır. Bilişim sistemleri, çok karmaşık bir yapıya sahip olan sağlık sistemlerinde de önemli çözümler ve pratik uygulamalar kazandırmaktadır. Sağlık bilişim sistemleri temelde uygulama farklılıkları bakımından klinik bilgi sistemleri ve teşhis tedavi sistemleri olarak iki grupta incelenebilmektedir. Sağlık bilişim sistemleri, sağlık alanında elde edilen veriler üzerinde çalışarak bu alanların geliştirilmesi için pratik uygulamalar üzerine bilgisayar teknolojilerinden yararlanarak çözümler üretmektedir. Klinik bilgi sistemleri, hastane yönetimi, hasta kayıt ve takibi, tıbbi görüntüleme ve laboratuvar analiz sonuçları kayıtları, tedaviye yönelik karar destek gibi alanlarda bilgisayar destekli sistemler geliştiren bilişim sistemleri olarak ifade edilebilmektedir. Teşhis ve tedavi sistemleri, tedavi ve teşhise destek sağlayan bilgisayar destekli sistemlerden oluşmaktadır. Sağlık bilgi sistemleri, sağlık alanındaki yapılan hareketlerin ortak bir havuzda birleşerek belirli bir veritabanı oluşturulmasına olanak sağlamıştır. Veritabanından elde edilen bilgiler doğrultusunda gerek hastane yönetimi aşamasında, gerekse de tip çalışanlarının hasta teşhis tedavi konusunda destek aldığı önemli bir sistem haline almıştır.

Çalışmada sağlık enformasyon sistemlerinin geliştirilerek teşhis ve tanı ve ileriye yönelik meydana gelebilecek hastalıkların tahminlerinde başarılı sonuçlar sağlanması amaçlanmıştır. İleriye yönelik tahminlemelerde istatistiksel yöntemler sıklıkla uygulanmaktadır. Stokastik modeller, belirli olasılıklar ve koşullar altında olasılık teoremlerine dayanarak tahminler için kullanılabilirlerdir.

Mevcut enformasyon sistemlerinin pek çoğu belirli algoritmalar, veritabanları ve varlık ilişki diyagramları ile kayıtları tutma ve düzenli bilgiler oluşturma işlemini üstlenmektedir. Çalışmada enformasyon sistemlerinin stokastik modeller ile ilişkilendirilerek ileriye yönelik tahminler üretebilme yeteneğinin kazandırılması hedeflenmiştir. Bilgi sistemleri ile ilişkilendirilecek yöntem saklı Markov modelleridir.

### 2. Saklı Markov Modeli

Saklı Markov Modeli, Markov zincirindeki durumlar arası geçiş olasılıklarının yanı sıra durumlara bağlı olarak gelişen gözlem olasılıkları olmak üzere iki stokastik süreçten oluşmaktadır. Saklı Markov modelinde bilinen ve bilinmeyen parametreler mevcuttur. Durumlar arası geçişler Saklı Markov modelinde gizlidir. Fakat bu durumlara bağlı olarak gelişen gözlemler mevcuttur ve gözlem olasılıkları bilinmektedir. Gözlem olasılıkları aracılığı ile sistemin hangi durumda olduğu ve durumlar arası geçişler tahmin edilmek istenmektedir. Saklı Markov modelinde geçiş olasılıkları dağılımı  $A$ , her bir durum için gözlem sembolleri olasılıkları  $B$  ve başlangıç durum dağılımı  $\pi$  olmak üzere kesikli parametrelili saklı Markov Modeli;

$$\lambda = (A, B, \pi)$$

olarak ifade edilir.

Saklı Markov modeli sistemin elverişliliği bakımından olasılıksal incelemeler gerçekleştirebilir. Bu incelemeler, saklı Markov modelinin ilgilendiği üç temel problem üzerindedir ve her bir problem çözümü için geliştirilmiş algoritmalar mevcuttur.

Saklı Markov modeli, bilgisayar bilimleri başta olmak üzere birçok alanda yaygın bir uygulama sahası bulmuştur.

Çalışmada, farklı yaş ve gruptan gelen hastaların hastalık belirtileri gözlemler olarak bilgi sistemine dahil edilmiştir. Belirtiler sonucu oluşan hastalıklar ise durumlar olarak nitelendirilmiştir. Belirtilenden yola çıkarak hastanın hangi hastalığa yakalanacağı olasılıksal olarak tahmin edilmiştir. Düzenlenen saklı Markov modeli, enformasyon sistemi ile ilişkilendirilerek hastaların geçmişte geçirdikleri hastalıklar ve tıbbi değerler dikkate alınarak gelecekte hangi hastalıklarla karşı karşıya kalabilecekleri tahmin edilmiştir.

### KAYNAKLAR

- [1] Rabiner, L., (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceeding of IEEE*. 77(2):257-286.
- [2] Baum, L.E., (1972). An Inequality and Associated Maximization Technique in Statistical Estimation of Probabilistic Functions of a Markov Process. *Inequalities* 3: 1-8.
- [3] Dynkin, E.B., (1961), *Theory of Markov processes*. Pergamon Press, London.
- [4] Ömürbek, N. and Altın, F.G., (2009). Sağlık Bilişim Sistemlerinin Uygulanmasına İlişkin bir Araştırma, *Isparta, SDÜ Fen Edebiyat Fakültesi Sosyal Bilimler Dergisi*, 19:211-232.
- [5] Güleş, H.K. and Özata M., (2005). *Sağlık Bilişim Sistemleri*, 1. Basım, Ankara, Nobel Yayın.

### ABSTRACT

#### HEALTH MANAGEMENT INFORMATION SYSTEMS DEVELOPMENT WITH STOCHASTIC MODELS FOR EARLY DETECTION AND DIAGNOSIS

In this study, Health information systems for early detection and diagnosis were developed using Hidden Markov models. Symptoms and diseases were defined as observations and states for hidden Markov model, respectively.

**Key Words:** Information Systems, Symptoms, Diseases, Hidden Markov Model.

### TOZ METALURJİSİYLE ÜRETİLMİŞ KENDİNDEN YAĞLAMALI BRONZ YATAKLARDA YOĞUNLUĞA, GÖZENEKLİLİĞE VE AŞINMAYA ETKİ EDEN PARAMETRELERİN DENEY TASARIMI YÖNTEMİYLE ARAŞTIRILMASI

Alper SOFUOĞLU<sup>1</sup>, Yusuf ATAK<sup>1</sup>, İbrahim USLAN<sup>1</sup>

<sup>1</sup>Gazi Üniversitesi, Mühendislik Fakültesi, Makine Mühendisliği, 06570, Maltepe, Ankara, Türkiye,

masofuoglu@gazi.edu.tr

## 1. GİRİŞ

Toz metalurjisi ile üretilmiş kendinden yağlamalı yataklardaki gözenekler yağ deposu görevi yapmaktadırlar. Mil dönmeye başladıktan sonra sistemde sıcaklık yükselmeye başlar. Yatak ile mil arasındaki yağ filminin hidrodinamik basınçtaki değişikliğin etkisiyle yağ gözeneklerden mile doğru emilir. Dönme işlemi durduğunda, yatak soğuyarak kılcalık etkisiyle yağ gözenekler tarafından emilir. Toz metalurjisiyle üretilen yataklarda malzeme ve sıkıştırma basıncı, yoğunluk, gözeneklilik, aşınma miktarını etkileyebilmektedir.

## 2. LİTERATÜR

Cho ve ark. [1] grafit, Sb<sub>2</sub>S<sub>3</sub>, MoS<sub>2</sub> katı yağlayıcıların otomobil sürtünme performansına etkilerini araştırmışlardır. Sb<sub>2</sub>S<sub>3</sub> ve grafitin sürtünme katsayısını kararlı hale getirdiğini, Sb<sub>2</sub>S<sub>3</sub> ve MoS<sub>2</sub> ilavesinin de aşınma direncini azalttığını görmüşlerdir.

Tüfekçi ve ark. [2] bronz ve demir esaslı kendinden yağlamalı yatakların aşınma ve sürtünme davranışlarını kıyaslamışlardır. Bronz yatakların daha iyi sürtünme ve aşınma performansı gösterdiğini görmüşlerdir. Deneylerde yük, kayma hızı ve sıcaklığı değiştirmişlerdir.

Bu çalışmada, toz metalurjisi yöntemiyle üretilmiş kendinden yağlamalı bronz yataklarda grafit katkısının, sıkıştırma basıncının ve mesafe miktarının yoğunluk, gözeneklilik ve aşınma miktarı üzerindeki etkileri deney tasarımı yöntemiyle incelenmiştir. Bu çalışmada Atak [3]'ün çalışmasından yararlanılmıştır. Deney sonuçlarının bazı değerleri Tablo-1'de gösterilmiştir. Atak [3]'ün yaptığı çalışmada, ilk olarak, ortalama toz boyutu 60 µm olan ön alaşım CuSn10 bronz tozlarına ağırlıkça % 1-2-3 oranında grafit tozları ilave edilerek homojen karışımlar oluşturmak için Turbula tipi karıştırıcıda 20 dakika süre ile karıştırılmıştır. Hazırlanan toz karışımları 100 MPa ile 250 MPa arasında 50 MPa aralıklarla sıkıştırılmıştır. Sinterlenmiş yatak malzemelerinin yoğunluğu ve gözenekliliği Arşimet ilkesi uygulanarak hesaplanmıştır. Sinterlenmiş numunelerin gözenek dağılımı ışık mikroskobu ile incelenmiştir. Bu çalışmada gerçek yoğunluğa, gözenek miktarına ve aşınma miktarına etki eden faktörler tam faktöriyel deney tasarımı yöntemiyle araştırılmıştır. Paket program olarak MINITAB 16.0 kullanılmıştır

Tablo 1. Deney sonuçları

Ağırlıkça %1 grafit içeren ve 100-150-200-250 Mpa basınç altında sıkıştırılmış yatak numunelerinin ölçüm ve hesaplama sonuçları

Yatak No	Basınç (Mpa)	Grafit İçeriği (%)	Yoğunluk (g/cm <sup>3</sup> )	Gözeneklilik (%)	Aşınma (mm)
100-100-1	100	10	8.50	15	0.05
100-100-2	100	20	8.50	15	0.05
100-100-3	100	30	8.50	15	0.05
100-150-1	150	10	8.50	15	0.05
100-150-2	150	20	8.50	15	0.05
100-150-3	150	30	8.50	15	0.05
100-200-1	200	10	8.50	15	0.05
100-200-2	200	20	8.50	15	0.05
100-200-3	200	30	8.50	15	0.05
100-250-1	250	10	8.50	15	0.05
100-250-2	250	20	8.50	15	0.05
100-250-3	250	30	8.50	15	0.05

## BİLDİRİ OTURUMLARI 4

### SESSION 4

#### Modelleme ve Simülasyon



### 3. SONUÇLAR

1-Grafit miktarının ve basınç miktarının gerçek yoğunluğa etkilerini araştırmak için tam faktöriyel deney tasarımı çalışması yapılmıştır.

2-Grafit miktarının ve basınç miktarının gözenek miktarına etkilerini araştırmak için tam faktöriyel deney tasarımı çalışması yapılmıştır.

3-Mesafenin, sıkıştırma basıncının ve grafit miktarının aşınma miktarına etkilerini araştırmak için tam faktöriyel deney tasarımı çalışması yapılmıştır.

4-Grafit miktarının ve basınç miktarının %95 güven aralığında gerçek yoğunluğu ve gözenek miktarını etkilediği, bu faktörlerle beraber mesafenin aşınma miktarını %95 güven aralığında etkilediği görülmüştür.

### KAYNAKLAR

[1] M. H. Cho, J. Ju, S. J. Kim, H. Jang, "Tribological Properties of Solid Lubricants (graphite, Sb2S3, MoS2) for Automotive Brake Friction Materials", Wear, 260: 855-860 2006.

[2] K.Tüfekçi, C. Kurbanoglu, E.Durak, R. F.Tunay, "Friction and Wear Properties of Cu and Fe-based P/M Bearing Materials", Journal of Mechanical Science and Technology, 20 (44): 513-521 2006.

[3] Y. Atak, "Kendinden Yağlamalı Grafit Katkılı Bronz Yatakların Aşınma Davranışının İncelenmesi" Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 2009.

### INVESTIGATION OF WEAR PROPERTIES OF GRAPHITE ADDED SELF-LUBRICATED BRONZE BEARINGS

In this study, the effects of graphite addition on wear performance of self lubricated bronze bearings produced by P/M technology were investigated using Design of Experiment. Using pin-on-disc test apparatus, self-lubricated bearing materials were tested at 1 m/s sliding velocity, 15 N load, and 2000 m distance in laboratory conditions. Wear rates were calculated. The results indicated that the wear performance of self-lubricated bronze bearing materials were dependent on the amount of graphite addition and compacting pressure. When the pressing pressure was increased, the density of the bearing materials increased, but percentage of porosity decreased. Results of wear tests showed that with adding graphite and increasing pressing pressure there was a decrease in wear rate.

**Key Words:** P/M, self-lubricated, bearing, bronze, graphite, wear, porosity.

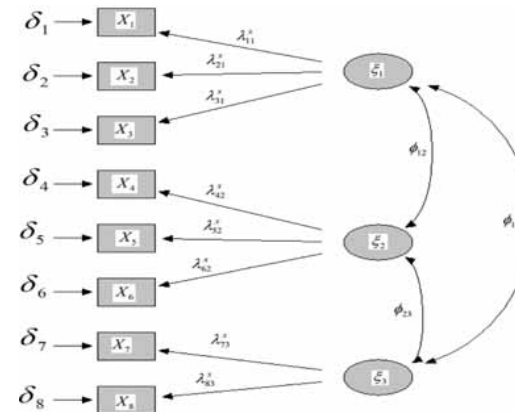
### DOĞRULAYICI FAKTÖR ANALİZİNDE ÖRNEKLEM HACMİ, TAHMİN YÖNTEMLERİ VE NORMALLİĞİN UYUM ÖLÇÜTLERİNE ETKİSİ

Murat DOĞAN, Özer ÖZAYDIN, Veysel YILMAZ  
Eskişehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 26040, Eskişehir, TÜRKİYE,  
mdogan@ogu.edu.tr, oozaydin@ogu.edu.tr, vyilmaz@ogu.edu.tr

### Giriş

Yapısal Eşitlik Modelleri (YEM), nedensel ilişkilerin tahmin edilmesinde ve test edilmesinde kullanılan bir yöntemdir. Doğrulayıcı Faktör Analizi (DFA) ise gizil ve gözlenen değişkenler arasındaki ilişkilerin ölçüm modelleriyle ilgilene YEM'in bir parçasıdır. Bu çalışmada, DFA uyum ölçütlerinin farklı koşullar altındaki karakteristik özelliklerinin belirlenmesi amacıyla MC simülasyonu yardımıyla yapılmıştır. Bu simülasyon çalışmasında, örneklem hacminin 7 farklı durumu (50, 100, 200, 400, 800, 1600, 4000), tahmin yönteminin 4 farklı durumu (En Çok Olabilirlik, En Küçük Kareler, Genelleştirilmiş En Küçük Kareler, Ağırlıklandırılmış En Küçük Kareler) ve dağılımsal koşulun 3 farklı durumu (çok boyutlu normallik varsayımı, az derecede normallikten sapma, orta derecede normallikten sapma) ele alınmıştır. Simülasyon çalışması, bu koşulların, CFA'da ve YEM'de en çok kullanılan 11 uyum ölçütüne ( $\chi^2$ , GFI, AGFI, IFI, MFI, NFI, NNFI, CFI, RMR, SRMR ve RMSEA) etkisini incelemek amacıyla EQS yazılımı yardımıyla oluşturulmuştur.

### Yöntem



Şekil 1. Çalışmada kullanılan CFA Modeli

Simülasyon çalışması Paxton ve arkadaşlarının (2001) yaptığı gibi 9 adımda oluşturulmuştur. Bu adımları; ilgilene araştırma sorusunun teorik olarak geliştirilmesi, soruya uygun modelin oluşturulması, çalışmaya uygun deneysel koşulların seçilmesi, kitle parametre değerlerinin seçilmesi, uygun yazılım programının seçimi, simülasyonun yürütülmesi, simülasyon çıktılarının uygun şekilde depolanması, sorun giderme ve doğrulama ve sonuçların özetlenmesi şeklinde özetlenebilir. Çalışmanın araştırma modeli Raykov ve Marcoulides (2006) tarafından oluşturulan ve Şekil 1'de verilen DFA modeli kullanılmıştır.

### Sonuç

Çalışma sonucunda ki kare uyum ölçüsü çok boyutlu normallik varsayımı altında örneklem hacmi arttıkça büyük değerler olarak gerçekçi olmayan sonuçlar verirken, normallik varsayımının olmadığı durumlarda ise örneklem hacminin büyümesinden çok fazla etkilenmemektedir. IFI ve NNFI normallik varsayımının sağlanmadığı az derecede ve orta derecede normal olmayan durumlarda tutarsız sonuçlar verdiği için kullanılmaması tavsiye edilebilir. NFI ve CFI uyum indeksleri ML ve LS tahmin yönteminde ve sadece normallik varsayımının sağlandığı durumlarda iyi sonuç vermektedir dolayısıyla sadece bu şartlar altında kullanılması önerilebilir.

Bu çalışma yurtdışında sık bir şekilde kullanılan EQS yazılımının Türkiye’de YEM üzerine ilk olarak bir simülasyon çalışmasında kullanılmıştır. Bu durumum YEM kullanan araştırmacılar için farklı bir bakış açısı olacağı değerlendirilmektedir.

#### KAYNAKLAR

[1] Paxton P., , Patrick J. Curran , Kenneth A. Bollen , Jim Kirby & Feinian Chen (2001), Monte Carlo Experiments: Design and Implementation, Structural Equation Modeling: A Multidisciplinary Journal, 8:2.

[2] Raykov, T. and Marcoulides, G.A., (2006), A first course in structural equation modeling-2nd ed., Lawrence Erlbaum Associates, Mahlah-New Jersey-London, 238 p.

#### ABSTRACT

#### INFLUENCE OF SAMPLE SIZE, ESTIMATION METHOD AND NORMALITY ON FIT INDICES IN CONFIRMATORY FACTOR ANALYSIS

Structural equation modeling (SEM) is a statistical technique for testing and estimating causal relations, they are suited to both theory testing and theory development. Confirmatory Factor Analysis (CFA) is a type of SEM that deals specifically with measurement models, that is, the relationships between observed and latent variables. In this study, Monte Carlo (MC) simulation is used to evaluate the characteristics of CFA fit indices under different conditions. A simulation study was conducted with EQS software. As a result of this study, all of the conditions discussed have influence on fit indices.

**Key Words:** Confirmatory Factor Analysis, Monte Carlo Simulation, EQS, Structural Equation Modeling, Fit Indices

#### SAĞLAM ÖLÇEK TAHMİN EDİCİLERİNİN ÇARPIKLIĞA BAĞLI OLARAK DAVRANIŞLARININ KARŞILAŞTIRILMASI

Celal AYDIN, Necla GÜNDÜZ\*

Gazi Üniversitesi Fen Fakültesi İstatistik Bölümü 06500 Ankara Türkiye  
aydinc@gazi.edu.tr, guduznecla@yahoo.com

#### Giriş

Dağılımın tam olarak belli olmadığı durumlarda, ölçek parametresine ilişkin tahmin edicileri, parametrik olmayan ölçüleri kullanarak elde etmek daha güvenlidir. Bu tür tahmin ediciler söz konusu olduğunda, sağlamlık büyük önem kazanmaktadır.

Bu çalışmada yer verilen ölçek tahmin edicileri, iyi bir sağlamlık ölçüsü olarak kabul gören kırılma noktası (breakdown point) değeri bakımından her üçü de %50 değerine sahiptir. Dağılımın yoğunlaştığı aralığı öngörmek bakımından iş görebilecek özellikte olan ve aynı derecede sağlam olan üç ölçek tahmin edicisi, dağılımın çarpıklığına karşı davranışları bakımından incelenmeye çalışılmıştır.

Ölçek tahmin edicileri hesaplanırken, lokasyon tahmin edicisi olarak, medyan istatistiğini kullanmak uygun olmaktadır ve  $\{x_1, x_2, \dots, x_n\}$  rassal örneği için medyan istatistiği;

$$Med_i x_i = Med\{x_1, x_2, \dots, x_n\} \quad \text{dir.}$$

Medyandan mutlak sapmaların medyanı olarak tanımlanan  $MAD_n$  istatistiği,  $\{x_1, x_2, \dots, x_n\}$  rassal örneği için;

$$MAD_n = Med_k |x_k - Med_j x_j| \quad \text{dir.}$$

$MAD_n$  istatistiği, gözlemlerin medyan değerlerinden olan uzaklıklarına odaklanmaktadır. Bu bakımdan, medyan merkezli,  $MAD_n$  uzunluklu simetrik aralık, simetri özelliği olan dağılımlarda, dağılımın yarısını kapsamaktadır [1].

İkinci olarak,  $S_n$  istatistiği,  $\{x_1, x_2, \dots, x_n\}$  rassal örneği için;

$$S_n = Med_i \{Med_k |x_k - x_i|\} \quad \text{dir.}$$

Her  $i$  için,  $\{x_l - x_i : l = 1, 2, \dots, n\}$  kümesinin medyanı bulunur. Bu şekilde elde edilen  $n$  tane medyan değerinin de medyan alınarak,  $S_n$  istatistiği elde edilir [2].

Tanımdan anlaşıldığı gibi,  $S_n$  istatistiği dikkatini, gözlemlerin bir merkezi değerden ne kadar uzak olduklarına çevirmek yerine, gözlemlerin birbirinden uzaklıklarına çevirmiş durumdadır.  $S_n$  istatistiği hesaplanırken, lokasyon tahminine ihtiyaç bulunmamaktadır.

Son olarak  $Q_n$  istatistiği, ikili mutlak farkların  $k$ 'inci sıra istatistiği olarak;  $h = \lceil n/2 \rceil + 1, \quad k = \binom{n}{2} \approx \binom{n}{2}/4$  olmak üzere;

$$Q_n = \{ |x_i - x_j| : i \leq j \}_{(k)} \quad \text{dir.}$$

$Q_n$  istatistiğinin hesaplanma zamanı,  $S_n$  istatistiğine göre hayli yüksektir ancak kolayca uygulanabilmektedir [2].

Gamma dağılımı iki parametrelili olup, simetrik olmayan bir ailedir. Dağılımın parametrelerinden birinin bir fonksiyonu çarpıklık ölçüsü olarak karşımıza çıkmaktadır [3].

Söz konusu tahmin edicilerin çarpıklığa bağlı olarak nasıl bir değişme gösterdiklerini görmek bakımından Gamma ailesinin uygun olduğu düşünüldü. Farklı çarpıklık derecelerindeki Gamma dağılımlarından, simülasyon yoluyla farklı örnek çaplı rassal örnekler elde edilerek, söz konusu tahmin edicilerin değerleri her rassal örnek için hesaplanarak tablolaştırıldı. Tahmin edicilerin çarpıklığa ve örnek çapına bağlı olarak davranışları incelenmeye çalışıldı.

#### KAYNAKLAR

- [1] Serfling, R., Mazumder, S., (2009), Exponential probability inequality and convergence results for the media absolute deviation and its modifications. *Statistics and Probability Letters* 79, 1767-1773.
- [2] Rousseeuw, J. P., Croux, C., (1993), Alternatives to the median absolute deviation. *Journal of the American Statistical Association* 88, 1273-1283.
- [3] Bannheka, B., M., S., G., Ekanayake, G., E., M., U., P., D., (2009), A new point estimator fort he median of gamma distribution. *Vidyodaya J. Of Sci.* 14, 95-103.

#### ABSTRACT

#### COMPARISONS OF THE BEHAVIOUR OF ROBUST SCALE ESTIMATORS IN RESPECT OF SKEWNESS

In the situations where the distribution is not certain, it is more appropriate to use nonparametric measures to obtain scale estimators. Robustness became important. Three scale estimators with the same degree of robustness and which are also capable for the prediction of condensation range of distribution are examined as a function of skewness in the distributional sense.

**Key Words:** Scale estimation, robustness, breakdown point, median absolute deviation, skewness.

#### MODEL BELİRLEMESİ, ÖRNEKLEM HACMİ VE TAHMİN YÖNTEMİNİN YAPISAL EŞİTLİK MODELLERİ UYUM ÖLÇÜLERİNE ETKİSİ

Rana ŞEN Veysel YILMAZ  
Eskişehir Osmangazi Üniversitesi, İstatistik Bölümü, 26040, Eskişehir, TÜRKİYE  
ranasen@ogu.edu.tr, vyilmaz@ogu.edu.tr

#### Giriş

Yapısal Eşitlik modelleri (YEM), gözlenen değişkenlerin doğrusal bileşimi olarak yazılabilen çok sayıda içsel ve dışsal gizil değişkeni birlikte ele alan bir modelleme yöntemidir. Bu yüzden, YEM araştırmalarında model uyumunun değerlendirilmesi zorlu bir konudur. YEM'de model uyumunun deneysel olarak değerlendirilmesi ve istatistiksel tahminlerin elde edilmesinde Monte Carlo (MC) simülasyonu yaygın olarak kullanılmaya başlanmıştır.

Bu çalışmada, model belirlenmesinin, örneklem hacminin ve tahmin yönteminin YEM'de kullanılan uyum ölçütlerine etkisi, bir MC simülasyonu tasarlanarak araştırılmıştır. Bu deneysel dizaynda, veriler EQS programı yardımıyla, bir kitle kovaryans matrisinden üretilmiş ve farklı model belirlenmesi (doğru model, az yanlış model, orta derecede yanlış model) durumlarındaki YEM modellerine uydurulmuştur. Çalışmada, örneklem hacminin 6 düzeyi (50, 100, 200, 400, 800 and 1600), tahmin yönteminin ise 3 düzeyi (En Çok Olabilirlik (ML), En Küçük Kareler (LS) ve Genelleştirilmiş En Küçük Kareler (GLS)) incelenmiştir. En çok kullanılan 11 uyum ölçütü (Ki Kare, GFI, AGFI, NFI, NNFI, CFI, IFI, MFI, RMR, SRMR, RMSEA) üzerinde çalışılmıştır.

#### Yöntem

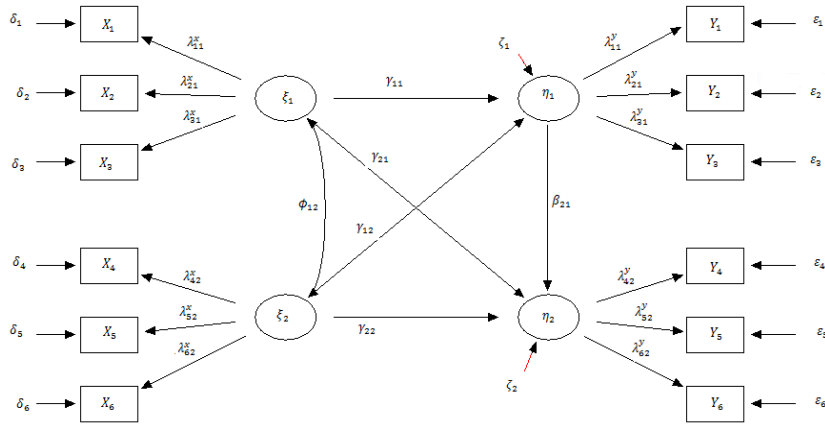
Yapısal eşitlik modelleri, yapısal modelin ve ölçüm modelinin birleşiminden oluşur. Yapısal model, gizil değişkenler arasındaki ilişkileri özetleyen yapısal eşitlikleri içerir. Yapısal model, matris notasyonuyla Eşitlik 1'de verildiği gibidir. Ölçüm modeli ise, gözlenen değişkenlerle bağlı oldukları gizil değişkenler arasındaki ilişkileri tanımlayan eşitlikleri içerir. Ölçüm modeli, matris notasyonuyla Eşitlik 2 ve Eşitlik 3'de verildiği gibidir.

$$\eta = B\eta + \Gamma\xi + \zeta \quad (1)$$

$$Y = \Lambda_y\eta + \varepsilon \quad (2)$$

$$X = \Lambda_x\xi + \delta \quad (3)$$

YEM çalışmalarında, özellikle son yıllarda sıkça rastladığımız Monte Carlo (MC) simülasyonu, belirli şartlar altında uyum ölçütlerinin performanslarının değerlendirilmesi amacıyla veri üretiminde kullanılan bir simülasyon yöntemidir (Paxton et.al., 2001). Çalışmada kullanılan araştırma modeli Şekil 1'deki gibidir.



## BURR XII DAĞILIMININ KONUM VE ÖLÇEK PARAMETRELERİNİN

### TAHMİNİ: MONTE-CARLO SİMÜLASYON ÇALIŞMASI

Fatma Gül AKGÜL<sup>1\*</sup>, Birdal ŞENOĞLU<sup>1</sup>

Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06100 Tandoğan, Ankara<sup>1</sup>  
fgakgul@ankara.edu.tr, senoglu@science.ankara.edu.tr

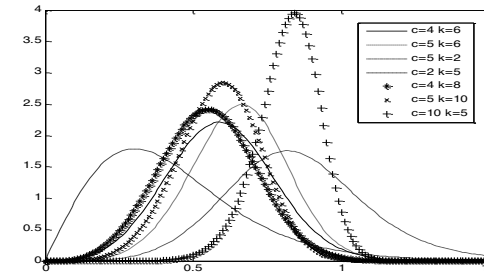
#### Giriş

Burr (1942) tarafından önerilen dağılım mühendislik, güvenilirlik, hidroloji, yaşam modelleri ve aktüerya gibi birçok alanda yaygın uygulama alanlarına sahiptir. Burr XII dağılımının olasılık yoğunluk fonksiyonu

$$f(x) = \frac{kc}{\sigma} \left(\frac{x-\mu}{\sigma}\right)^{c-1} \left(1 + \left(\frac{x-\mu}{\sigma}\right)^c\right)^{-(k+1)}, x > \mu, c, k, \sigma > 0, -\infty < \mu < \infty \quad (1)$$

şeklinde ifade edilir. Burada,  $\mu$  konum,  $\sigma$  ölçek,  $c$  ve  $k$  ise şekil parametreleridir.

Şekil parametreleri  $c$  ve  $k$ 'nin farklı değerleri için Burr XII dağılımının olasılık yoğunluk fonksiyonuna ait grafikleri aşağıda verildiği gibidir. Şekil parametrelerinin aldığı değerlere göre Burr XII dağılımı simetrik veya çarpık olabildiği gibi, kısa kuyruklu veya uzun kuyruklu da olabilir. Bu durum uygulamada verileri modellerken esneklik sağlar.



Şekil 1. Burr XII dağılımının farklı şekil parametreleri için olasılık

yoğunluk fonksiyonu grafikleri

Literatürde, Burr XII dağılımının parametrelerinin tahmin edicileri ile ilgili birçok çalışma mevcuttur, bkz. Wingo (1983), Hossain ve Nath (1997) ve Shao (2004).

Bu çalışmada, Burr XII dağılımının şekil parametrelerinin bilindiği varsayımı altında konum ve ölçek parametrelerinin en küçük kareler (least squares-LS), momentler (method of moments-MM), en çok olasılık (maximum likelihood-ML) ve uyarlanmış en çok olasılık (modified maximum likelihood-MML) tahmin edicilerinin etkinlikleri,  $c$ ,  $k$  ve  $n$ 'nin farklı değerleri için, Monte Carlo simülasyon çalışması ile karşılaştırılmıştır. Tahmin edicilerinin etkinlikleri karşılaştırılırken

$$Def(\hat{\mu}, \hat{\sigma}) = MSE(\hat{\mu}) + MSE(\hat{\sigma}) \quad (2)$$

kullanılmıştır.  $Def$  birden fazla tahmin edicinin etkinliğinin birlikte değerlendirilmesi amacıyla kullanılan ve  $MSE$ 'lerin toplamı olarak ifade edilen bir kavramdır.

#### Şekil 1.Araştırma Modeli

##### Bulgular

Bu çalışmada, model belirlemesine en duyarlı uyum ölçütünün SRMR, örneklem hacmine en duyarlı uyum ölçütlerinin ise GFI, AGFI ve NFI olduğu bulunmuştur. Model belirlemesi, örneklem hacmi ve tahmin yöntemi üçlü etkileşimine duyarlı uyum ölçütlerinin ise sadece GFI, AGFI ve NFI olduğu gözlenmiştir.

#### KAYNAKLAR

[1] Paxton, P., Curran, P.J., Bollen, K.A., Kirby, J. and Chen, F. (2001), Monte carlo experiments: design and implementation, Structural Equation Modeling: A Multidisciplinary Journal, 8, 2, 287-312.

#### ABSTRACT

##### EFFECTS OF MODEL MISSPECIFICATION, SAMPLE SIZE AND ESTIMATION METHODS ON STRUCTURAL EQUATION MODELING FIT INDICES

In this study, a Monte Carlo simulation study was conducted to investigate the effects on structural equation modeling (SEM) fit indices of, model specification, sample size and estimation method. Based on balanced experimental design, samples were generated from a population covariance matrix and fitted to structural equation models with different degrees of model misspecification (true, slightly misspecified or moderately misspecified models) with EQS software. Six levels of sample size (50, 100, 200, 400, 800 and 1600) and three levels of estimation method (Maximum Likelihood (ML), Least Square (LS) and Generalized Least Square (GLS)) were examined.

**Key Words:** Structural Equation Modeling, Fit Indices, Model Misspecification, Monte Carlo Simulation

Diğer çalışmalardan farklı olarak bu çalışmada Tiku (1967) tarafından önerilen MML yöntemi karşılaştırmalara dâhil edilmiştir. Bildiğimiz kadarıyla daha önce Burr XII dağılımının konum ve ölçek parametreleri tahmin edilirken MML yöntemi kullanılmamıştır. Olabilirlik denklemlerinin logaritmasının ilgili parametrelere göre türevleri alınıp sıfıra eşitlendiğinde analitik çözüme ulaşılamazsa iteratif yöntemler tercih edilir. Ancak iteratif yöntemler kullanılırken birden fazla köke yakınsama, yanlış köke yakınsama veya hiç yakınsamama v.b. gibi problemlerle karşılaşılabilir. MML yöntemi bahsedilen problemlerden kaçınmak ve parametrelerin tahmin edicilerinin açık çözümlerini elde etmek için önerilen bir parametre tahmin yöntemidir.

Monte-Carlo simülasyon çalışmasının sonucunda ML ve MML yöntemi ile bulunan tahmin edicilerin etkinliklerinin diğer yöntemlere göre daha yüksek olduğu görülmüştür. Bunun yanı sıra, literatürden alınan gerçek bir veri seti kullanılarak Burr XII dağılımının konum ve ölçek parametreleri tahmin edilmiştir.

#### KAYNAKLAR

- [1] Burr, I.W. (1942), Cumulative frequency functions, *Annals of Mathematical Statistics* 13, 215–232.
- [2] Hossain, A.M. and Nath, S.K. (1997), Estimation of parameters in the presence of outliers for a burr XII distribution, *Communications in Statistics-Theory and Methods*, 26(3), 637-652.
- [3] Shao, Q. (2004), Notes on maximum likelihood estimation for the three-parameter Burr XII distribution, *Computational Statistics & Data Analysis*, 45, 675–687.
- [4] Tiku M. L. (1967), Estimating the mean and standard deviation from a censored normal sample, *Biometrika*, 54, 155-165.
- [5] Wingo, D.R. (1983), Maximum likelihood methods for fitting the Burr Type XII distribution parameters to life test data, *Biometrical Journal*, 25, 77–84.

#### ABSTRACT ESTIMATION OF THE LOCATION AND THE SCALE PARAMETERS OF BURR XII DISTRIBUTION: MONTE-CARLO SIMULATION STUDY

In this study, when the shape parameters are assumed to be known, least squares estimators (LS), method of moment estimators (MM), maximum likelihood estimators (MLE), modified maximum likelihood estimators (MML) are used to estimate the location and the scale parameters of Burr XII distribution. We compare the performances of these estimators via Monte Carlo simulation study for different sample sizes and shape parameter values with regards to their joint efficiency.

**Key Words:** Burr XII distribution, least squares, method of moments, maximum likelihood, modified maximum likelihood, efficiency.

#### Kİ-KARE UYUM İYİLİĞİ TESTİNDE MİNİMUM SINIF SAYISININ T-TESTİ KULLANILARAK BENZETİMLE BELİRLENMESİ

Orhan KESEMEN\*, Buğra Kaan TİRYAKI  
Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, Trabzon, TÜRKİYE  
oke semen@gmail.com, bugrakaantiryaki@gmail.com

#### Giriş

Ki-kare istatistiği uyumun yeterliliği için karar vermek amacıyla kullanılır. “Uyum” terimi gözlenen örneklem dağılımlarının ya da deneysel sonuçlarla elde edilen dağılımların, beklenen ya da normal, binom, Poisson ve düzgün dağılım gibi kuramsal dağılımlarla karşılaştırılması için kullanılır. Beklenen frekansların eğrisi gözlenen frekansların eğrisi üzerine çizilir ve ki-kare istatistiği uyumun yeterli olup olmadığını belirler [1].

Ki-kare uyum iyiliği testi yapılırken sınıf sayısı birçok problemde frekans tablosu içinde hazır olarak verilir. Sınıf sayısını bulmak için ayrı bir hesaplama yapılmaz. Bu tür uyum iyiliği testi problemlerinde sınıf sayısı bilindiğinden serbestlik derecesi kolaylıkla hesaplanır ve test işlemi gerçekleştirilir. Ancak gerçek hayat problemlerinde frekans tablosu, testi yapan kişi tarafından oluşturulur. Sınıf sayısı belirlenirken beklenen frekansın %20’den fazlasının 5’ten küçük olmamasına dikkat edilir. Fakat beklenen frekansın 5’ten küçük olmaması varsayımını gerçekleştirirken, sınıf sayısının gereğinden fazla seçilmesi beklenen frekansın 5’ten daha az olmasına, gereğinden az seçilmesi ise testin gücünün azalmasına, 2.tip hatanın artmasına neden olmaktadır.

#### Yöntemler

Frekans tablosu oluşturulurken, sınıf sayısının belirlenmesinde genel kabul edilebilir bir kural olmamakla birlikte, testi yapan kişinin deneyimine bağlı kalınır ve veriyi en iyi temsil eden yöntem kullanılır. Genel olarak sınıf sayısı 5 ile 20 arasında olması kabul edilir. Sınıf sayısının belirlenmesi için bazı kurallar ve formüller geliştirilmiştir.

- **Sturges Kuralı**  
$$k = [1 + 3,3 \log(n)] \quad (1)$$

n: gözlem sayısı, k: sınıf sayısı

- **Karekök Seçim Formülü**  
$$k = \sqrt{n} \quad (2)$$

n: gözlem sayısı, k: sınıf sayısı

- **Rice Kuralı**  
$$k = [2\sqrt[3]{n}] \quad (3)$$

n: gözlem sayısı, k: sınıf sayısı

- **Doane Formülü**  
$$k = 1 + \log_2(n) + \log_2\left(1 + \frac{|g_1|}{\sigma_{g1}}\right) \quad (4)$$

$$\sigma_{g1} = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}} \quad (5)$$

n: gözlem sayısı, k: sınıf sayısı,  $g_1$ : dağılımın 3. momentinin kestirimi

#### Önerilen Yöntem

Ki-kare uyum iyiliği testinde minimum sınıf sayısı t-testi kullanarak benzetimle belirlenmiştir. Benzetim için sınıf sayısı 2 ile 50 aralığında, örneklem sayısı ise 30'dan başlayıp 10'ar artışla 100'e kadar, sonra 100'den başlayıp 100'er artışlarla 1000 kadar alınmıştır. Ve her bir örneklem sayısı için tüm sınıfların ki-kare değerleri hesaplanmıştır. Hesaplanan ki-kare değerleri 0.01, 0.05, 0.10 anlamlılık düzeylerinde test edilmiştir. Kabul edilmeyen hipotezlerin anlamlı olup olmadığı t-testi ile test edilmiştir ve anlamlı olan minimum sınıf sayısı belirlenmiştir.

#### KAYNAKLAR

- [1] Prof.Dr. Fikri AKDENİZ (2007), Olasılık ve İstatistik, Nobel Kitabevi.
- [2] Sturges, H. A. (1926). "The choice of a class interval". Journal of the American Statistical Association: 65–66.
- [3] Doane DP (1976) Aesthetic frequency classification. American Statistician, 30: 181–183
- [4] Online Statistics Education: A Multimedia Course of Study (<http://onlinestatbook.com/>). Project Leader: David M. Lane, Rice University (chapter 2 "Graphing Distributions", section "Histograms")

#### ABSTRACT

#### DETERMINING MINIMUM NUMBER OF CLASS FOR CHI-SQUARE GOODNESS OF FIT TEST USING T-TEST WITH SIMULATION

In this study, minimum number of class for chi-square goodness of fit test was determined with simulation. For the simulation, number of classes were taken range of 2 to 50 and the number of the sample were taken starting from 30 up to 100 with 10 increments, then from 100 up to 1000 with 100 increments. Also, each samples of the chi-square values were calculated for the number of all classes. Calculated chi-square values were tested 0.01, 0.05, 0.10 significance levels. Moreover, rejected hypotheses were tested by t-test. Lastly, the minimum significant number of classes were determined.

**Key Words:** Frequency distribution, chi-square goodness of fit test, number of bins

## BİLDİRİ OTURUMLARI 4

### SESSION 4

#### *Kalkınma Bakanlığı Özel Oturumu 1*

## BÖLGESEL BÜYÜME ODAKLARI BELİRLEME ÇALIŞMASI

Leyla BİLEN KAZANCIK  
Kalkınma Bakanlığı Necatibey Caddesi No: 110/A 06100 Ankara Türkiye  
lbilen@dpt.gov.tr

### Çalışmanın Kapsamı ve Amacı

Ülkemizde planlı kalkınma dönemiyle birlikte, bölgesel gelişme, bölgelerarası gelişmişlik farklarının azaltılarak kalkınmanın yaygınlaştırılması öncelikli politikalar arasında yer almıştır. Dokuzuncu Kalkınma Planının “Bölgesel Gelişiminin Sağlanması” ve “Rekabet Gücünün Geliştirilmesi” eksenlerinde başta az gelişmiş bölgelerde olmak üzere, büyüme ve çevrelere hizmet verme potansiyeli yüksek “Cazibe Merkezleri” (Bölgesel Büyüme Odakları) belirlenerek, yerel dinamiklere ve içsel potansiyele dayalı gelişme ortamının oluşturulması, kamu yatırım uygulamalarında ve hizmet arzında mekansal önceliklendirme ve odaklanmanın sağlanması amaçlanmaktadır. Bu amaca yönelik olarak bir bölgesel büyüme odağında olması gereken kriterler tespit edilmiş ve çok değişkenli istatistiksel yöntemler kullanılarak Bölgesel Büyüme odaklarının belirlenmeye çalışılmıştır.

### Teorik Çerçeve

Bölgesel Büyüme aracı olarak cazibe merkezleri politikasının kökleri 1960'lı yıllara kadar uzanmaktadır. Büyüme kutupları kuramına (growth pole theory) göre cazibe merkezi olarak belirlendiği kentlere, bölgesel mekânsal yapıyı dönüştürmek amacıyla kamu yatırımları yoğun bir şekilde yönlendirilir. Kamu desteğiyle yaratılan bu “yığılma” ekonomisinin bölgenin tamamı için yaratacağı yeni iş imkânları, üretim kapasitesinin artması gibi dışsallıklar sayesinde bölgesel ölçekte bir refah artışı sağlanması hedeflenmektedir.

Bölgesel büyüme Odakları (Cazibe merkezleri) yaklaşımı ile gelişme potansiyeli bulunan belirli bir kent merkezinin desteklenmesi, kısa dönemde ilgili merkezin bulunduğu bölge içinde bir gelişmişlik farkı oluşmasına yol açacağı öngörülmektedir. Bu sayede, belirlenen kent merkezlerinin kalkınmasına ivme kazandırılacak ve daha sonra bu gelişme çevre merkezlere de sıçrayarak iç göç bölge içerisinde tutulmasına ve ulusal düzeyde bölgesel gelişmişlik farklarının azaltılmasına katkıda bulunacaktır.

Cazibe merkezlerinin belirlenmesinde çok boyutlu ölçekleme, temel bileşenler analizi, diskriminant analizi yöntemleri kullanılmıştır. Çok boyutlu ölçekleme yöntemi (ÇBÖ), nesne ya da birimler arasında gözlemlenen benzerlikler ya da farklılıklardan oluşan uzaklık değerlerine dayalı olarak bu nesnelere tek ya da çok boyutlu uzaydaki gösterimini elde etmeyi amaçlayan, böylece nesnelere arasındaki ilişkilerin belirlenmesini sağlayan çok değişkenli bir istatistiksel analiz yöntemidir. ÇBÖ'nin amacı, mümkün olduğunca az boyutla, nesnelere yapılarını (uzaklık değerleri kullanılarak) orijinal şekle yakın bir şekilde ortaya koymaktır. Bu nedenle ÇBÖ'nün boyut indirgemeyi hedefleyen bir Q tekniği olduğu söylenebilir. ÇBÖ; kümeleme ve diskriminant analizi gibi gruplamayı amaçlayan Q teknikleri arasında yer alırken, aynı zamanda boyut indirgeme özelliği nedeniyle R teknikleri arasında da yer almaktadır.

Diskriminant analizi, birimleri en az hata ile ait oldukları kitlelere ayırmak için geliştirilmiş istatistiksel bir yöntemdir. Genel olarak birimlerin gruplanmasında bazı matematiksel eşitliklerden faydalanılır. Diskriminant fonksiyonu olarak adlandırılan bu eşitlikler, birbirine en çok benzeyen grupları belirlemeye olanak sağlayacak şekilde grupların ortak özelliklerini belirlemek amacıyla kullanılmaktadır. Diskriminant analizi aracılığıyla elde edilen diskriminant (ayırıcı) fonksiyonları, tahmin değişkenlerinin doğrusal bileşenlerinden oluşmaktadır. Diskriminant fonksiyonları gruplar arası farklılığa etki eden değişkenlerinin hangileri olduğunu ortaya çıkarır. Diskriminant analizinin bir diğer önemli ise, gruplardan herhangi birisine ait olan ancak hangi gruptan geldiği bilinmeyen bir birimin ait olduğu grubu en az hata ile saptamasıdır.

### Uygulama ve Sonuçlar

Türkiye'nin görece az gelişmiş Doğu Anadolu, Güneydoğu Anadolu, İç Anadolu ve Karadeniz bölgelerinde (TR7, TR8, TR9, TRA, TRB, TRC Düzey 1 bölgeleri) yer alan Endüstriyel Büyüme odağı olmayan iller çalışma kapsamında yer almıştır. Bölgesel Büyüme Odakları olabilecek illeri belirlemek üzere; ilin tahakkuk eden vergisinin ülke içindeki payı, kişi başına tahakkuk eden vergi, sanayi sektörü istihdamının Ülke sanayi sektöründeki payı, hizmetler sektörü istihdamının, Ülke hizmetler sektöründeki payı, şehir nüfusunun ülke şehir nüfusu içindeki payı, kişi başına elektrik tüketimi, elektrik tüketimin Türkiye içindeki payı, Sosyo-Ekonomik Gelişmişlik Endeksi, öğretim elemanı sayısının ülke içindeki payı, DHMİ yurt içi ve yurt dışı yolcu sayılarının ülke içindeki payı değişkenleri kullanılmıştır. Farklı yöntemlerden elde edilen sonuçların bir karşılaştırması yapılmıştır. Analizler sonucunda **Trabzon, Diyarbakır, Samsun, Erzurum, Van, Elazığ, Malatya, Şanlıurfa, Zonguldak ve Sivas illeri bölgesel büyüme odakları olarak belirlenmiştir.**

## KAYNAKLAR

- [1] Kalkınma Bakanlığı (2013), Bölgesel Gelişme Ulusal Stratejisi (Taslak Doküman), Ankara.
- [2] Selçuk Seratesen, “Bölgesel Gelişimde “Yeni” Bir Politika Aracı: Cazibe Merkezleri”, TEPAV. Değerlendirme Notu. [http://www.tepav.org.tr/upload/files/1271312063r1374.Bolgesel\\_Gelisimde\\_Yeni\\_Bir\\_Politika\\_Araci\\_Cazibe\\_Merkezleri.pdf](http://www.tepav.org.tr/upload/files/1271312063r1374.Bolgesel_Gelisimde_Yeni_Bir_Politika_Araci_Cazibe_Merkezleri.pdf) Erişim tarihi 03.10.2013.
- [3] Kuzey Anadolu Kalkınma Ajansı (2012), Cazibe Merkezleri Özet Raporu.
- [4] KARA, M., 2008. "Bölgesel Rekabet Edebilirlik Kavramı ve Bölgesel Kalkınma Politikalarına Yansımaları", (DPT, Planlama Uzmanlığı Tezi), Ankara, 2008, <http://ekutup.dpt.gov.tr/bolgesel/karam/rekabet.pdf>, 31 Ocak 2010.

## ABSTRACT

### REGIONAL GROWTH POLES DETERMINATION STUDY

In our country, with the period of the planned development, regional development and reducing regional disparities place among the primary policies. In Ninth Development Plan, particularly in underdeveloped regions, "Growth poles" including high potential growth and providing services their surroundings is determined and aim to pave the way for development based on local dynamics and potential.

To this end, the regional growth poles are determined using multivariate methods among the provinces which include the relatively less developed regions Eastern Anatolia, Southeastern Anatolia, Central Anatolia and Black Sea regions (TR7, TR8, TR9, TRA, TRB, TRC Level 1 regions). As a result of analysis, **Trabzon, Diyarbakır, Samsun, Erzurum, Van, Elazığ, Malatya, Şanlıurfa** and **Sivas** is identified as regional growth poles.

**Key Words:** (“Regional Development”, “Growth Pole Theory” Principle Component”, “Multi Dimensional Scaling”)

## TR 71 BÖLGESİ İLÇELERİ SOSYO-EKONOMİK GELİŞİMİŞLİK ANALİZİ

Sevgi SARAÇ, Gonca BUDAK

Ahiler Kalkınma Ajansı, Planlama, Programlama ve Koordinasyon Birimi,  
Atatürk Bulvarı no:57 Merkez-NEVŞEHİR  
sevgisarac@ahika.gov.tr

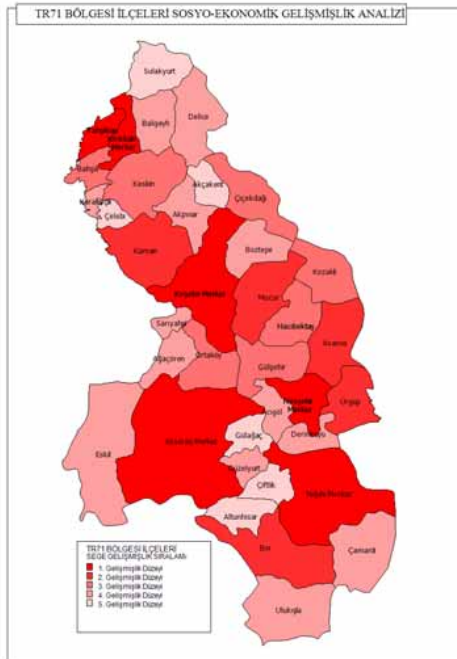
TR71 Bölgesinde yer alan Aksaray, Kırıkkale, Kırşehir, Nevşehir ve Niğde (5 il) illerinin (37 ilçe) ilçelerini kapsayan bu çalışma ile, çeşitli alanlardan seçilen değişkenler baz alınarak, ilçelerin sosyo-ekonomik gelişmişlik seviyelerini tespit etmek ve bu tespit doğrultusunda gelişmişlik sıralamalarını yapmak amaçlanmıştır. Çalışma kapsamında ayrıca, ilçelerin veri setindeki değişken grupları bazında gelişmişlik sıralamaları yapılmıştır.

Çalışmada, sosyo-ekonomik gelişmişliği olumlu ve olumsuz yönde etkilediği ya da sosyo-ekonomik gelişmeden olumlu ya da olumsuz etkilendiği düşünülen 35 farklı değişken kullanılmıştır. Değişkenler ilişkili oldukları gelişmişlik boyutuna göre demografi, eğitim, sağlık, rekabetçi kapasite ve istihdam, tarım, mali ve yaşam kalitesi göstergesi gruplarında sınıflandırılmıştır.

Çalışma, istatistiki anlamlılık testleri yapılabilen ve birçok ülkede değişik araştırmacılar tarafından benzer amaçlarla yaygın olarak kullanılan, temel bileşenler analizi (principal components analysis) tekniği kullanılarak gerçekleştirilmiştir.

Çalışma sonucunda, TR71 Bölgesi kapsamındaki beş merkez ilçenin en üst sıralarda yer aldığı görülmektedir. Sıralamada, gelişmişlik endeks değeri 2,098 olan Aksaray merkez ilçesi ilk sırayı alırken gelişmişlik endeks değeri 1,57 olan Kırıkkale'nin Çelebi ilçesinin son sırayı aldığı görülmüştür. Demografik göstergeler ve sağlık göstergeleri bakımından Kırıkkale merkez ilçesi ilk sırayı alırken, eğitim göstergelerinde Kırıkkale'nin Yahşihan ilçesi, rekabetçi kapasite, tarım ve mali göstergeler bakımından Aksaray merkez ilçesinin ilk sırada yer aldığı dikkat çekmektedir.

İlçelerin gelişmişlik endeksi değeri göz önünde bulundurularak ilçeler ayrıca beş ayrı sınıfa ayrılmıştır. En gelişmiş olan birinci kademede 6, ikinci kademede 5, üçüncü kademede 7, dördüncü kademede 13 ve en az gelişmiş olan beşinci kademede 6 ilçe yer almıştır.



Harita 1. TR71 Bölgesi İlçeleri Gelişmişlik Sıralaması

Çalışmanın son bölümünde, ilçeler gösterge gruplarındaki değişkenlerde aldıkları değerlere göre değerlendirilmiş ve son olarak ilçelerin bulunduğu il içerisinde incelemesi yapılmıştır.

## KAYNAKLAR

- [1] BARTLETT, M.S. (1950), Tests of Significance in factor analysis, British Journal Of Psychology Statistical Section.
- [2] DUNTEMAN, G. J. (1989), Principal Component Analysis, Thousands Oaks, CA: Sage Publications.
- [3] JOHNSON, R. A. . (1992), Applied Multivariate Data Analysis, Prentice Hall, New Jersey.
- [4] ÖZASLAN, M. & DİNÇER, B. (2004). İlçelerin Sosyo-Ekonomik Gelişmişlik Sıralaması Araştırması, Ankara: DPT.
- [5] TATLIDİL, H. (1996). Uygulamalı Çok Değişkenli İstatistiksel Analiz, Cem Ofset, Ankara.

## ABSTRACT

### SOCIO-ECONOMIC DEVELOPMENT ANALYSIS OF TR71 REGION PROVINCES

The goal of this study is to identify the level of development of TR71 region's provinces (that are located in Aksaray, Kırıkkale, Kırşehir, Nevşehir, Niğde) and to rank by taking various variables at district level as a basis. In the study, 35 different indicators grouped under major categories such as; demography, education, health, competitive capacity, employment, agriculture and finance were employed in principal component analysis in order to measure the level of development of districts.

**Key Words:** Principal Components Analysis, Socio-Economic Development Index



## KATKISALLIK DEĞERLENDİRMESİ: ÇUKUROVA VE İZMİR KALKINMA AJANSLARININ KOBİ DESTEKLERİ

Zeyneb ERSAYIN  
Kalkınma Bakanlığı Necatibey Caddesi No: 110/A 06100 Ankara Türkiye  
zeyneb.ersayin@kalkinma.gov.tr

### Çalışmanın Kapsamı

Bu çalışmada, müdahalelerin amaçladığı değişiklikleri yaratma derecesini araştıran katkısallık değerlendirme ele alınmıştır. Öncelikle katkısallığın kavramsal çerçevesi ortaya konulmuş, katkısallığın unsurları ve türleri incelenmiştir. Daha sonra Türkiye’de KOBİ’lere yönelik destekler sunmaya başlayan kalkınma ajanslarından Çukurova ile İzmir Kalkınma Ajanslarının KOBİ’lere yönelik olarak uyguladıkları ilk programlar olan İktisadi Kalkınma ve KOBİ Mali Destek Programlarının katkısallıkları değerlendirilmiştir.

Söz konusu değerlendirme kapsamında proje bazlı bir yaklaşım tercih edilmiş ve desteklenen KOBİ’lere anket uygulanmıştır. Anket sonuçlarının analiz edilmesiyle programların katkısallıkları değerlendirilmiş ve katkısallığı etkileyebilecek faktörler belirlenmiştir. Elde edilen sonuçlar doğrultusunda kalkınma ajanslarının destek uygulamalarına ilişkin öneriler getirilmiştir.

### Katkısallık Değerlendirmesinin Kavramsal Çerçevesi

Müdahalelerin amaçladığı net değişiklikleri meydana getirebilme derecesine katkısallık denilmektedir. Katkısallığın tanımında birbiriyle iç içe geçmiş iki husus ön plana çıkmaktadır; müdahaleden kaynaklanma ve değişiklik. Müdahaleden kaynaklanma hususu katkısallık değerlendirmesinde “müdahale olmasaydı ne olacaktı?” sorusu ile belirlenmeye çalışılmaktadır. Müdahalenin olmadığı durumda da gerçekleşeceği düşünülen değişiklikler müdahaleden kaynaklanmış sayılmamakta<sup>1</sup> ve dara kaybı (deadweight) olarak adlandırılmaktadır.

Katkısallığın tanımındaki ikinci husus olan, müdahalenin gerçekten bir değişikliğe neden olup olmadığının belirlenmesinde dara kaybına ek olarak, hedef gruptan sapmalar (sızıntı etkisi-leakage), destek almayan bölge ya da gruplarda meydana gelen olumsuz etkiler (de-plasman-displacement), ikame (substitution) ve çarpan (multiplier) etkilerinin de dikkate alınması gerekmektedir. Katkısallığın unsurları olarak adlandırılan bu hususların değerlendirilmesiyle hedef grupta meydana gelen gerçek değişiklikler görülebilecektir.

Diğer yandan, katkısallık genel olarak, girdi ve çıktı katkısallıkları ile davranışsal katkısallık olmak üzere üç türde ele alınmaktadır<sup>2</sup>. Ayrıca bu üç katkısallık türünün ortak yanlarını kendinde toplayan proje katkısallığı da bu sınıflandırmaya dahil edilmekte ve kamu desteği olmadan projenin yapıp yapılmayacağı sorusuna cevap aramaktadır<sup>3</sup>. Müdahalelerin yararlanıcılar üzerinde meydana getirdiği öğrenme etkilerini ayrı bir katkısallık türü olarak ele alan araştırmacılar da bulunmaktadır. Söz konusu çalışmada proje katkısallığına odaklanılmış, ayrıca “istihdam edilen kişi” göstergesi üzerinden çıktı katkısallığı da değerlendirilmiştir. Çıktı katkısallığının değerlendirilmesinde farklı alanlardaki projelerin performans göstergelerinin gruplandırılmasının mümkün olmaması nedeniyle, programların amaçlarıyla örtüşür şekilde, en genel gösterge olan istihdam artışına odaklanılmıştır.

### Örnek Uygulama ve Elde Edilen Sonuçlar

Uygulanan anket ile öncelikle firmaların büyüklüğü, faaliyet gösterdikleri sektör gibi temel bilgiler alınmış, daha sonra katkısallığın unsurlarına dair çıkarımlarda bulunulmasını sağlayacak sorular yöneltilmiştir. Anket sonuçları istatistiksel olarak analiz edilerek firmaların genel özellikleri belirlenmiştir. Daha sonra katkısallık ve unsurlarına dair oranlar hesaplanmıştır. Katkısallık oranının hesaplanmasında kullanılan temel formül şu şekildedir:

$$\text{Katkısallık} = (1 - \text{dara kaybı}) \times (1 - \text{deplasman}) \times (1 - \text{Sızıntı}) \times \text{Çarpan} (1)$$

Yapılan hesaplamada projelerin ortalama katkısallık oranı 0,53 olarak bulunmuştur. Her bir proje için hesaplanan katkısallık oranları bu projelere aktarılan hibe tutarları için kullanıldığında, 22,5 milyon TL’nin 11,4 milyon TL’sinin katkısallık olduğu görülmektedir. 11,1 milyon TL ile ise herhangi bir değişiklik yaratılmamış olmaktadır. Bu durumda girdi katkısallığı 0,51 olarak hesaplanmaktadır. Çıktı katkısallığı kapsamında, istihdam edilen 593 kişiden 246’sının katkısallık olduğu sonucu elde edilmiştir.

### KAYNAKLAR

- [1] Davenport, S., C. Grimes, J. Davies, “Research Collaboration and Behavioural Additionality: A New Zealand Case Study”, Technology Analysis & Strategic Management, Vol.10, No.1, March 1998, pp. 55-67.
- [2] Tokila, A., M. Haapanen, J. Ritsila (2008), Evaluation of Investment Subsidies: When is Deadweight Zero?, International Review of Applied Economics, Vol.22, No.5, 2008, pp.585-600.
- [3] UK English Partnership (2008), Additionality Guide, A Standard Approach to Assessing the Additional Impact of Interventions, UK.

### ABSTRACT

#### ADDITIONALITY EVALUATION: ÇUKUROVA AND İZMİR DEVELOPMENT AGENCY’S SME SUPPORTS

In this study, additionality evaluation which explores the realization degree of the intended changes owing to public interventions has been discussed. Firstly, conceptual framework of additionality has been put forward. As development agencies have begun to provide supports to SMEs in Turkey, additionality of Economic Development and SME Financial Support Programmes of Çukurova and Izmir Development Agencies has been evaluated.

Within the evaluation study, a survey has been conducted to the supported SMEs. By analyzing the survey results, additionality of the programmes has been evaluated and factors that may affect additionality have been determined.

**Key Words:** (“Evaluation”, “Additionality”, “Deadweight”, “Leakage Effect”, “Substitution Effect”, “Displacement”, “Multiplier Effect”, “Financial Support”, SME.)

<sup>1</sup> Tokila at al, 2008:587

<sup>2</sup> Davenport at al, 1998:56

<sup>3</sup> Davenport at al, 1998:56

## SÜRDÜRÜLEBİLİR BÖLGESEL KALKINMANIN ÖLÇÜLMESİ: “SÜRDÜRÜLEBİLİRLİK ÇEMBERLERİ VE ÜÇGENİ”

Leyla BİLEN KAZANCIK, Hasan ŞANALMIŞ, Tuğba DENİZ  
Kalkınma Bakanlığı Necatibey Caddesi No: 110/A 06100 Ankara Türkiye  
tdeniz@kalkinma.gov.tr

### 1. Çalışmanın Kapsamı

Bu çalışma ile günümüzün önemli bir kalkınma paradigması; “Sürdürülebilir Kalkınma” için bölgesel düzeyde ölçüm yöntemi geliştirilmiştir. Çalışma kapsamında öncelikli olarak Sürdürülebilir Kalkınmanın tarihsel süreçteki gelişimi ve teorik altyapısı incelenmiş olup, Türkiye’deki Düzey 2 Bölgeleri bazında sürdürülebilirlik endeksleri, sürdürülebilirlik çemberi ve üçgenleri oluşturulmuştur.

### 2. Kavramsal Çerçeve

1970li yıllardan itibaren<sup>4</sup> ekonomi, toplum ve çevre arasındaki dengeyi gözeterek gelişmeyi hedefleyen “Sürdürülebilir Kalkınma” nosyonu, 1987 yılında Birleşmiş Milletler Dünya Çevre ve Kalkınma Komisyonu tarafından yayımlanan “Ortak Geleceğimiz (Brudtland Raporu)” ile tüm dünyada yaygın olarak kullanılmaya başlanmıştır<sup>5</sup>. Türkiye’de de ülke kalkınması için önemli bir belge olan Kalkınma Planlarında Sürdürülebilirliğin eklenmesi, 1990-1994 yıllarını kapsayan Altıncı Kalkınma Planı ile mümkün olmuştur. Daha sonraları tüm dünyada farkındalığı artırmak amacıyla 1992 yılında gerçekleştirilen Rio Konferansı ile sürdürülebilirliğin çevresel ve ekonomik unsurlarının yanında sosyal boyutunun da üzerinde durulmuştur. 2002 yılındaki Rio+10 ve 2012 yılındaki Rio+20 Konferansları ile de ekonomik büyümenin halen çevresel sorunları artırdığı ve toplumlardaki sosyal adaleti, dayanışmayı gözardı ettiği vurgulanmaya devam etmiştir.

Esasında 1970’lerden sonra gündeme getirilen Sürdürülebilir Kalkınma kuramının araç üretmemesi ve tüm dünyada teorinin uygulamadaki ilgisizliği bugünkü küresel çevre sorunlarının alt yapısını hazırlamış<sup>6</sup> ve gelecekte de gerekli önlemlerin alınmaması durumunda bu sorunların devam etmesinin sözkonusu olacağını işaret etmektedir. Bu bağlamda, son yıllarda tartışmanın yönü “Sürdürülebilir Kalkınmanın Ölçülmesi” konusuna yönelmiştir. Özellikle küreselleşme ile birlikte yerleşmenin de ön plana çıkması sonucu sürdürülebilirliğin bölgesel bazda irdelenmesi ana odak olmuştur.

Çalışmada bölgeler düzeyinde ölçümü gerçekleştirebilmek için 30 değişkenden oluşan bir gösterge seti oluşturulmuş ve 2 tür metod uygulanmıştır. Bu metodlardan birincisi; “Sürdürülebilirlik Çemberi”dir. Kullanılan bir diğer teknik ise temel bileşenler analizidir. Bu analiz yöntemine göre diğer metotta kullanılan 30 gösterge ile sürdürülebilirlik endeksleri oluşturulmuştur. Sürdürülebilirlik üçgeni ise sürdürülebilir kalkınmanın 3 temel boyutu olan çevre, sosyal ve ekonomik değerler üzerinden bölgelerin eğilimini ortaya koymuştur.

### 3. Uygulama ve Elde Edilen Sonuçlar

Çalışmada Türkiye’deki düzey 2 bölgeleri baz alınarak 30 değişken kullanılmıştır. Bu değişkenler ise çevre altyapısına yönelik altı, sosyal ve ekonomik yapıya yönelik 12’şer adet göstergesi kapsamaktadır.

Yapılan analizler ile Türkiye’nin planlama bölgeleri olarak belirlenen Düzey 2 bölgelerinin sürdürülebilirlik çember ve üçgenleri oluşturulmuş ve hangi bölgelerin hangi alanlarda sürdürülebilir olduğu ortaya çıkartılmıştır.

## KAYNAKLAR

- [1] ATKINSON, Dubourg, Hamilton, Munasinghe, Pearce, Young, Measuring Sustainable Development: Macroeconomics and the Environment, Edward Elgar, Cheltenham, UK, Lyme, US, 1997.  
[2] BM Dünya Çevre ve Kalkınma Komisyonu, Ortak Geleceğimiz, Rapor, 1987  
[3] KELEŞ, Ruşen, Kentleşme Politikası, 12. Baskı, Ankara, İmge Kitabevi, 2012.

<sup>4</sup>1970li yılların başında “Sürdürülebilir Kalkınma” için bu türden bir ilgi, "Büyümenin Sınırları" ve 1972’de düzenlenen “Stockholm Konferansı”nda ortaya sürülen önemli tartışmalar ile başlamıştır. Şanalmiş, 2007, s.6.

<sup>5</sup>Keleş, Hamamcı, 2009, s.244.

<sup>6</sup>Şanalmiş, 2007, s.6

[4] MEADOWS, D.H., Meadows, D.L., Randers, J., Behrens Iii, W.W., The Limits to Growth- A Report for the Clup of Rome’s Project on the Predicament of Mankind, A Potomac Associates Book, Second Edition, 1975.

[5] OECD, “Assessing Regional Sustainability; A Benchmark Approach”, 2005.

## ABSTRACT

### MEASUREMENT OF SUSTAINABLE REGIONAL DEVELOPMENT: SUSTAINABILITY WHEELS AND TRIANGLES

In this study, an important development paradigm; “Sustainable Development” is measured in regional level. First of all, the theoretical framework and historical background of “Sustainable Development” is assessed. All over the world, this concept has become widespread from the time that Brudtland Report 7 is published. However, Measurement of Sustainable Development has been perceived more serious in recent times.

Within the study, two models; namely, “Sustainability Wheel” method and “Principle Components Analysis” are used to measure sustainability levels of Turkey’s level 2 regions. In this methods, indicator set composed of 30 variables are used to evaluate which regions are sustainable or not. Finally, with the help of sustainability triangles, regions can be classified according to three areas; social, environmental and economic.

**Key Words:** Sustainability, Development, Sustainability Indicators, Sustainability Wheels, Sustainability Index.

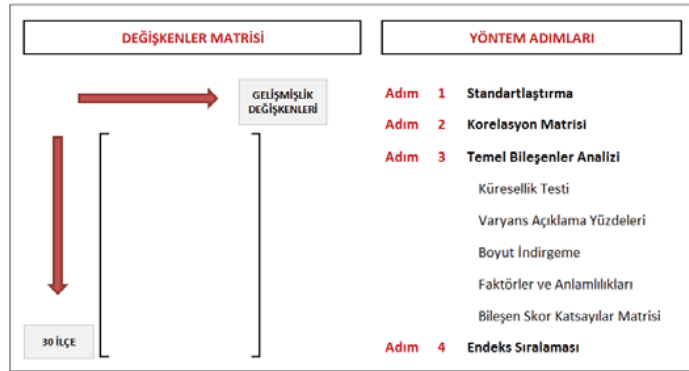
### TRC3 BÖLGE İÇİ REKABET ANALİZİ

Alper DEMİR

Dicle Kalkınma Ajansı, Batman Yatırım Destek Ofisi, Batman, TÜRKİYE, alper.demir@dika.org.tr

#### Analiz Yöntemi (Temel Bileşenler Analizi)

Temel Bileşenler Analizi; bir değişkenler setinin varyans-kovaryans yapısını, bu değişkenlerin doğrusal birleşimleri vasıtasıyla açıklayarak, veri indirgenmesi ve yorumlanmasını sağlayan, çok değişkenli bir istatistik tekniğidir. Çalışmada bu analiz tekniği kullanılarak TRC3 Bölgesi (Mardin, Batman, Şırnak ve Siirt) ilçelerinin gelişmişlik düzeyi ile ilgili bir kanaatin oluşabileceği sosyoekonomik göstergelerin (değişkenlerin) oluşturduğu tek bir bileşen ile ilçelerin diğer ilçeler arasındaki gelişmişlik pozisyonu elde edildi.



Şekil 1. Değişkenler Matrisi ve Yöntem Adımları

#### Gelişmişlik Endeksi Oluşturma

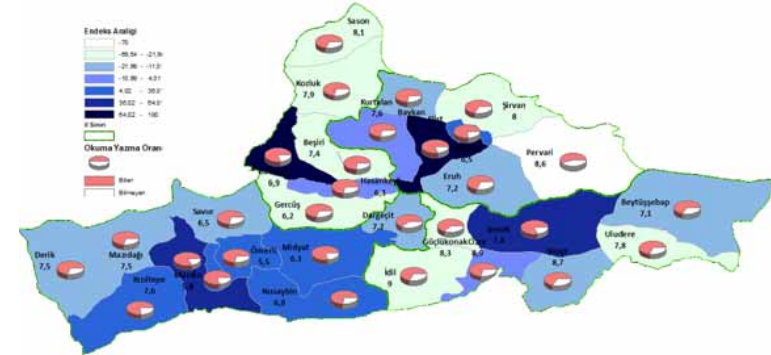
İlçelerin sosyal ve ekonomik gelişmişlik düzeyleri ile tarım ve hayvancılık sektöründeki gelişmişlik düzeyleri endeks değerleri oluşturularak belirlenmeye çalışılmıştır. 12 sosyal yapı değişkeni ile sosyal gelişmişlik, 11 ekonomik yapı değişkeni ile de ilçelerin ekonomik gelişmişlik endeksi belirlenmiştir. Ayrıca toplam 23 adet ilçe değişkeni ile elde edilen sosyoekonomik gelişmişlik endeksi ile beraber bu üç kategoriye ilişkin endeks değerlerine bağlı ilçe sıralamaları oluşturulmuştur.

Tarımsal göstergeler ile bölge ilçelerinin analiz edildiği kısımda da, her birinde altı adet olmak üzere tarım alanı, büyükbaş hayvancılık ve küçükbaş hayvancılık gelişmişliğini etkileyen değişkenler ile sektörlerle ilişkin gelişmişlik endeksleri belirlenmiştir. Ayrıca elde edilen gelişmişlik endeksleri ile beraber bu üç kategoriye ilişkin endeks değerlerine bağlı ilçe sıralamaları oluşturulmuştur.

#### Çizelge 1. Örnek analiz çıktısı: İlçelerin Sosyal gelişmişlik analiz çıktısı

Temel Bileşen	Öz Değeri	Varyans Açıklama Oranı %	Değişkenlerin 1. Temel Bileşendeki Ağırlıkları
1	5,74	47,86	Nüfus 0,1166 Şehirleşme oranı 0,1440 Nüfus artış hızı 0,0880 Nüfus yoğunluğu 0,1298 15+ yaş okuma yazma oranı 0,1442 15+ yaş kadın okuma yazma oranı 0,1344 Yaş bağımlılık oranı -0,1397 Kişi başına düşen Şartlı Nakit Transferi değeri -0,1010 Ortalama toplam hane halkı sayısı -0,0827 Tarım sektörü istihdam oranı -0,1052 Elektrik enerjisi kayıp-kaçak oranı -0,0843 Kişi başına ödenen çevre vergisi değeri 0,1469
2	2,30	19,16	
3	1,35	11,27	

Tüm endeks oluşturma aşamalarında varyans açıklama oranı ile baskın çıkan birinci temel bileşen, tüm değişkenleri temsil edebilen en yetkin bileşen olarak belirlenmiştir.



Şekil 2. Örnek görünüm: TRC3 Bölgesi İlçelerinin Sosyal Gelişmişlik Görünümü

#### KAYNAKLAR

Diñer, Özasan, Kavasoglu, İllerin ve Bölgelerin Sosyo-Ekonomik Gelişmişlik Sıralaması, Kalkınma Bakanlığı, 2003  
Diñer, Özasan, İlçelerin Sosyoekonomik Gelişmişlik Sıralaması Araştırması, Kalkınma Bakanlığı, 2004  
Türkiye İstatistik Kurumu (TÜİK) Veritabanı

#### ABSTRACT

#### COMPETITIVE ANALYSIS IN TRC3 REGION

In this study it is aimed to identify development levels of TRC3 (Mardin, Batman, Şırnak, Siirt) districts among other districts by creating a single component consisting relevant socio-economic indicators.

**Key Words:** Principal component analysis (PCA), development ranking

## BÖLGESEL DÜZEYDE MEKÂNSAL KARARLARIN ALINMASI SÜRECİNDE İSTATİSTİKİ SAHA ARAŞTIRMALARI KONYA KARAMAN ÖRNEĞİ

Fuat KARAGÜNEY\*, Nurten ELVAN KÖKSOY, İsmail ARAS, Mehmet Emin YİĞİTBAŞI, Mehmet BUĞRA AHLÂTÇI  
Mevlana Kalkınma Ajansı, Medrese Mahallesi Ulaşbaba Caddesi No:28 Selçuklu/Konya/Türkiye,  
E- posta: bilgi@mevka.org.tr

### Giriş

Küreselleşme sürecinde bölgelerin rekabet gücünü artırabilmek için bölge içindeki yerleşmeler arasında uzmanlaşma ve iş bölümüne gereksinim vardır. Çünkü bölgelerin ekonomik yapısında önemli değişimler yaşanırken, mekânda da önemli dönüşümlerin gerçekleştiği görülmektedir. Bu değişimlerin ilki metropoliten alanların giderek yayılarak genişlemesi şeklinde ortaya çıkarken yeni ortaya çıkan mekânsal ölçeğin ana odağın çevreinde yer alan farklı sektörlerde uzmanlaşmış yerleşmeleri içine alan bir kent bölge tanımının ortaya çıktığı görülmektedir. Merkez kentlerin konumlarını ve rekabet güçlerini çevrelerindeki, farklı işlevleri üstlenen yerleşimlerle kurdukları işbirliği ağları ile artırdıkları ve kent bölgeye dönüştükleri görülmektedir.8

Diğer yandan bölge içi gelişmişlik farklarının azaltılması bölgesel gelişme sorununun temel bileşenlerinden bir tanesidir. Toplamda 37 ilçeden oluşan Konya Karaman Bölgesi'nde ilçeler arasında ciddi gelişmişlik farkları yaşanmaktadır. Bölgenin sahip olduğu geniş yüzölçümü nedeniyle kırsal ve kentsel alanlar arasındaki fiziki erişim imkânlarının sınırlı kalması ve buna bağlı olarak da mal, hizmet ve insan akımlarında kaliteli ve verimli bir hizmetin sunulamamasıdır.

### Mekânsal Kademelenme Analizi

Bu iki olguyu dikkate alan bu analiz çalışmasında ilk olarak Konya Karaman Bölgesinde yer alan ilçeler arası insan hareketliliğini nicelikselsel olarak izleyebilmek adına 33 ilçede toplam 842 adet "karar örnekleme" yöntemiyle anket çalışması yürütülmüştür. Çalışmada Konya Metropolitan Alanını oluşturan Selçuklu, Meram ve Karatay ilçeleri ile Karaman kent merkezinde anket yapılmamasının nedeni bu ilçelerde yaşayanların zaten ihtiyaç duydukları hizmetleri kendi yaşadığı ilçeden karşılıyor olmasıdır. Saha çalışmalarının ışığında Alman coğrafyacı Walter Christaller tarafından geliştirilen Merkez Yerler Teorisi esas alınarak bölgedeki ilçeler arasındaki insan hareketliliğinin ne yönde olduğu ortaya konmuştur.

### Konya Karaman Bölgesi İlçeler Arası SEGE Analizi

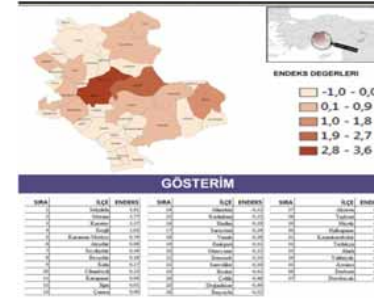
İkinci aşamada istatistikî analiz yöntemlerinde temel bileşenler analizi kullanılarak ilçeler arası yaşanan gelişmişlik farkları analiz edilmiştir. Daha sonra bu gelişmişlik farklarının ortaya çıkmasındaki muhtemel sorun alanları irdelenerek bölge planında bölge içi gelişmişlik farklarının azaltılmasına dair oluşturulacak hedef ve stratejilerin ne yönde olması gerektiğine dair potansiyeller ve fırsatlar analiz edilmiştir.

Temel bileşen analizi ile ulaşılmış istenilen ilk sonuç;  $X_1, X_2, \dots, X_p$  gibi p tane değişkeni, önemli bir bilgi kaybına neden olmaksızın, bu değişkenleri temsil edebilen daha az sayıda değişkene indirgemek ve değişkenlere etki eden genel nedensel faktörleri elde etmektir. Daha sonra indirgenmiş yeni değişkenler ile çalışmanın amacı doğrultusunda çeşitli sonuçlara ulaşılabilmektedir.  $X_1, X_2, \dots, X_p$  vektörlerinin standartlaştırılmış hali olan  $Z_1, Z_2, \dots, Z_p$  vektörlerinin p tane doğrusal birleşimi, ya da temel bileşeni;

$$\begin{aligned} Y_1 &= (a_1)t Z = a_{11} Z_1 + a_{21} Z_2 + \dots + a_{p1} Z_p \\ Y_2 &= (a_2)t Z = a_{12} Z_1 + a_{22} Z_2 + \dots + a_{p2} Z_p \\ Y_p &= (a_p)t Z = a_{1p} Z_1 + a_{2p} Z_2 + \dots + a_{pp} Z_p \end{aligned}$$

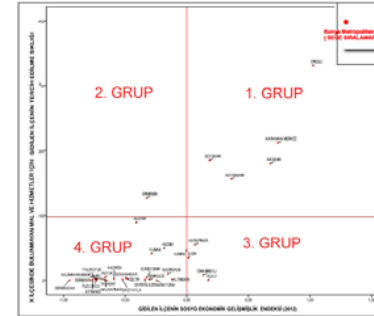
İlçe bazında yapılan bu SEGE analizi Bölgedeki ilçelerin gelişmişlik düzeylerinin sosyo-kültürel ve ekonomik değişkenler yardımıyla ölçülmesi ve birbirleriyle analitik olarak karşılaştırılması TR 52 ilçeleri arasındaki gelişmişlik farklılıklarının azaltılmasına yönelik politikaların oluşturulması bakımından önem taşımaktadır. Ayrıca bu sayede TR52 Düzey2 Bölgesi ilçelerinin sosyal gelişmişlik endekslerini elde edilebilen verilerle oluşturarak soyut bir kavram olan gelişmişlik kavramını yıllar içinde izlemeye olanak sağlanmış olacaktır.

\* A. Eraydın,2001



Harita 1. İlçelerin Sosyo Gelişmişlik Endeksi, 2012

Son olarak ise kademelenme analizi ve temel bileşenler analizi kullanılarak yapılan ilçelerin SEGE analiz sonuçları çaprazlanarak benzer özellikler gösteren mekânsal tipolojiler tanımlanmış ve bölge içinde farklı yerleşim yerlerine farklı müdahale biçimleri geliştirilmiştir.



Şekil 1: Konya Karaman Bölgesi İlçelerinde SEGE ve Kademelenme İlişkisi

### KAYNAKLAR

- [1] Huavari, J., Kangasharju, A. ve Alanen, A. (2001). Constructing an Index for Regional Competitiveness, Pellervo Economic Research Institute Working Papers, No:44, Helsinki.
- [2] Industrial Development Report , UNIDO, 2011
- [3] Ayda Eraydın, KÜRESELLEŞME-YERELLEŞME VE FARKLILAŞAN KENTLER . "Cevat Geray'a Armağan", 2001

**Anahtar Kelimeler:** Bölgesel Gelişme, Kademelenme, Sosyo ekonomik Gelişmişlik, Mekansal Planlama,

THE TRANSLATED GAMMA PROCESS AND FINITE TIME RUIN PROBABILITY: AN  
APPLICATION OF TURKEY COMPULSORY TRAFFIC INSURANCE

Başak BULUT KARAGEYİK\*, Şule ŞAHİN\*

\*Hacettepe University, Faculty of Science, Department of Actuarial Science, 06800, Beytepe- ANKARA.  
basakbulut@hacettepe.edu.tr , sule@hacettepe.edu.tr

**Abstract**

In this study, we present an approximation to calculate finite time survival probabilities and ruin probabilities under the assumption that the aggregate claims process is the translated gamma process. An application of survival and ruin probability calculations based on compulsory traffic insurance data in Turkey has been discussed.

**Key Words:** Gamma Process, Translated Gamma Approximation, Ruin Probability

**1. Introduction**

Modelling the paid and outstanding claims distribution has an important role in non-life actuarial analysis. Amount of claims in the past gives information about the boundary of the future claims. Modelling the claim amounts also helps to allocate the reserve amount at the end of the period. Many different approximation methods have been used for modelling the claim amount distribution. Each of the approximation methods has some limitations, advantages and disadvantages in comparison with the others. The first step in modelling the claim amount distribution is to decide the correct approach.

Ruin probability for finite time can be shown as  $\psi(u, t)$  under the assumption that annual premiums and claims processes do not vary in time. This probability is a significant tool for the insurance company. In case of starting with an appropriate initial surplus, it is an indication whether the insurer's premium and claims processes have been calculated correctly. High ruin probability may cause not only emergence of reinsurance treaties but also upgrade of the risky policies premium amounts or to withdraw cash from capital of the insurance company.

**2. Translated Gamma Distribution**

Gamma distribution has two parameters and its density function is given below:

$$f(x) = \frac{x^{a-1} e^{-\frac{x}{\theta}}}{\theta^a \Gamma(a)} \quad x, a, \theta > 0$$

Translated gamma distribution has a similar structure with gamma distribution but it is assumed to be shifted by some amount  $k$ . The probability density function of translated gamma distribution is given below;

$$f(x) = \frac{(x-k)^{a-1} e^{-\frac{x-k}{\theta}}}{\theta^a \Gamma(a)} \quad x, a, \theta > 0$$

Aggregate claim distribution is fitted to the translated gamma distribution by matching first three moments. The fit of the translated gamma distribution is substantially better than normal distribution for the right tail of the distribution. On the other hand, translated gamma distribution fits poorly the left tail of the distribution and it gives the probabilities even for some negative values (Hardy, 2004).

Since the claim amount has positive values and a right skewed tail distribution, translated gamma approximation can be preferred to other approximations.

**3. Ruin Probability for the Translated Gamma Distribution**

**BİLDİRİ OTURUMLARI 4**

*SESSION 4*

*Actuarial Sciences*

The gamma process was introduced into the actuarial literature by Dufresne, Gerber and Shiu (1991). The idea of the calculation of the survival probability by the translated Gamma distributions was originally discussed by Seal (1978) and then Dickson and Waters (1993) improved the method.

$\{S_G(t)\}_{t>0}$  is a *Gamma*  $(\alpha, \beta)$  process and  $k$  is a constant.  $\delta_{SG}(U, t)$  denotes the probability of non-ruin up to time  $t$ . Since the first three moments of the approximating translated gamma process match those of the compound Poisson process for all values of  $t$ , we would expect  $\delta_{TG}(U, t)$  to be a reasonable approximation to

$\delta(U, t)$  (Dickson and Waters, 1993).

$\delta(U, t) = \Pr(S(\tau) \leq U + c\tau \text{ for all } \tau, 0 < \tau < t)$  by

$\{S(t)\}_{t>0}$  is approximated by translated gamma process  $\{S_{TG}(t)\}_{t>0}$ .

$$S_{TG}(t) = S_G(t) + k$$

$$\begin{aligned} \delta_{TG}(U, t) &= \Pr(S_{TG}(\tau) \leq U + c\tau \text{ for all } \tau, 0 < \tau < t) \\ &= \Pr(r(S_G(\tau) + k\tau) \leq U + c\tau \text{ for all } \tau, 0 < \tau < t) \\ &= \Pr(S_G(\tau) \leq U + (c - k)\tau \text{ for all } \tau, 0 < \tau < t) \end{aligned}$$

The main equations of the survival probabilities are;

$$\delta_{SG}(0, t) = F_{SG}(ct, t) - \frac{1}{c} F_{SG}(ct, t + 1)$$

$$\delta_{SG}(U, t) = F_{SG}(U + ct, t) - c \int_0^t f_{SG}(U + cs, s) F_{SG}(c(t - s), t - s) ds + \int_0^t f_{SG}(U + cs, s) F_{SG}(c(t - s), t - s + 1) ds$$

where  $f_{SG}(x, t)$  and  $F_{SG}(x, t)$  are the density function and distribution function, respectively of a Gamma  $(t, 1)$  random variable (Dickson and Waters, 1993).

#### REFERENCES

- [1] Dickson D.C.M, Waters H., 1993, Gamma Processes and finite time survival probabilities, Astin Bulletin, Vol 23, No 2, 259-272 pp.
- [2] Dufresne F., Gerber H.U., Shiu E.S.W., 1991, Risk Theory and the gamma process, Astin Bulletin, Vol 22, 177-192 pp
- [3] Hardy M. R., 2004, Approximating the Aggregate Claims Distribution, Encyclopedia of Actuarial Science, Wiley London, 76-86 pp.
- [4] Seal, H.L., 1978, From Aggregate claims distribution to probability of ruin, ASTIN Bulletin 10, 47-53 pp.

#### COMPARISON OF DIFFERENT METHODS TO COMPUTE THE GREEKS

Bilgi YILMAZ

Middle East Technical University,  
Institute of Applied Mathematics,  
Financial Mathematics, Ankara-TURKEY  
ybilgi@metu.edu.tr

Bülent Alper İNKAYA

Middle East Technical University,  
Institute of Applied Mathematics,  
Financial Mathematics, Ankara-TURKEY  
ainkaya@metu.edu.tr

Yeliz YOLCU OKUR  
Middle East Technical University,  
Institute of Applied Mathematics,  
Financial Mathematics, Ankara-TURKEY  
yyolcu@metu.edu.tr

#### Introduction

The Greeks in finance are the partial derivatives of a financial quantity with respect to any of the model parameters. These derivatives could serve to measure the stability of the financial quantity under study (e.g. delta is the derivative of an option price with respect to the initial price) or to hedging a certain payoff (Higa and Montero, 2004). The Greeks are useful tools in finance which help to understand how the option reacts to a change in the parameters. The information gained from calculated Greeks of options help the investors in their portfolio management decisions. However the calculation of Greeks are not always straightforward, because of some technical difficulties. For example, the structure of the payoff function of some contingent claims are very complex. For this reason computing their Greeks could be cumbersome. On the other hand, sometimes the payoff function does not have a closed form and one may have to estimate the Greeks numerically, which could be a time consuming and fraught workout. There are essentially three methods used to compute the Greeks: The Finite Difference Method, the Malliavin Calculus and the Finite Element Method.

The most widely used method to compute the Greeks is the finite difference method. This method requires to compute the financial quantity of interest at two nearby points and approximate the differential of the payoff function at that point. The problem here is, the identification of "two nearby points" is not clear. L'Ecuyer and Perron (1994) attempt a solution to resolve this issue asymptotically. Since this method is deeply related with the Kernel Density Estimation method in Statistics, it is also called the Kernel Density Estimation Method.

The Malliavin calculus, also known as Stochastic Calculus of Variations or Calculus in infinite dimensions, was introduced by Paul Malliavin in 1976 (Heno, 2005). It is first used as a tool to prove results in Calculus through the use of probabilistic theory and it is an area of research which for many years has been considered highly theoretical and technical from the mathematical point of view. In recent years, it played a major role in applications of Mathematical and Computational Finance. It can also be applied to Monte Carlo simulations, and therefore it is used in computation of the Greeks.

The finite element method is a numerical method to solve partial differential equations (PDE). This method is entirely deterministic. Although the system of difference equations can be usually solved quickly in low dimensions, the method is not suitable to compute the Greeks that are not directly related to the derivatives computed in the PDE. Cases where it can be applied successfully are the calculation of delta and gamma. In other cases it involves increasing amounts of recalculations which can be cleverly reduced in certain cases.

The main aim of this study is give the details of the three methods given above and compute the Greeks of European options in Black-Sholes environment via Malliavin Calculus. The secondary aim is to compare the methods using the Monte Carlo Simulation Method.

## KAYNAKLAR

- [1] Henao, W. (2005), Malliavin Calculus in Finance: Review of Faster Methods for the Calculation of Sensitivities, Mathematics for Finance Practitioners seminar.
- [2] Higa A. K. and Montero, M. (2004), Malliavin Calculus in Finance, Handbook of Computational and Numerical Methods in Finance p: 111-175.
- [3] L'Ecuyer P. and Perron G. (1994), On the convergence rates of IPA and FDC derivative estimators. Operations Research 42, p: 643-656.

## ABSTRACT

### COMPARISON OF DIFFERENT METHODS TO COMPUTE THE GREEKS

In this study, first we give a basic introduction to Malliavin Calculus and its usage within the area of Monte Carlo simulations in finance. Instead of giving an overview of the theory, we focus on a single application, that is, the computation of the Greeks in the Black-Scholes environment and their Monte Carlo simulation. We then compare the results with the Finite Difference Method and Finite Element Method.

**Key Words:** Malliavin Calculus, Greeks, Monte Carlo methods, Black-Sholes model, integration by parts formula.

## RISK CAPITAL ALLOCATION: IMPORTANCE OF COPULA SELECTION

Uğur KARABEY

Department of Actuarial Sciences Hacettepe University 06800-Beytepe, Ankara  
ukarabey@hacettepe.edu.tr

### 1. Introduction

Financial companies have to hold cash reserves as a cushion against unforeseen losses. This reserve is called risk capital. In order to determine the risk capital of a company (possibly multi-line) we first need to define the aggregate loss of the company. Assume that  $X_i$  represents the loss of line (or sub-portfolio)  $i$  then the aggregate loss can be defined as  $X = \sum_i X_i$ . We here do not make any assumptions, such as independence

or identical distribution on  $X_i$ 's. They are hardly independent or identically distributed in real life problems; contrarily it is easy to observe dependence between risks in any insurance portfolio. For example, different business lines are exposed to the common factor risks such as interest-rates, inflation, economic crises etc. or a catastrophic event can have big effects on many different business lines. Thus, both the marginal distributions and dependency structure between them needs to be studied carefully to determine risk capital  $\rho(X)$  of the company. In real life, the marginal distributions of individual business lines are easy to obtain whereas the joint distribution of these are unavailable in most cases. Therefore, we need tools to construct the joint distribution of the company with using these known marginals. The most effective tool for this kind of task is a copula.

On the other hand, risk capital allocation is mainly required for internal risk management in financial companies[1, 2].

The purpose of this study is to analyze the importance of copula selection in internal risk management. For a stock portfolio, risk contributions are determined for different choices of copulas.

### 2. Risk Measures

Many risk measures exist in the literature. In the following commonly used risk measures are defined. Value at Risk (VaR) measures the maximum potential loss of a given portfolio over a prescribed holding period at a given confidence level  $\alpha$  where  $\alpha \in (0,1)$  and it can be described as,

$$\text{VaR}_\alpha(X) = -\inf\{x \in \mathbb{R} : P(X \leq x) > \alpha\}. \quad (2.1)$$

$$\text{ES}_\alpha(X) = -E[X | X \leq -\text{VaR}_\alpha(X)] \quad (2.2)$$

is called expected shortfall (ES) at level  $\alpha$  ( $\alpha$  usually close to 1) which is defined as an average of VaRs of  $X$  at level  $\alpha$  and higher. Expected shortfall belongs to the family of coherent risk measures and it provides better approach for risk management. It is more sensitive to the shape of the loss distribution and it counts tail of the loss distribution completely. For more details, see [2]. Further risk measures are also considered.

### 3. Risk Capital Allocation

There are many motives behind the risk capital allocation. Firstly, by comparing different losses on capital for each component, it is often possible to answer if a component is worth to keep or not. Secondly, since capital is defined as a risk measure of whole company, one can assess the riskiness of each component's position by splitting this capital, and compare one to another. Another motive is, allocation provides a useful device for

assessment of performance of managers, which can be linked to their compensations. Last but not least, insurers may want to use the allocation in pricing [2, 3, 4]. Many allocation methods are occupied in this study.

#### 4. Scenario and a Simulation Study

In order to show how theory works, we used a stock portfolio of five companies and data covers the period from February 2003 to February 2013. We assume that given total wealth is £1,000,000 which is equally weighted to sub-portfolios. We consider a one-period framework; therefore between time 0 and T no trading is possible. Note also that we study 1 year time horizon, thus  $T=1$ . We assume 'risk' to be given by a random variable  $X$  representing a cash flow at time T. Precisely, the corresponding loss variable at time T can be defined by  $X = X(T) - X(0)$ . Individual stock returns are modelled by geometric Brownian motions (GBM), i.e. price processes  $X_i(i = 1, \dots, 5)$  with

$$X_i(t) = X_i(0) \exp\left(\left(\mu_i - \frac{\sigma_i^2}{2}\right)t + \sigma_i \sqrt{t} Z\right), t \geq 0 \quad (4.1)$$

where  $\mu_i$  is the drift,  $\sigma_i$  is the diffusion coefficient,  $X_i(0)$  is the price of the  $i$ -th asset at time 0 and  $Z$  is a standard normally distributed random variable.

#### REFERENCES

- [1] J. Dhaene, A. Tsanakas, E. A. Valdez, and S. Vanduffel, (2010), Optimal capital allocation principles, Journal of Risk and Insurance, Forthcoming.
- [2] U. Karabey, (2013), Risk Capital Allocation and Sensitivity Analysis in Non-Life Insurance Companies, Manuscript submitted for publication.
- [3] M. B. Neil, (2007), Capital allocation by percentile layer. Enterprise Risk Management Symposium, Society of Actuaries.
- [4] A. E. Valdez and A. Chernih, (2003), Wang's capital allocation formula for elliptically contoured distributions. Insurance: Mathematics and Economics, 33:517-532.

**Key Words:** risk capital allocation, copula, risk measures

#### AN EARLY WARNING MODEL FOR INSURANCE COMPANIES IN TURKEY

Gizem OCAK, Doc. Dr. Sevtap SELCUK-KESTELDr. Ahmet GENC

Midde East Technical University Institute of Applied Mathematics, -Undersecretariat of Treasury

gizem.ocak@metu.edu.tr ;skestel@metu.edu.tr; ahmet.genc@hazine.gov.tr

Insurance companies have responsibilities both to their policy holders and state which are framed by national and in the last decade also by the international regulations. In policyholders' aspect, the insurance company should have adequate amount of reserve to fulfill the demand resulting from the claims, and attaining the certain level of financial stability for its existence in the insurance sector. The quantification of this liability is the main concern of Solvency II regulations. There are many methods concerning measuring the liability of an insurance company to be solvent. One of them is to use financial ratios in order to demonstrate weak and strong side of company. This method can be used for comparing market average ratios by the ratios set by the state and federal laws. Financial ratios also can be used in some models as random variables to give signals of company's future as an early warning system. In addition, early preventions and precautions are also required in order the company to be solvent. It is important to emphasize that ratios vary with respect to the type of insurance policy and the benefit promised. Hence, the choice of financial ratio as determinant is considerable to make judgments on the insolvency pattern.

The aim of this study is to determine the solvency of a non-life insurance company based on Turkish insurance sector data. An earlier study done by Genc (2006) is taken as the guiding literature to re-evaluate the existing situation based on recent development. Furthermore, logit model is used to examine financial failure by Isseveroglu and Guenme (2009). The technical and financial balance sheets, income statements of companies in Turkish insurance sector are taken into account. The ratios derived are evaluated by using the discriminant analysis and a comparative analysis of different linear model approaches are applied to select the best model which should be taken as an early warning model. In the selection of the model accuracy a differentiation grade on the failure or non-failure is implemented. The data set contains financial details of 59 companies and during 1994-2011. The ratios will be chosen with respect to the liquidity, the profitability and the credibility criteria. An accuracy study on the selected model is implemented by using an appropriate optimization model based on the Solvency II requirements. Since financial details of company are constructed of ratios, constraints are determined according to ratio basis. Risk-free rate refers to the rate which firm used in technical provisions and interest rate refers to currently reel rate of return of market at that time period. The aim is to find the parameters which minimize the squared deviation between the realized and estimated ratios based on interest rate, asset liability ratio, risk management ratio, size of the firms, credibility ranking of the company the year earlier, and so on. The optimization process will enable us determine how much load has to be added to each constraints taken into account.

As mentioned above, guiding literature study done by Genc (2006) contains significant results for Turkish insurance sector. The model with five ratios is selected as the best alternative one. 15 of 46 were determined as insolvent companies according to the results of the methodology chosen for the period covering 2004. The most noteworthy and useful result is the ratio which serves as a capability of claim payment explains dependent variable even solely as a random variable. This study upgrades the results by using alternative statistical methods to show the impact of latest developments in insurance sector, changes in the economic indicators and the government regulations.

#### KAYNAKLAR

- [1] Genç, A. (2006), Sigorta Şirketlerinin Mali Analizi, Erken Uyarı Sistemi ve Derecelendirme, Ankara.
- [2] İşseveroğlu G., Gücenme Ü. (2009), Early Warning Model With Statistical Analysis Procedures In Turkish Insurance Companies, Bursa.



[3] Sharpe G. I., Stadnik A. (2006) , A Statistical Early Warning Model Of Financial Distress In Australian General Insurers, Sydney.

#### ABSTRACT

#### AN EARLY WARNING MODEL FOR INSURANCE COMPANIES IN TURKEY

In Turkey, insurance companies have some obligations to be solvent and for this reason they are regularly audited according to defined constraints. There are some models and methods in order to analyze company performances. One of them is an early warning model that is constructed by using some financial ratios. The aim of the study is to determine how Solvency requirements affected the financial stability of Turkish insurance companies last years. The proposed model takes into account the financial ratios related to liquidity, profitability, and other factors regarding to the country specific properties. Historical data on companies' financial indicators are evaluated based on comparative linear model estimation methods to determine the company's financial position which functions as an early warning indicator. To determine the adequacy of the estimators, an optimization model is run to quantify the impact of each factor, such as, interest rate, asset liability ratio, market indicators and operational risk indicators on the estimated model.

**Key Words:** Discriminant analysis, logit regression, non linear optimization, Solvency II

#### MARKOV AGEING APPROACH IN TURKISH MORTALITY ESTIMATION

Fulya AKAT, Doc. Dr. Sevtap SELCUK-KESTEL

Middle East Technical University,  
Institute of Applied Mathematics,  
Financial Mathematics, Ankara-TURKEY  
fulya.akat@metu.edu.tr

Middle East Technical University,  
Institute of Applied Mathematics,  
Actuarial Science, Ankara-TURKEY  
skestel@metu.edu.tr

#### Introduction

The improvement of medical technology, life standards cause higher expected life time and lower mortality rates leading to increasing population in older ages. Turkey like other countries is also exposed to low fertility and longer life time. Mortality rates are the key components to make inferences and predictions on future developments in the population. The well known models in estimating the mortality rates are DeMoivre, Gompertz, Makeham and Weibull which depends on the historical realizations in the population. Additionally, recent developments, such as Lee-Carter (Pitacco, 2003) model yield mortality rates by using self- generating set of iterations when the past information on the population does not fit to the models above.

Mortality projections should represent the change in future's expected life time, for this reason, their precision play an important role in actuarial modeling of life insurance and pension funds (Vaupel, 1998). The most commonly used mortality tables in Turkey are CSO80, SGK 2008 and TRSH 2010. This study questions how much information is gained in measuring the mortality risk when a Markov Aging approach is modified for estimating the mortality risk in Turkey. The properties of this model are having the heterogeneity, resembling the same structure of the rectangular shape in the survival function and assigning the higher mortality rates in young ages of the population.

#### 2. Markov Aging Approach

A finite-state continuous-time Markov process is to model ageing process where each age is represented by the states (Lin and Liu 2007). The transition from one state to another is the indication of aging and physiological distortion. The main advantage of this approach is that it gives time until death time with phase-type distribution and it is easy to add longevity risk factor in this model. The parameters in the transition matrix are estimated by minimizing the sum of weighted square errors:

$$F = \sum_{x=0}^w (q_x + \hat{q}_x)^2 S(x) \quad (1)$$

where  $q_x$  is the observed death rate,  $S(x)$  denotes the survival probability at age  $x$ , and  $\hat{q}_x$  is the corresponding model value for  $q_x$ . The Markov process consists of a set of transition state and a single absorbing state which represent death. The time of staying at state  $i$  before moving to state  $i+1$  follows an exponential distribution with mean  $1/\lambda_i$ . The time until absorption is defined as  $\tau = \inf\{t \geq 0 | X_t = n + 1\}$  is said to have phase-type distribution (Faddy, 1995). The survival function  $\hat{S}(x)$  having the survival time is the time of death or absorption. Because of that the survival time follows a phase-type distribution which represented as  $(\alpha, \Lambda)$

$$\hat{S}(x) = \alpha \exp(\Lambda x) e \quad (2)$$

where  $\Lambda$  is the transition matrix, estimates  $S(x)$  given in Equation 1. The initial distribution is  $\alpha = (1, 0, \dots, 0)$  and  $e$  is the column vector of ones.

The graphs of survival function of mortality table for Turkish population have rectangular shape for both male and female. Rectangular shape of survival function indicates that the life expectancy has a great tendency to respond to a change in the force of mortality. Additionally, mortality rates of the young ages behave differently comparing with other age states. The data used in this study is collected from TUIK (Turkish Statistical Institute). TUIK has published periodic demographic data set for each year since 2007. Mortality tables which

are TRSH2010, CSO80 and SGK 2010 are also used for analyzing how much they represent accuracy current mortality rates in Turkey.

#### KAYNAKLAR

- [1] Vaupel, J.W.(1998), Demographic Analysis of Aging and Longevity, Germany, American Economic Association.
- [2] Pitacco, E. (2003), Survival Models in Actuarial Mathematics: From Halley to Longevity Risk, Italy, Invited lecture, 7th Int. Congress Insurance: Mathematics and Economics.
- [3] Faddy, M.J.(1995), Phase-Type Distribution for Failure Times, Mathematical Computing Modeling, Great Britain.
- [4] Lin,X.S. and Liu, X. (2007) Markov Aging Process and Phase-Type Law of Mortality, Toronto, University of Toronto.

#### ABSTRACT

##### MARKOV AGEING APPROACH IN TURKISH MORTALITY ESTIMATION

Turkey is exposed to the longevity risk despite having young population and this has an enormous impact on pension liabilities. As the accumulation from early retirement and other country specific reasons cause financial deficit on Turkish Government budget, the actuarial valuation of the pension system does not account longevity risk which causes more uncertainty. This study aims to introduce Markov Aging Model to estimate the mortality rates and compared to the mortality tables commonly used in Turkey and the valuation of longevity risk is done based on the proposed model. The parameters optimizing the best mortality estimates are determined and the valuation on a hypothetical pension portfolio is performed.

**Key Words:** Markov Ageing, Longevity, Turkish Pension System

##### OPTIMAL RETENTION FOR PROFIT MAXIMIZING UNDER VaR LEVELS CONSTRAINTS

Murat BÜYÜKYAZICI  
Hacettepe University, Faculty of Science,  
Department of Actuarial Sciences, 06800 Beytepe, Ankara, TÜRKİYE  
muratby@hacettepe.edu.tr

#### Introduction

When an insurer seeks reinsurance protection, the insurer is faced with the classic trade-off between the retained loss and the reinsurance premium. If the retention is small, then it is expected to be low retained loss to the insurer but higher premium payable to the reinsurer. On the other hand, if the retention is large, then it is expected to be large retained loss but lower cost of the reinsurance premium. This implies that the insurer is faced with a problem of determining the optimal retention.

Using well known financial risk measures such as the value-at-risk (VaR), and the conditional-tail-expectation (CTE), Cai and Tan [1] calculated the optimal retention for stop-loss reinsurance under the expected value premium principle. Subsequently; Tan, Chengguo and Zhang [2] extended Cai and Tan's [1] results in two directions. One is to expand the class of reinsurance contracts by considering the quota-share reinsurance, in addition to stop-loss reinsurance. The other is to examine the optimality of the stop-loss and the quota-share reinsurance under many other premium principles. As a result, they have expressed "Because of the complexity of the optimization problem for the stop-loss reinsurance, there are a few premium principles for which we are unable to determine analytically if the optimal reinsurance exist or not". One of these premium principles is standard deviation principle. Computer simulation is frequently used in evaluating such a complex optimization case [3].

In practice, an insurer not only is concerned with reducing risk exposure, but is also interested in maximizing the profit [2]. We incorporated both of these goals into the optimal reinsurance models by maximizing profit under different VaR measure constraints. We also obtained efficient frontier to describe the relationship between the insurer profit and the VaR of the insurer total cost.

In our stochastic model of a non-life insurance company's profit, the stochastic nature of the aggregate loss  $X$  is incorporated by a non-negative random variable has exponential, Pareto and lognormal distributions with mean  $E(X) = 1000$ . Our objective is to maximize the profit of insurer  $S$ , for obtaining efficient frontier to describe the relationship between the insurer profit and the  $VaR_T$  risk measure. Under the stop-loss reinsurance model, with different  $VaR_T$  risk level constraints, the optimal retention  $d^*$  is the solution to the following optimization problem:

$$\max_{d \in (0, \infty)} E[S(d, \varepsilon)] \quad (1)$$

In efficient frontier analysis, a suitable lowest bound of  $VaR_T$  is needed to be determined by solving the following optimization problem:

$$\min_{d \in (0, \infty)} E[VaR_T(d, \alpha, \varepsilon)] \quad (2)$$

The resulting optimal retention  $d^*$  ensures that  $VaR_T$  is minimized for a given confidence level  $1 - \alpha$ .

We propose practical solutions for the determination of optimal retentions in stop-loss reinsurance under VaR risk measure using the Crystal Ball 11.1.2 package program which incorporates OptQuest simulation optimizer. The minimum values of  $VaR_T$  used in efficient frontier analysis are calculated by solving (2) and optimal retentions maximizing insurer's profit for each test points are calculated by solving (1) with simulation optimization. We compare premium principles for three distributions with respect to the insurer's profit for the

same risk level. We also compare the distributions for two premium principles with respect to the insurer's profit for the same risk level. For example, under expected value premium principle, for almost all risk levels Pareto distribution gives the highest profit and exponential distribution gives the lowest. As the risk level increases profits for exponential and lognormal distributions increase more rapidly than that for Pareto as seen in Figure 1.

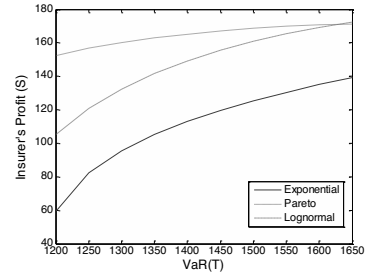


Figure 1. Comparison of distributions for expected value premium principle

Computational results show that the assumed loss model and premium principle used are effective factors for determining optimal retention limit, so the insurer's expected profit. Finding analytically optimal solution to some complex real-world problems is extremely hard and even sometimes impossible. Therefore, simulation optimization approach can be used to find optimal retention limit for better profit with their acceptable risk level.

#### REFERENCES

- [1] J. Cai, S.K. Tan, Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measures, *ASTIN Bull.* 37 (1) (2007), 93-112.
- [2] K. S. Tan, W. Chengguo, Y. Zhang, VaR and CTE criteria for optimal quota-share and stop-loss reinsurance, *North Amer. Actuarial J.* 13 (4) (2009), 459-482.
- [3] E. Tekin, I. Sabuncuoglu, Simulation optimization: A comprehensive review on theory and applications. *IIE Transactions*, 36 (2004), 1067-1081.

## BİLDİRİ OTURUMLARI 5

### SESSION 5

#### İstatistik Teori 2

## MEME KANSERİ VERİSİNİN BAYESÇİ SAĞKALIM ANALİZİ İLE İNCELENMESİ

Esin AVCI\* Nural BEKİROĞLU Meral YAY

Türkiye İstatistik Kurumu, 34353, İstanbul, TÜRKİYE, esinavci@hotmail.com

Marmara Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim ABD, 34668, İstanbul, TÜRKİYE,  
nural@marmara.edu.tr

Mimar Sinan Güzel Sanatlar Üniversitesi, Fen Edebiyat fakültesi, İstatistik Bölümü, 34433, İstanbul,  
TÜRKİYE, myay@msgsu.edu.tr

### Bayesci Sağkalm Analizi

Sağkalm en kısa tanımıyla yaşamda kalma süresidir. Daha genel bir ifade ile belirli bir başlangıç değerinden (doğum, tedavi başlangıcı veya işverenin iş vermesinden) belirli bir olguya ulaşmaya (ölüm, hastalığın nüks etmesi veya işten çıkarılma) kadar geçen süre olarak tanımlanır. Sağkalm verilerinin analizi ve modellenmesi, 17. yy da aktüerya ve demografinin gelişmesiyle şekillenmeye başlamıştır. Sağkalm analizi konusundaki ilk önemli gelişme, Kaplan ve Meier'in 1958 yılında yayımlanan "sağkalm eğrisinin tahmini" makalesiyle kendini göstermiştir (Kaplan ve Meier, 1958). Bu çalışmanın ardından, sağkalm süresine etki eden risk faktörlerinin belirlenmesi amacıyla geliştirilen yeni yöntemler araştırmacıların ilgisini çekmiş ve araştırmaların sayısı gün geçtikçe artmıştır. Son yıllarda yapılan araştırmalarda Bayesci sağkalm analizi üzerinde odaklanılmıştır..

Bayesci yaklaşım, Bayes teoremine dayanarak geliştirilmiştir ve yaklaşıma göre parametreler klasik yaklaşımdaki gibi sabit olarak değil, olasılığa bağlı olarak tanımlanır. Dolayısı ile her bir parametreye ilişkin bir dağılım söz konusudur. Ayrıca veriyi modellemede önsel (prior) bilgiye başvurma esnekliği nedeniyle klasik yöntemlere göre oldukça avantajlıdır (İbrahim ve diğ., 2001; Wong ve diğ., 2005). Önsel bilginin elde edilmesi, Bayesci çıkarsamada önemli rol oynar. Önsel dağılımın sonuçları ne kadar değiştireceği model seçim kriterleri ile tespit edilmelidir. Önsel dağılımlar, açıklayıcı olan (informative) ve açıklayıcı olmayan (noninformative) olmak üzere iki temel gruba ayrılır. Bayesci yaklaşımda, denklem (1)'de de gösterildiği gibi önsel dağılımdan ve veriden gelen bilgi birleştirilerek sonsal dağılım elde edilir. Sonsal bilginin elde edilmesindeki en önemli sorun, dağılımının hesaplama güçlüğüdür. Bu nedenle Markov Zinciri Monte Carlo (MCMC) yöntemleri yardımıyla oldukça kolaylaşmaktadır (Gilks ve diğ., 1996). Yöntem ücretsiz olarak indirilebilen WinBUGS paket programı yardımıyla kolaylıkla uygulanmaktadır. BUGS (Spiegelhalter ve diğ., 1996), özel olarak MCMC yönteminin tam olasılık modellerine uygulanması için kullanılan bir programdır.

$$f(\theta|y) \propto L(y|\theta)f(\theta) \quad (1)$$

Burada  $f(\theta)$  önsel dağılımı,  $L(y|\theta)$  olabilirlik fonksiyonunu ve  $f(\theta|y)$  sonsal dağılımı ifade etmektedir.

Bayesci sağkalm analizi, (1) denklemi ile tanımlanan sonsal dağılım denklemine (2) denklemi ile tanımlanan sağkalm süresine ait olabilirlik fonksiyonunun  $f(\theta)$  parametrelere ait önsel dağılımların birleştirilmesi olarak tanımlanabilir.

$$L(\theta|D) = \prod_{i=1}^n f(t_i|\theta)^{v_i} S(t_i|\theta)^{(1-v_i)} \quad (2)$$

Burada  $t = (t_1, \dots, t_n)'$ 'nin sağkalm zamanını,  $v = (v_1, \dots, v_n)'$ 'nin durum değişkenini ve  $D = (n, y, v)$ 'nin gözlenen verileri göstermektedir.

Bu çalışmada, açıklayıcı olan ve olmayan önsel bilgiye dayalı Bayesci Sağkalm Analizi (sırasıyla BSA-I ve BSA-II olarak tanımlanmaktadır) İstanbul Üniversitesi Tıp Fakültesi Meme Cerrahi Kliniğinde tedavi gören 9 yıl izlenli 458 meme kanseri hastasının sağkalm süresine yaş, aile hikayesi, menepoz durumu, tümör büyüklüğü, TNM evresi, histolojik tip, östrojen reseptör durumu, progesteron reseptör durumu, tedavi türü (radyoterapi ve hormon terapisi) ve aksiller lenf nodu tutulumu prognostik faktörler olarak belirlenmiştir.

Ölüm gerçekleşen 69 hastanın sağkalm sürelerine ait dağılım incelenmiş ve Lognormal dağılıma sahip olduğu olasılık yoğunluk, Hazard fonksiyonları grafiği ve verilerin logaritmaları alınarak normallik testi yardımıyla saptanmıştır.

Sağkalm süresine etki eden prognostik faktörler için açıklayıcı önsel dağılımların elde edilmesi için daha önce yapılan araştırmalar incelenmiş ve bilgiler derlenmiştir. Açıklayıcı olmayan önsel dağılım için  $N(0,0, 0,0001)$  kullanılmıştır. Bu bilgiler doğrultusunda WinBUGS'ta macro yazılmıştır. Sonsal dağılımın yakınsama sağlanması ve başlangıç değerlerinin etkisinin en aza indirilmesi için, Markov zincirinde elde edilen örneklemin başlangıç kısmı çıkarılmıştır. Bu amaçla 5000 iterasyonun ilk 2500 iterasyonu çıkarılmış ve elde edilen sonuçlarda MC hatanın, sonsal standart hataya oranı %5'ten küçük olmasından dolayı yeterli iterasyon sayısına ulaştığı görülmüştür. Parametrelerin iz grafikleri salınım gösterdiğinden sonsal dağılıma yakınsama hızının yüksek olduğu saptanmıştır. Sağkalm süresine etki eden (anlamlı bulunan) parametrelerin güven aralıkları "0" değerini içermemesine göre saptanmaktadır. Bu bağlamda tümör büyüklüğü, TNM evresi IIA, TNM evresi IIB, östrojen reseptör durumu ve aksiller lenf nodu tutulumu değişkenlerinin sağkalm süresi üzerindeki etkilerinin önemli olduğuna hem BSA-I'in hem de BSA-II'de saptanmıştır. AIC kriterlerine göre yapılan karşılaştırma sonucunda, BSA-I'in BSA-II'den daha küçük değere sahip olduğu ve dolayısı ile tercih edilmesi gerektiğine karar verilmiştir. Ayrıca BSA-I yöntemi ile elde edilen parametre tahminlerine ilişkin standart hataların, BSA-II için elde edilen standart hatalardan daha küçük olduğu gözlenmiştir.

### KAYNAKLAR

[1] Gilks, W.R., Richardson, S., Spiegelhalter D.J. (1996), Markov Chain Monte Carlo in Practice. Chapman-Hall, London.

[2] İbrahim, J.G., Chen, M.H., Sinha, D. (2001), Bayesian Survival Analysis, Springer-Verlag, New York.

### ABSTRACT

#### EXAMINED BREAST CANCER DATA BY BAYESIAN SURVIVAL ANALYSIS

Breast cancer is one of the most common cancer and the leading death cause among women in the world. The aim of this study was to define prognostic factors of breast cancer using Bayesian survival models.

**Key Words:** Bayesian survival analysis, Lognormal distribution, Breast cancer, WinBUGS

## MCMC YÖNTEMİNİN DİNAMİK DOĞRUSAL MODELLERE UYGULANMASI

Hatice Yağmur GÜRKAN, Gül ERGÜN\*

Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800, Ankara, TÜRKİYE  
h.yagmur.gurkan@gmail.com, gul@hacettepe.edu.tr

Bu çalışmada, üç farklı dinamik doğrusal modelin R'da dlm paketi kullanılarak işletilmesi ele alınmıştır. Bu modeller: durgun model, doğrusal büyüme modeli ve mevsimsel etkili ikinci dereceden polinomiyal model olmak üzere birleşik modeldir. Çalışmada ençok olabilirlik yöntemi, ileriye doğru filtreleme geriye doğru örnekleme algoritması ve/veya Gibbs örnekleme yöntemi incelenen her bir modele uygulanmıştır; hata terimlerine ait bilinmeyen gözlem ve sistem varyans tahminlerinin yanı sıra Kalman filtresi sonuçları elde edilmiştir. Çalışmada ele alınan her model için aşağıdaki adımlar uygulanmıştır. Bunlar:

1. Bilinmeyen varyansların ençok olabilirlik tahminlerinin elde edilmesi,
  - 1.1. Ençok olabilirlik tahminlerinin standart hataları düşük ise, bu tahminlerin modelde kullanılması ve Kalman filtresi sonuçlarının elde edilmesi,
    - 1.1.1. Filtre sonuçlarının ileriye doğru filtreleme geriye doğru örnekleme algoritmasının bir parçası olarak kullanılması ve geriye örnekleme kapsamında dlmBSampler fonksiyonu kullanılarak yeni verilerin türetilmesi,
      - 1.1.2. Türetilen bu verilerin Gibbs örnekleme için bir kaynak olarak kullanılması ve bilinmeyen varyansların Bayesci tahminleri olan Monte Carlo tahminlerinin elde edilerek analizin sonlandırılması ve sonuçların özetlenmesi.
    - 1.2. Ençok olabilirlik tahminlerinin standart hataları yüksek ise, Gibbs örnekleme kullanılarak bilinmeyen varyansların Monte Carlo tahminlerinin elde edilmesi,
      - 1.2.1. Bu tahminlerin modele yerleştirilerek Kalman filtresi sonuçlarının elde edilmesi,
      - 1.2.2. Kalman filtresi uygulaması sonrası analizin sonlandırılması ve sonuçların özetlenmesi.

Bu özetle sadece durgun modele ait sonuçlar yer almaktadır. Dinamik doğrusal modellerin en yakın hali olan durgun model  $\{F=1, G=1, V, W\}$  dörtlüsü ile aşağıdaki gibi ifade edilir:

$$y_t = \mu_t + v_t, \quad v_t \sim N(0, V) \quad (1)$$

$$\mu_t = \mu_{t-1} + w_t, \quad w_t \sim N(0, W) \quad (2)$$

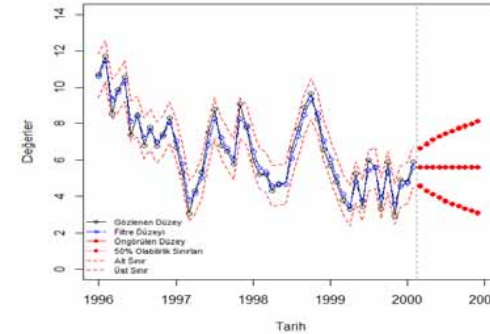
Uygulamada, durgun model R'da tanımlanırken  $\mu_0 \sim N(10, 2)$  alınmış ve durgun modelden üretilmiş yapay bir seri üzerinde yukarıda tanımlanan adımlar izlenerek tahminler yapılmıştır.

**Çizelge1. Durgun model için ençok olabilirlik tahminleri**

PARAMETRE	ENÇOK TAHMİNİ	OLABİLİRLİK	STANDART HATA
V	0.6215211		0.3658219
W	1.285503		0.5667107

Ençok olabilirlik tahminleri kullanılarak elde edilen durgun modele Kalman filtresi uygulanmış ve sonuçları Şekil 1'de verilmiştir.

Gibbs örnekleme için işletilebilmesi için bilinmeyen varyansların önsel dağılımlarının belirlenmesi ya da bu parametrelere birer önsel dağılım atanması gerekmektedir. Bu çalışmada, d-ters Gamma modeli bilinmeyen varyansların önsel dağılımları için ele alınmış ve önsel dağılımların parametreleri Petris'in [2] çalışması referans alınarak belirlenmiştir. Gibbs örnekleme için 10000 iterasyonda çalıştırılarak ilk 5000 değer yakılmış ve elde edilen Monte Carlo tahminleri Çizelge 2'de verilmiştir.



**Şekil 1. Durgun Modele Ait Kalman Filtresi ve Öngörü Sonuçları Grafiği**

**Çizelge2. Durgun model için Monte Carlo tahminleri**

PARAMETRE	MONTE CARLO TAHMİNİ	STANDART HATA
V	1.64E-01	5.24E-05
W	5.01E-02	1.38E-04

Çalışmanın sonucunda, ele alınan modeller için Bayesci yaklaşımla elde edilen Monte Carlo tahminlerinin ençok olabilirlik tahminlerine göre daha düşük standart hatalara sahip oldukları görülmüştür. Buna göre bilinmeyen varyanslar söz konusu olduğunda, Kalman filtresinin kullanılması için daha düşük standart hatalı tahminler veren Bayesci yaklaşımın tercih edilmesi daha yerinde bir tercih olacaktır.

### KAYNAKLAR

- [1] Petris, G. (2010), An R Package for Dynamic Linear Models, Journal of Statistical Software, 36(12), 1-16.
- [2] Petris, G., Petrone, S., Campagnoli, P. (2009), Dynamic Linear Models with R. Springer-Verlag.
- [3] West, M., Harrison, J. (1997), Bayesian Forecasting and Dynamic Models, Springer.

### ABSTRACT

#### AN APPLICATION OF MCMC METHOD TO DYNAMIC LINEAR MODELS

This study deals with implementing three different types of dynamic linear models using dlm package in R. These models are the steady model, the linear growth model and the combined model as a second order polynomial model with a seasonal effect. Several methods such as maximum likelihood, forward filtering backward sampling and/or Gibbs sampling are applied for each model considered in the study. Datasets are generated for the first two models; Turkey Cost of Living Index (Wage Earners) series is used for the combined model. The estimations of the unknown variance components besides, the Kalman filter results are obtained.

**Key Words:** Dynamic Linear Models, Bayesian Inference, Kalman Filtering, Forward Filtering Backward Sampling, Gibbs Sampling, R, dlm.

## OECD ÜLKELERİNİN ENERJİ PERFORMANSLARINI BAYESCI STOKASTİK SINIR ANALİZİ İLE DEĞERLENDİRMESİ

Mehmet Ali CENGİZ\*, Naci MURAT\*\*, Talat ŞENEL\*\*\*

\* Ondokuz Mayıs Üniv., Fen-Ed.Fak., İstatistik Böl., Samsun, TÜRKİYE E-mail: macengiz@omu.edu.tr

\*\* Ondokuz Mayıs Üniv., Endüstri Mühendisliği, Samsun, TÜRKİYE E-mail: nacimurat@omu.edu.tr

\*\*\* Ondokuz Mayıs Üniv., Fen-Ed.Fak., İstatistik Böl., Samsun, TÜRKİYE E-mail: tlsenel@omu.edu.tr

Stokastik sınır analizi (SSA) sınır fonksiyonlarını tahmin etmek ve üretim etkinliğini ölçmek için sıkça kullanılan bir yöntemdir. SSA ekonometrik yöntemleri kullanan parametrik bir yaklaşımdır. Stokastik sınır yaklaşımı, maliyet, kâr ve üretim gibi açıklanan değişkenlerle; girdi, çıktı ve çevresel faktörler gibi açıklayıcı değişkenler arasında bir ilişki kurmakta ve hata terimini göz önüne almaktadır.

Bir modelin stokastik sınır üretim fonksiyonu aşağıdaki gibi ifade edilmektedir.

$$\ln(y_i) = x_i\beta + v_i - u_i \quad , i=1,2,\dots,n$$

$$\beta = (\beta_0, \beta_1, \dots, \beta_K)'$$

Burada  $\beta = (\beta_0, \beta_1, \dots, \beta_K)'$  ((K+1)x1) boyutlu bilinmeyen parametre vektörü ve  $y_i$ , i. karar verme biriminin çıktısını temsil eder.  $x_i$ , (K+1) lik bir satır vektördür, bu vektörün ilk elemanı "1" dir ve geri kalan elemanları i. karar verme birimi tarafından kullanılan K tane girdi miktarının logaritmasıdır.  $u_i$  ise negatif olmayan rastgele değişken olup karar verme birimlerinin teknik etkinsizliğiyle ilişkilidir. Son olarak  $v_i$  ise istatistiksel gürlüğü açıklamak için rastgele hatadır ([1], [2]).

Son yıllarda stokastik sınır analizinde Bayesci yaklaşımın kullanımı sözkonusudur ([4],..., [9]) Bayesci yaklaşımın kullanıldığı alanların başında Enerji sektörü gelmektedir. Enerji tüketimi ulusal ekonominin gelişiminde belirleyici bir faktördür ve ülkelerin gelişmişlik düzeyleriyle doğru orantılıdır. Ekonomik olarak gelişmiş ülkeler ihtiyacı olan enerjiyi kendi kaynaklarını kullanarak üretirler. Enerji tüketimi ve ekonomik büyüme karşılıklı olarak birbirlerini etkiler. Bu çalışmada OECD ülkelerinin enerji tabanlı gelişmişlik performansları Bayesci stokastik sınır analiziyle ölçülmüş ve elde edilen sonuçlar klasik stokastik sınır analiziyle elde edilen sonuçlarla karşılaştırılmıştır.

### KAYNAKLAR

- [1] Aigner, D.J., Lovell, C.A.K., Schmidt, P. (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models", Journal of Econometrics, 6 (1): 21-37.
- [2] Meeusen, W., van den Broeck, J. (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error", International Economic Review, 18 (2): 435-444.
- [3] U.S. Energy Information Administration (2012), "International Energy Outlook", EIA, Washington.
- [4] Koop, G., J. Osiewalski and M.F.J. Steel (1997), "Bayesian Efficiency Analysis Through Individual Effects: Hospital Cost Frontier." Journal of Econometrics, 76, 77-105.
- [5] van den Broeck, J., G. Koop, J. Osiewalski and M.F.J. Steel. (1994), "Stochastic Frontier Models: A Bayesian Perspective." Journal of Econometrics, 61, 273-303.
- [6] Griffin, J. E. and Steel, M. F. J. (2004), "Semiparametric Bayesian Inference for Stochastic Frontier Models." Journal of Econometrics 123, 121-152.
- [7] Kumbhakar, S.C. and E.G. Tsionas. (2005), "Measuring technical and allocative inefficiency in the translog cost system: a Bayesian approach." Journal of Econometrics, 126, 355-384.
- [8] Cengiz, M.A., Şenel, T., Terzi, Y., Murat N. (2012). A Bayesian computation for stochastic frontier analysis, Book of Abstracts, 11th International Conference on Data Envelopment Analysis.
- [9] Cengiz, M.A., and Murat N. (2012). Assessing Convergence Of MCMC Algorithms for Stochastic Frontier Analysis, Book of Abstracts, 11th International Conference on Data Envelopment Analysis

### ABSTRACT

## OECD COUNTRIES' ENERGY PERFORMANCE EVALUATION USING BAYESIAN STOKASTIC FRONTIER ANALYSIS

Starting from seminal study by Aigner et al. (1977), theoretical literature on stochastic frontier has grown vastly. The range of applications of the techniques described is huge. The economic meaning of a frontier is to represent the best-practice technology in a production process or in a particular economic sector. Cost frontiers describe the minimum level of cost given a certain output level and certain input prices. Production frontiers represent the maximum amount of output that can be obtained from a given level of inputs. The gap between the actual and the maximum output is a measure of inefficiency and an important issue in many application fields, such as production studies. More recently, a large amount of interest has been devoted to the use of Bayesian methods for making inference in stochastic frontier models. The Bayesian inference in stochastic frontier was first proposed by van den Broeck et al. (1994). Koop et al. (1997) developed Bayesian inferential procedures to be applied to panel data. Griffin and Steel (2004) adopts a semiparametric Bayesian approach for inference. Kumbhakar and Tsionas (2005), Cengiz and Murat (2012) and Cengiz et al. (2012) introduce the use of MCMC methods in this context.

Energy consumption is a decisive factor in the development of the national economy and it is directly proportional to the development level of the countries. Economically developed countries produce their needed energy using by their own sources. Energy consumption and economic growth mutually influences each other. This study's aim is to measure OECD countries' energy performances using by Bayesian stochastic frontier analysis.

**Key words:** Stochastic frontier, Bayesian, Energy, OECD

## BAYESÇİ YAPISAL EŞİTLİK MODELLERDE FARKLI TİPTE ÖLÇEKLERİN ANALİZİ İÇİN EŞİK DEĞER YAKLAŞIMI

Naci MURAT<sup>1</sup>, Mehmet Ali CENGİZ<sup>2</sup>, Erol TERZİ<sup>3</sup>

<sup>1</sup>Mühendislik Fakültesi, Ondokuz Mayıs Üniversitesi, Samsun, Türkiye, e-mail: nacimurat@omu.edu.tr

<sup>2</sup>Fen Edebiyat Fakültesi, Ondokuz Mayıs Üniversitesi, Samsun, Türkiye, e-mail: macengiz@omu.edu.tr

<sup>3</sup>Fen Edebiyat Fakültesi, Ondokuz Mayıs Üniversitesi, Samsun, Türkiye, e-mail: eroltr@omu.edu.tr

Kantitatif araştırma tekniği pek çok alanda yaygın olarak kullanılmaktadır. Çoğu zaman veriler sıralı kategorik değişkenler şeklinde elde edilmektedir. Bu değişkenlere örnek olarak; tutum ölçekleri, likert ölçekler, derecelendirme ölçekleri gibi ölçekler verilebilir [1,2]. Kategorik olarak toplanan verilerin analizinde, verilere normal dağılımdan çekilmiş sürekli değerler gibi muamele edilmektedir. Bu yaklaşım histogram dağılımlarının simetrik olduğu durumlarda ciddi problemlere yol açmaz. Oysa pek çok araştırmada uç değerlerde birikmeler olur. Bu birikmeler histogramların çarpık olmasına yol açar[3-5]. Bu tür kesikli verilerin analizinde, eşik değerli normal dağılımdan gelen gizli sürekli değişken olarak kabul etme yaklaşımını kullanılmaktadır. Bazı tutumlarla ilgili yapılan sorgulamalarda ölçek 3'lü iken, bazı durumlarda 5'lidir. Eşik değer sayısı sıralı kategorik sayısının bir eksiği kadardır. Farklı ölçeklerin kullanıldığı durumlarda eşik değer sayıları da farklı olacak ve analiz aşaması tıkanacaktır. Literatürde farklı ölçekler için uygulanmış bir eşik değer yaklaşımı bulunmamaktadır.

Bu çalışmada, farklı ölçeklerin kullanıldığı araştırmalarda veri analizi için eşik değerli normal dağılımdan gelen gizli sürekli değişken olarak kabul etme yaklaşımının Bayesci analizi kullanılmıştır. Farklı ölçeklerde lineer dönüşüm tekniği kullanılmış ve ölçeklerden biri diğerine dönüştürülmüştür. Tüm veriler için geliştirilen ortak ölçekten sonra eşik değer yaklaşımı uygulanmıştır. Sıralı kategori sayısının bir eksiği kadar belirlenen eşik değerler ortak bir normal dağılımla ilişkilendirildiği gibi ayrıca parametrelerin kolayca yorumlanmasını sağlamıştır. Analizde sıralı kategorik değerlerin sayısal değerleri sadece ilgili kategoriyi temsil için kullanılmıştır. Veri analizinde bu değerlere ilişkin frekans oranları ile işlem gerçekleştirilmiştir.

**Anahtar Kelimeler:** Yapısal Eşitlik Modelleri, Bayesci Yaklaşım, Kesikli Veri Analizi, Farklı Ölçekler

[1] Likert R. (1992), A Technique for the Measurement of Attitudes, Archives of Psychology,

140: 1-55

[2] Jamieson S. (2004), Likert Scales: How to (Ab)Use Them, Medical Education,

38:1217-1218.

[3] Lee S. Y., Poon W. Y. and Bentler P. M. (1990a), Full maximum likelihood analysis of structural equation models with polytomous variables, Statistics and Probability Letters, 9, 91-97.

[4] Lee S. Y., Poon W. Y. and Bentler P. M. (1990b), A three-stage estimation procedure for structural equation models with polytomous variables, Psychometrika, 55, 45-51.

[5] Lee S. Y., Song X. Y., Skevington S. and Hao Y. T. (2005), Application of structural equation models to quality of life. Structural Equation Modeling – A Multidisciplinary Journal, 12, 435-453.

### ABSTRACT

### TRESHOLD APPROACH FOR ANALYSIS OF DIFFERENT TYPES OF SCALE IN BAYESIAN STRUCTURAL EQUATION MODELS

Quantitative research techniques are widely used in many fields. Most of the time sequential data is obtained in the form of categorical variables. Examples of such variables are attitude items, Likert items, rating scales and the like. For analysis of obtained in the form of categorical variables, one common approach is to treat the assigned integers as continuous data from a normal distribution. This approach may not lead to serious problems if the histogram of the observations is symmetrical. However, many studies are storage extreme

values. This concentration leads to skewed histogram. A better approach for assessing these kinds of discrete data is to treat them as observations that are coming from a hidden continuous normal distribution with a threshold specification. The threshold value is one less than the number of categorical sequential number. Threshold value will be different from the number of cases using different scales and analysis process will get stuck. There isn't a threshold approach is applied for the different scales in the literature.

In this study, a threshold approach for analysis of different types of scales in Bayesian structural equation models was introduced. One likert scale was converted to another scale using linear transformation at different scales. Thus, a common scale was developed.

**Key Word:** Structural Equation Modelling, Bayesian Approach, Discrete Data Analysis, Different Scales

## BAYESÇİ COX REGRESYONDA MCMC YAKINSAMA KRİTERLERİNİN DEĞERLENDİRİLMESİ

Nesrin ALKAN\*, Yüksel TERZİ\*\*, Mehmet Ali CENGİZ\*\*

\* Sinop Üniv., Fen-Ed.Fak., İstatistik Böl., Sinop, TURKEY E-mail:nesrin.alkan@gmail.com

\*\* Ondokuz Mayıs Üniv., Fen-Ed.Fak., İstatistik Böl.,Samsun, TURKEY E-mail:yukselt@hotmail.com

\*\* Ondokuz Mayıs Üniv., Fen-Ed.Fak., İstatistik Böl.,Samsun, TURKEY E-mail:macengiz@hotmail.com

### 1. Bayesci Cox Regresyon Modeli

Cox regresyon, sağkalm analizinin en çok kullanılan yöntemlerindedir. Son zamanlarda yapılan çalışmalar incelendiğinde Bayesci sağkalm analizinin uygulamalarında bir artış gözlenmiştir. Daha önceki çalışmalarda araştırmacılar zor teorisi olması nedeniyle Bayesci analizleri medikal çalışmalarda çok sık kullanmamışlardır. Fakat bilgisayar teknolojisinin gelişmesi ile birlikte Bayesci yaklaşımlara ilgi artmış ve Bayesci Cox regresyon yöntemi geliştirilmiştir. Parametre hakkındaki Bayesci çıkarsama parametrenin sonsal dağılımına dayanmaktadır. Sonsal dağılım, olabilirlik fonksiyonu ile önsel dağılımın çarpımından elde edilir. Bayesci Cox Regresyonda sonsal dağılım tahmin edilirken olabilirlik fonksiyonu olarak kısmi olabilirlik fonksiyonu kullanılır ve parametrenin sonsal dağılımına dayanan çıkarsamada bulunulur. Bayesci Cox Regresyon modelinde sonsal dağılım eşitlik (1)'de verildiği gibidir.

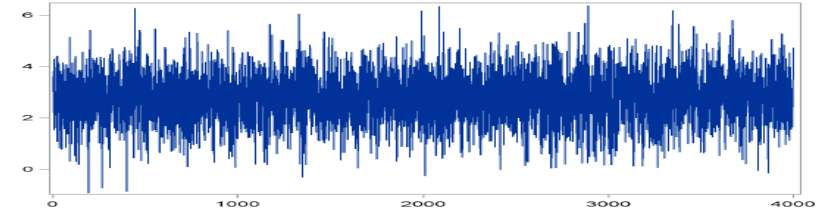
$$P(\beta / y) \propto L_p(y / \beta)P(\beta) \quad (1)$$

Burada  $L_p(y / \beta)$ , regresyon katsayısı  $\beta$  parametrelili kısmi olabilirlik fonksiyonudur ve eşitlik (2)'de verildiği gibidir.

$$L_p(y / \beta) = \prod_{i=1}^k \frac{\exp(\beta' x_i)}{\sum_{j \in R(t_i)} \exp(\beta' x_j)} \quad (2)$$

### 2. Markov Zincirinin Yakınsama Testleri

Markov Zinciri Monte Carlo (MCMC) yöntemi hedeflenen bir sonsal dağılımdan örneklem çekmek ve bu örneklemeleri kullanarak ilgilenilen herhangi bir miktarı hesaplayan iteratif bir yöntemdir. Simülasyona dayalı bu yöntemde dikkat edilmesi gereken iki önemli nokta vardır. Birincisi Markov zincirinin durağan hale veya istenen sonsal dağılıma gelip gelmediğine karar vermek, ikincisi ise Markov zincirinin durağan hale gelmesini sağlayan iterasyon sayısını belirlemektir. Markov zincirinin yakınsaması hedeflenen sonsal dağılıma ulaşılması anlamına gelir. Bayesci analizlerde çıkarsamalar yakınsama sağlanmış sonsal dağılımdan elde edilir. Eğer zincir yakınsamazsa geçerli çıkarsamalar elde edilemez. Bütün parametreler için yakınsamanın kontrol edilmesi gerekir. Yakınsama testleri bu önemli noktada yardımcı olur. Yakınsama testleri hedef dağılımdan çıkarsamaları elde etmek için Markov zincirinin ne kadar uzunlukta çalıştırılması gerektiğini değerlendiren yöntemlerdir. Yakınsamanın değerlendirilmesinde iz grafikleri ile görsel analizi yapılır. İz grafikleri, yakınsamanın değerlendirilmesinde çok kullanışlıdır. İz grafikleri hedef dağılıma yakınsama olup olmadığını gösterir. Zincirde ilerleme devam ederken noktaların dağılımı değişmiyorsa zincir hedef dağılıma ulaşmış sayılır.



Şekil 1. Yakınsama sağlamış bir zincirin iz grafiği

Yakınsamanın değerlendirilmesinde ayrıca Gelman-Rubin, Geweke, Heidelberger-Welch, Raftery-Lewis, Autocorrelation ve Monte Carlo standart hata yöntemleri de kullanılır. Literatürde MCMC yönteminde standart bir yakınsama kriterinin kullanımına ilişkin çalışmalar bulunmamaktadır.

Bu çalışmada Bayesci Cox regresyon modelinin parametreleri için farklı iterasyon sayılı MCMC yakınsama kriterlerinin performansları gerçek bir veri seti kullanılarak karşılaştırıldı. İterasyon sayıları 100'den başlayarak bütün yöntemlerin yakınsama sağladığı noktada durduruldu. Böylece yöntemlerin farklı iterasyon sayılarında nasıl davrandıkları ve birbirleriyle olan benzerlikleri veya farklılıkları ortaya çıkarıldı.

### KAYNAKLAR

- [1] Ibrahim J.G., Chen M.H. and Sinha D. (2001). Bayesian Survival Analysis. Springer-Verlag, s: 290-317, New York.
- [2] Geweke J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments. In Bayesian Statistics 4 (ed JM Bernardo, JO Berger, AP Dawid, and AFM Smith). Clarendon Press, Oxford, UK.
- [3] Heidelberger P. and Welch P.D. (1983). Simulation Run Length Control in the Presence of an Initial Transient, Operations Research, 31, 1109-1144.
- [4] Raftery A.E. and Lewis S.M. (1996). The Number of Iterations, Convergence Diagnostics and Generic Metropolis Algorithms, in W.R. Gilks, D.J. Spiegelhalter, and S. Richardson, eds., Markov Chain Monte Carlo in Practice, London, UK: Chapman & Hall.

### ABSTRACT

#### ASSESSING CONVERGENCE DIAGNOSTICS IN MCMC FOR BAYESIAN COX REGRESSION

The Markov Chain Monte Carlo (MCMC) is the most common method for the Bayesian analysis. The basic idea of MCMC is to generate samples from target posterior distribution and uses these samples to calculate any relevant quantities of interest (Gilks et al, 1996). Researchers have to decide whether the Markov Chain is stable or the chain has reached desired posterior distribution. Additionally they have to specify the number of iterations which keep after the Markov Chain is stable. Convergence diagnostics help to resolve these issues. Convergence diagnostics are methods for assessing how long to run a Markov chain in order to obtain observations from the stationary distribution. Several statistical diagnostic tests can help to assess Markov chain convergence. This study aimed to compare performances of convergence diagnostics test by using real data set.

**Key words:** Convergence criteria, Markov Chain Monte Carlo, Bayesian Cox regression



GEÇİKMESİ DAĞITILMIŞ MODELLERDE BAYESYEN LIU VE RIDGE TİPİ TAHMİN  
EDİCİLERİN KARŞILAŞTIRILMASI

Nimet TÜRKER<sup>1</sup>, Berrin GÜLTAY<sup>2</sup>, Selahattin KAÇIRANLAR<sup>3</sup>

1 Çukurova Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü , 01330, Adana, nturker@cu.edu.tr

2 Onsekiz Mart Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü , Çanakkale,  
berrin\_erkan@hotmail.com

3Çukurova Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü , 01330, Adana, skacir@cu.edu.tr

ÖZET

Gecikmesi sonlu dağıtılmış modeli göz önüne alalım:

$$y_t = \delta + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} + u_t \quad t = p + 1, \dots, T$$
$$= X\beta + u \quad (1)$$

Çoklu iç ilişki problemi olması durumunda, (1) deki  $\beta$  ların en küçük kareler (EKK) yöntemiyle tahmin edilmesi bazı dezavantajlara sahiptir.

Bu problemleri çözmek için  $\beta_i$  üzerinde lineer stokastik olmayan önbilginin geliştirilmesi Almon (1965) tarafından verilmiştir.

Bununla birlikte, Almon metodunun uygulanmasıyla bazı problemlerle yüz yüze kalabiliriz. Bu problemleri gidermek için stokastik ön bilgi altında Ridge tipi tahmin edici yaklaşımı alternatif olarak ele alınmıştır. Bu tahmin edicilerin aynı zamanda Bayesyen tahmin ediciler olduğu gösterilmiştir ( Maddala (1974), Vinod ve Ullah(1981)).

Özkale ve Kaçıranlar (2007) iki parametrelili Liu tahmin edici tanımlamıştır. Gruber (2012) bu tahmin edicinin, ridge tahmin edici gibi, bir Bayesyen tahmin edici olduğunu göstermiştir.

Bu çalışmada amacımız, Almon ile Ridge ve Liu tipi Bayesyen tahmin edicilerin birleştirilmesiyle elde edilen alternatif tahmin edicileri karşılaştırmaktır.

KAYNAKLAR

[1] Almon, S. (1965), The Distributed Lag between Capital Appropriations and Expenditures, *Econometrica*, Vol 30, 96-178.

[2] Liu, K. (1993), A New Class of Biased Estimate in Linear Regression, *Commun. Stat. –Theory Meth.*, 22(2), 393-402.

[3] Maddala, G.S. (1974), Ridge Estimator for Distributed Lag Models, NBER Working Paper Series, no:69.

[4] Vinod, H.D. and Ullah, A. (1981), *Recent Advances in Regression Methods*, Marcel Dekker, New York.

[5] Özkale, M.R. and Kaçıranlar, S.(2007), The Restricted and Unrestricted two-parameter estimators. *Commun. Stat. –Theory Meth.*, 36,,2707-2725.

ABSTRACT

COMPARISONS OF BAYESIAN LIU AND RIDGE ESTIMATORS FOR DISTRIBUTED LAG MODELS

The finite distributed lag model is as follows:

$$y_t = \delta + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} + u_t \quad t = p + 1, \dots, T$$
$$= X\beta + u \quad (1)$$

In the presence of multicollinearity problem, the OLS estimator of the  $\beta$ 's in (1) has some disadvantages.

To tackle these problems, a generalization of the linear nonstochastic prior on  $\beta_i$  is given by Almon (1965). However, we may face some problems applying the Almon method. To circumvent these problems, the ridge type estimator approach under stochastic smoothness priors is considered as an alternative.

It is also demonstrated that these estimators are special cases of the linear Bayes estimator ( Maddala (1974), Vinod and Ullah (1981)). Özkale and Kaçıranlar (2007) proposed a two-parameter variation of the Liu estimator. It is demonstrated that two-parameter Liu estimator is a special case of the linear Bayes estimator as the ridge estimator ( Gruber (2012)).

Our aim in this study is to introduce alternative estimators defined by combining the Almon estimator with the Liu and ridge type bayesian estimator. We also compare their properties in some detail.

**Key Words:** Finite distributed lag model; Almon estimator; Bayesian estimator; ridge estimator; Liu estimator

**KONİK IŞINLI BİLGİSAYARLI TOMOGRAFİ GÖRÜNTÜLERİNDEKİ ÇENE  
PATOLOJİLERİNİN RASGELE YÜRÜYÜŞLE BELİRLENMESİ**

Ercüment YILMAZ\*  
Karadeniz Teknik Üniversitesi, Enformatik Bölümü, 61080, TRABZON, TÜRKİYE  
ercument@ktu.edu.tr

Temel KAYIKÇIOĞLU  
Karadeniz Teknik Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği Bölümü, 61080,  
TRABZON, TÜRKİYE  
tkayikci@ktu.edu.tr

Saadettin KAYIPMAZ, Ömer Said SEZGİN  
Karadeniz Teknik Üniversitesi, Diş Hekimliği Fakültesi, Oral Diagnoz ve Radyoloji ABD, 61080, TRABZON,  
TÜRKİYE  
skayitmaz@yahoo.com, omersaidsezgin@gmail.com

**Giriş**

Bu çalışmanın konusu; Konik Işınli Bilgisayarlı Tomografi (KIBT) cihazlarıyla elde edilen tıbbi görüntülerde çene patolojilerinin rasgele yürüyüş yöntemi ile tespit edilmesidir. Diş hekimleri bu amaçla geliştirilen bir yazılım aracılığıyla 3 boyutlu görüntüler içerisinde hastalıklı bölgeleri tespit ederek bölütleyebilecekleri gibi, matematiksel ve olasılıksal bir yöntem olan rasgele yürüyüş yöntemini kullanarak bu bölgeleri en yakın doğrulukta ve daha kısa sürede de işaretleyebileceklerdir.

**Hastalıklı Bölgenin Hekim Tarafından Tespit Edilmesi**

KTÜ Diş Hekimliği Fakültesi Oral Diagnoz ve Radyoloji kliniğine gelen hastalara ait KBIT görüntülerinde, çene bölgelerinde hastalıklı patolojiler tespit edilen 30 kişiye ait kesit dosyaları, bu çalışmada kullanılan veri setlerini oluşturmaktadır. DICOM formatındaki [1] dosyalarda saklanan üç boyutlu görüntü bilgilerinin hekim tarafından incelenmesine imkan tanıyan özel bir yazılım hazırlanmıştır. Hekimin çalıştığı eksen boyunca patolojinin büyüklüğüne göre bir hasta için 20 ile 150 adet görüntü kesiti işaretlenerek bu bölgelere ait koordinatlar üç boyutlu görüntü matrisinde saklanmıştır.

**Rasgele Yürüyüş Yöntemi**

“Rasgele Yürüyüş” görüntü bölütleme işlemlerinde kullanılan, temelde olasılığa ve graf teorisine dayalı bir yöntemdir [2]. Yöntemde iki farklı gruba ait “nesne” ve “arka plan” etiketli tohumlar bir görüntünün üzerindeki uygun veri noktalarına (piksel) yerleştirilir ve her piksel bir düğüm olarak kabul edilerek bu piksellerin koordinatları ve parlaklık değerlerini içeren Laplace Matrisi oluşturulur. Matris üzerinde etiketlenmemiş bir düğüm için rasgele yürüyüş başlatılır ve önceden etiketlenmiş düğümlerin her birine ulaşma olasılıkları ayrı hesaplanır. Bu olasılık hesabında, matristeki komşu düğümler arasındaki kenar ağırlıklarının elde edilebilmesi için tipik Gauss ağırlık fonksiyonu (1) kullanılmaktadır.

$$w_{ij} = \exp(-\beta(g_i - g_j)^2) \quad (1)$$

Rasgele yürüyüşün başlatıldığı düğüm için hesaplanan olasılık, ulaşılan tohumlardan hangisi için daha yüksek çıkarsa o piksel bu tohuma ait etiket ile işaretlenir. Bu işlem görüntü üzerindeki bütün pikseller için tekrar edilerek “nesne” ve “arka plan” bölgeleri tespit edilir ve böylece bölütleme işleminin gerçekleştirilmesi sağlanır.

**Hastalıklı Bölgelerin Rasgele Yürüyüş Yöntemi ile Tespit Edilmesi**

Aynı görüntülerde işaretlenmek istenilen alanın tespit edilebilmesi için, yazılım ekranında ilgili görüntünün üzerindeki uygun veri noktalarına (piksel) tohumlar yerleştirilir. Her bir kesit için rasgele yürüyüş yöntemi çalıştırılarak işaretlenecek olan bölge otomatik tespit edilir ve bu bölgeye ait koordinatlar üç boyutlu görüntü matrisinde saklanır.

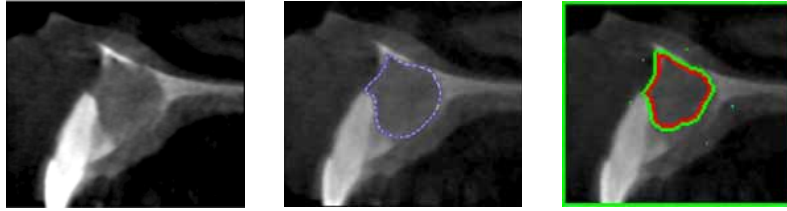
**Sonuçların Karşılaştırılması**

Hastalıklı bölgeye ait görüntüler üzerinde (Şekil 1.a) rasgele yürüyüş yönteminden elde edilen sonuçlar (Şekil 1c) hekim tarafından işaretlenmiş (Şekil 1.b) ve referans olarak kabul edilen görüntüler ile karşılaştırılmıştır.

**BİLDİRİ OTURUMLARI 5**

*SESSION 5*

*Veri Analizi ve Modelleme*



Şekil 1. a) Hastalıklı bölgeye ait görüntü, b) Hekim tarafından işaretlenmiş görüntü, c) Rasgele yürüyüş ile işaretlenmiş görüntü

İki farklı yolla işaretlenmiş ve bilgileri üç boyutlu matrislerde saklanmış olan bölütleme sonuçlarının kıyaslanmasında (2) numaralı hata fonksiyonundan faydalanılmıştır.

$$Hata = 1 - \frac{\sum(R_1 \cap R_2)}{\sum(R_1 \cup R_2)} \quad (2)$$

#### KAYNAKLAR

- [1] D. Grauer, L.S.H. Cevidanes, W.R. Proffit, "Working with DICOM craniofacial images" American Journal of Orthodontics and Dentofacial Orthopedics, 136(3): 460–470, 2009
- [2] L. Grady. "Random walks for image segmentation". IEEE TPAMI, 28(11):1768–1783, 2006

#### ABSTRACT

#### DETECTION OF PATHOLOGIES OF JAWS IN CONE BEAM COMPUTER TOMOGRAPHY IMAGES USING RANDOM WALKS

Subject of this study is to detect pathologies of jaws in medical images obtained by CBCT devices. Performance of the method was tested with comparing the regions marked by dentists with regions detected by random walks algorithm.

**Key Words:** CBCT, Pathologies of Jaws, Random Walks, Gaussian Weighting Function

#### REGRESYON MODELİ İÇİN HEMEN HEMEN YANSIZ GENELLEŞTİRİLMİŞ LIU M TAHMİN EDİCİSİ

Hasan ERTAŞ\* Selahattin KAÇIRANLAR

\*Çukurova Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 01330, Adana, TÜRKİYE,  
hertas@cu.edu.tr

Çukurova Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 01330, Adana, TÜRKİYE, skacir@cu.edu.tr

#### Giriş

Çoklu lineer regresyon modeli;

$$y = X\beta + \varepsilon \quad (1)$$

y, n×1 tipinde yanıt değişkenin vektörü, X, n×p tipinde açıklayıcı değişkenlerin matrisi, β, p×1 tipinde bilinmeyen parametrelerin vektörü ve ε, E(ε) = 0 ve Var(ε) = σ<sup>2</sup>I olan n×1 tipinde rastgele hataların vektörüdür. (1) denkleminin En Küçük Kareler (EKK) tahmin edicileri;

$$\hat{\beta} = (X'X)^{-1}X'y \quad (2)$$

ile verilir

(1) ifadesi ile verilen regresyon modelinde, EKK yönteminin uygulanması için belli varsayımlar gerekmektedir. Bu varsayımlardan özellikle; tüm gözlemlerin kestirilmiş regresyon doğrusu üzerinde aynı etkiye sahip ve açıklayıcı değişkenlerin lineer bağımsız olması çok önemlidir. Bu koşullar sağlanmadığında, sapan gözlemlerin varlığı ve çoklu iç ilişki olarak adlandırılan iki problemle karşılaşılır.

Veri setinde sapan değerlerin bulunması durumunda EKK tahmin edicisi olumsuz etkilenmekte ve yapılan tahminler güvenilir olmamaktadır. Bu sorunu ortadan kaldırmak için Dayanıklı Regresyon yöntemi kullanılır. Bu yöntemde sapan değerlerden etkilenmeyen dayanıklı kestiriciler üretilmeye çalışılır. Dayanıklı regresyonda bir çok kestirim yöntemi önerilmiştir. Bunlardan, literatürde en yaygın kullanılanı M kestiricilerdir ( $\hat{\beta}_M$ ) [4]. M-kestiriciler y-yönündeki sapan değerlere karşı etkili (Bozulma noktası % 30) fakat X-yönündeki sapan değerlere karşı etkili değildir (Bozulma noktası % 0).

Diğer yandan açıklayıcı değişkenler arasında lineer bağımlı olması çoklu iç ilişki problemine neden olmaktadır. Bu problemin etkisini minimuma indirmek için önerilen kestirim yöntemleri yanlı tahmin edicilerin ortaya çıkmasına ve incelenmesine neden olmuştur. Literatürde yanlı tahmin edicilerle ilgili birçok araştırma yapılmış olup, bazıları şu şekildedir: Ridge tahmin edicisi (RE) [3] ve Liu tahmin edicisi (LE) [5]. RE ve LE tahmin edicileri y-yönündeki sapan değerlere karşı hassas olduğunda Silvapulle (1991) de M-kestiricilere dayalı robust ridge (RME) ve Arslan ve Billor [2] M-kestiricilere dayalı Liu (LME) tahmin edicilerini tanımlayıp üstünlüklerini incelemiştir.

Akdeniz ve Kaçıranlar LE nin yanlılığını azaltmak için hemen hemen yansız Liu tahmin edicisini önermiştir (AUGLE) [1]:

$$\tilde{\beta}_L^* = [I - (X'X + I)^{-2}(I - D)^2]\hat{\beta} = W(D)\hat{\beta} \quad (3)$$

burada; D = diag(d<sub>1</sub>, d<sub>2</sub>, ..., d<sub>p</sub>), 0 < d<sub>i</sub> < 1, i = 1, 2, ..., p. AUGLE y-yönündeki sapanlardan olumsuz etkilendiği için hemen hemen yansız genelleştirilmiş Liu M-tahmin edicisi önerilmiştir (AUGLME):

$$\tilde{\beta}_{LM}^* = W(D)\hat{\beta}_M \quad (4)$$

ile verilir.

Bu çalışmada yeni tanımlanan AUGLME nin; Hata kareler Ortalaması kriterine (MSE) göre AUGLE, LME, LE ve M tahmin edicilerine göre üstünlükleri incelenmiş olup elde edilen teorik sonuçlar bir nümerik örnek ile desteklenmiştir.

#### KAYNAKLAR

- [1] Akdeniz F. and Kaciranlar S. (1995), On the almost unbiased generalized Liu estimator in lineer regression, Communications in Statistics 22, 393- 402.
- [2] Arslan O. and Billor N. (2000), Robust Liu estimator for regression based on an M-estimator, Journal of Applied Statistics, 27,39-47
- [3] Hoerl, A. E.; Kennard, R. W., (1970), Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics 12 (1), 55-67.
- [4] Huber P. J. (1981), Robust Statistics (New York, Wiley)
- [5] Liu K. (1993), A New Class of Biased Estimate in Linear Regression, Comm. Statist. Theory Methods, 22 (2), 393-402.

#### ABSTRACT

##### A ALMOST UNBIASED GENERALIZED LIU M-ESTIMATOR FOR REGRESSION MODEL

In regression analysis, the presence of multicollinearity among independent variables is a common problem which exhibits serious undesirable effects on the ordinary least squares (OLSE). In addition, outliers in data set can strongly distort OLSE estimators and lead to unreliable results. In the existence of multicollinearity and outliers in data sets, it is suggested using biased methods based on robust estimator [2]. In this paper, to overcome the combined problem of multicollinearity and outliers in the y-direction, we propose a new almost unbiased generalized Liu M-estimator (AUGLME). We show using theoretic results that our new estimator advantages over almost unbiased generalized Liu estimator (AUGLE) [1], Liu-type M-estimator (LME) [2] and M-estimator (M) [4]. Moreover, a numerical example is given to show the theoretical results.

**Key Words:** Robust Regression, Biased Estimation, Multicollinearity,Outliers

#### AŞIRI KÜÇÜK ÖRNEKLEM PROBLEMİNDE BİLGİ KARMAŞIKLIĞI KRİTERİ (ICOMP) VE DÜZENLEŞTİRİLMİŞ KOVARYANS YAPILARI İLE BOYUT İNDİRGEMEYE ALTERNATİF BAKIŞ: GEN VERİSİ ÖRNEĞİ

Esra PAMUKÇU<sup>a\*</sup> Hamparsum BOZDOĞAN<sup>b</sup>

<sup>a</sup>Fırat Üniversitesi, Fen Fakültesi İstatistik Bölümü 23119 ELAZIĞ  
epamukcu@firat.edu.tr

<sup>b</sup>: University of Tennessee, College of Business Administration, Department of Statistics, Operations, and Management Science 231 Stokely Management Center Knoxville, TN 37996-0532

bozdogan@utk.edu

#### Materyal ve Metot

Gerçek yaşamdaki bir çok problemde, kovaryans matrisleri, kötü şartlandırılmış (ill-conditioned), pozitif tanımlı olmayan hatta singüler yapıya sahip olabilmektedirler. Bu özellikle yüksek derecede ilişkili tahmin edicilerin olduğu regresyon modellerinde ve aşırı derecede küçük örneklem probleminin  $n \ll p$  olduğu durumlarda ortaya çıkabilir [1]. Bu çalışmanın amacı, yukarıda bahsi geçen problemlerin PCA analizinden elde edilecek sonuçların güvenilirliği üzerinde etkisi olacağı için, bunların üstesinden gelebilmek amacıyla bazı stratejiler geliştirmektir.

Bu amaçla, Khan ve ark. (2001)'de [2] yaptıkları çalışma sonucunda elde edilmiş bir veri seti olan SRBCT(Small Round Blue Cell Tumor) gen veri seti, analize alınmıştır. Veri setinde; EWS: Ewing's sarcoma-23 gözlem, BL: Burkitt Lymphoma -8 gözlem, NB: Neuroblastoma-12 gözlem, RMS: Rhabdomyosarcoma-20 gözlem olmak üzere bu tümör çeşitlerine ait 2300 gen incelenmiştir.

İlk olarak, klasik PCA analizinde kullanılan örnek kovaryans matrisi S yerine, Maksimum Entropi Kovaryans Matrisi hesaplanmıştır [3]. Yeni elde edilen kovaryans matrisi için özdeğerler incelendiğinde, örnek kovaryans matrisi S' de ortaya çıkan negatif özdeğer içermesi probleminin üstesinden gelindiği görülmüştür. İlave olarak, bu kovaryans matrisi için singülerliğe yakın olup olmadığının incelenmesi de gerekmektedir. Bunun için, Bozdogan tarafından singülerliğin tanımı için önerilen  $\kappa(\Sigma) = \frac{1}{\sigma_{\min}} / \sigma_{\max}$  sayısı kullanılmıştır [4]. Kovaryans matrisinde singülerliğin de üstesinden gelmek için önerilen bazı düzenleme yöntemleri mevcuttur [1]. Bu çalışmanın uygulama aşamasında Thomaz Stabilizasyonu ile Convex Sum toplamı ile ifade edilen düzenleme yöntemlerinin bir hibrit formu olarak STA\_CSE yapısı kullanılmıştır.

STA\_CSE kovaryans matrisi kullanılarak PCA analizi yapıldığında, veri ne kadar yüksek boyutlu olursa olsun, boyut sayısı kadar PC elde edilebilmekte ve klasik PCA uygulanması sonucunda ihmal edilebilecek yüksek boyutlardaki bazı gizli bilgilere de ulaşılabilmektedir.

Bundan sonraki aşama, önemli olan PC sayısına karar vermektir. Literatürde model seçim kriterleri olarak bilinen çeşitli kriterlerler, burada önemli PC sayısına karar vermek için kullanılmıştır. Bunlar ICOMPPEU\_C1F, ICOMPPEU\_mis, CAIC, CICOMP ve AIC kriterleridir. Her birisi için hesaplamalar yapılmış ve çalışma kapsamında hangi kriterin PC sayısının seçiminde kullanılabileceğine grafikler ve açıklayıcı oranları kullanılarak karar verilmiştir.

Son aşama olarak, elde edilen PC'lerin, veriyi ne kadar iyi temsil edebildiğini anlamak için bir sınıflama prosedürü kullanılmak istenmiştir. Bunun için LDA ve QDA sınıflayıcıları kullanılmıştır. İhtiyaç olması halinde ise RDA kullanılmıştır.

#### 2. Sonuçlar

Tablo-1: Bilgi Kriterleri için toplu sonuçlar

Bilgi Kriteri	PC sayısı	Açıklayıcılık
ICOMP_PEU_C1F	12	%95
ICOMP_PEU_MISS	32	%97

CAIC	6	%93
CICOMP	6	%93
AIC	17	%96

Yukarıdaki tabloda bilgi kriterlerinin seçtiği PC sayıları görülmektedir. Seçilen PC'ler ile sınıflama yapılmış oldukça başarılı sonuçlar elde edilmiştir. Bu sonuçlar özellikle LDA sınıflama prosedüründe daha açık bir şekilde kendini göstermektedir. Orijinal veri ilk 12 gen ile %17,46 oranında bir hatalı sınıflama yüzdesi elde edilirken, ME\_STA\_CSE kovaryans yapısı kullanılarak oluşturulan PC skorları %1,587 oranında bir hatalı sınıflama yüzdesi elde edilmiştir. QDA için, orijinal veri singülerlik hatası vermeden maksimum %15,87 oranında hatalı sınıflama verirken, önerilen yeni yöntem sayesinde bu oran %7,937'ye kadar gerilemiştir. Sonuç olarak, aşırı derecede küçük örnekleme sahip bir gen veri seti, hiçbir transpoz alma işlemi uygulamadan analize dahil edilmiş ve klasik yöntemlerle hesaplanamayan özdeğerlerin hepsine ulaşılabilmektedir. Ayrıca yapılan hesaplamalar ile, kullanılan düzgünleştirme ve düzenleme işlemleri sonucunda kovaryans matrisleri için önemli bir problem olan singülerlik durumunun üstesinden gelindiği görülmüştür.

#### KAYNAKLAR

- [1] Bozdoğan, H., Howe, J.A., 2012, Misspecified multivariate regression models using the genetic algorithm and information complexity as the fitness function, European Journal of Pure and Applied Mathematics. 5(2): 211-249
- [2] Khan et.al. 2001, Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat. Med., 7(6):673-679
- [3] Fiebig D.G., 1984, On the maximum entropy approach to undersized samples, Applied mathematics and computation, 14:301-312
- [4] Bozdoğan, H., Information Complexity and Multivariate Learning in High Dimensions with Applications in Data Mining. Forthcoming book.

#### ABSTRACT

**AN ALTERNATIVE APPROACH TO DIMENSION REDUCTION WITH INFORMATION COMPLEXITY CRITERIA AND SMOOTHED COVARIANCE MATRICES FOR UNDERSIZED SAMPLE PROBLEM: A GENE EXPRESSION DATA ANALYSIS**

When the number of observations is much smaller than the number of features an undersized sample problem occurs. In this case the usual sample covariance matrix  $S$  degenerates and it cannot be computed. This is a challenging and unresolved problem in the literature. In this paper we introduce a new and novel approach using maximum entropy covariance matrix and its hybridized smoothed covariance estimators along with the information complexity (ICOMP) criterion of Bozdoğan to reduce the dimensionality of the data.

**Key Words:** Dimension Reduction, Information Complexity, Smoothed Covariance Matrices, Gene Expression Data Analysis

#### RETİNA GÖRÜNTÜLERİNİN İSTATİSTİKSEL OLARAK KALİTE DEĞERLENDİRMESİ

Uğur ŞEVİK\*, Tolga BERBER  
Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü 61080  
TRABZON, TÜRKİYE  
usevik@ktu.edu.tr, tberber@ktu.edu.tr

Hidayet ERDÖL

Karadeniz Teknik Üniversitesi, Tıp Fakültesi, Göz Hastalıkları AD, 61080, TRABZON, TÜRKİYE  
herdol@ktu.edu.tr

#### 1. Giriş

Görüntü kalite değerlendirmesi, medikal görüntülerin analizi yapılmadan önce uygulanacak yöntemlerin ve algoritmaların etkin sonuç verebilmeleri için görüntü ön işlemeden önce yapılması gereken ilk aşamadır. Fundus kameralarla elde edilen görüntü, çok karanlık, çok aydınlık, çok kapağının kapanması, çok bulanık, kirli lens, yanlış odaklanma, objektif karşısında hastanın olmayışı ve objektif kapağının kapalı olması gibi istenmeyen görüntüler olabilir. Bu gibi retina dışı görüntülerin otomatik tespit sistemlerine alınması, sistemin ne ön işleme aşamasında ne de lezyon tespit ve sınıflandırma aşamasında doğru sonuçlar vermeyecektir. Bu nedenle, retina analizi için otomatik teşhis ve tanı koyma sistemlerinde, mutlaka önceden bir kalite değerlendirmesi yapılması gerekmektedir. Çünkü sistem otomatik çalışmakta ve her kararı kendisi vermektedir. Sisteme retina görüntüsü dışında herhangi bir görüntü veya yeterli bilgi içermeyen görüntü geldiğinde, bunu algılayıp analiz sürecine girmesine izin vermemelidir. Bu işlem sayesinde analiz sürecinin neden olduğu zaman kaybı engellenir ve sistemin doğru karar verme oranı artırılmış olur.

#### 2. İstatistiksel Öznitelik Çıkarımı

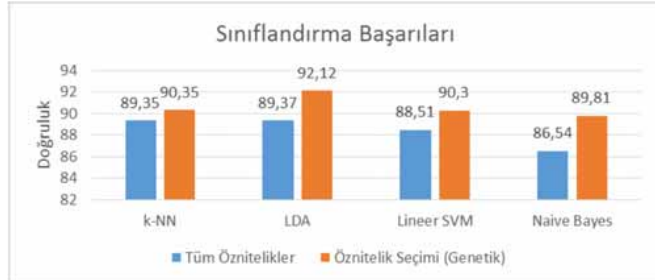
Retina kameralarla elde edilen retina görüntüleri iyi, kötü ve retina dışı olmak üzere üç sınıfa ayrılmıştır. Veri setimizde, 125 iyi, 69 kötü ve 22 retina dışı görüntü olmak üzere toplamda 216 görüntü vardır. Her sınıftaki görüntülerin yarısı test, diğer yarısı da eğitim verisi olarak kullanılmıştır. Gri-seviye retina görüntülerinin histogramlarından elde edilen istatistiksel öznitelikler ortalama, standart sapma, yumuşaklık, çarpıklık, tekdüzelik, basıklık, parlaklık olasılıkları ve ortalama karşıtlık olarak sıralanabilir [1]. Bunun yanı sıra Haralick ve diğerlerinin [2] önermiş olduğu yöntemle, dokunun uzamsal parlaklık bağımlılıkları kullanılarak istatistiksel öznitelikler çıkarılmıştır. Bu bağlamda öznitelikleri çıkarılacak doku, parlaklık değerlerinin  $d$  uzaklığında ve  $\theta$  açısında birlikte görülmelerini temsil eden Gri Seviye Birlikte Görülme Matrisinin (GLCM; Gray Level Co-occurrence Matrix) bazı istatistiksel özelliklerinin hesaplanması ile 14 istatistiksel öznitelik çıkarılmıştır [2]. Tüm bu öznitelikler, eğitim kümesindeki görüntülerden toplanarak sınıflandırıcıların eğitilmesi için kullanılmıştır.

#### 3. Kullanılan Sınıflandırıcılar

Retina görüntülerinin eğitim kümesinden çıkarılan istatistiksel öznitelikler, gözetimli öğrenme yöntemi kullanılarak sınıflandırıcılar eğitilmiştir. Retina görüntülerinin kalite değerlendirmesinde literatürde sınıflandırma için yaygın olarak kullanılan Lineer Ayrıştırma Analizi (LDA), k-En Yakın Komşu (k-NN), Naïve Bayes ve Destek Vektör Makinesi (SVM)

sınıflandırıcıları kullanılmış ve performans karşılaştırması yapılmıştır [3], [4], [5]. Sınıflandırıcılar eğitilirken iki yöntem kullanılmıştır. İlk olarak, tüm öznitelikler kullanılarak eğitim sağlanmış ve ikinci olarak da genetik algoritma yöntemi ile öznitelik seçimi (feature selection) yapılmıştır. Bu iki yöntem ile eğitilen sınıflandırıcıların performansları Grafik 1'de verilmiştir.

#### 4. Sonuçlar



Grafik 1. Tüm öznitelikler ve genetik algoritma ile öznitelik seçimi yapılarak elde edilen performans değerleri.

Buna göre, tüm özniteliklerin kullanımı yerine etkin olanların seçilerek sınıflandırıcılarda kullanılması başarı oranını yükseltmiştir. Deneyler sonunda LDA sınıflandırıcısı, tüm öznitelikleri kullanarak %89,37 ve öznitelik seçimi yapılarak %92,12 doğruluk oranı ile diğer sınıflandırıcılardan daha başarılı olduğu görülmektedir.

#### KAYNAKLAR

- [1] Southard T. E. and Southard K. A., (1996), "Detection of simulated osteoporosis in maxillae using radiographic texture analysis.," IEEE transactions on bio-medical engineering, vol. 43, no. 2, pp. 123-32.
- [2] Haralick R. M., Shanmugam K., and Dinstein I. H., (1973), "Textural features for image classification.," IEEE Transactions on Systems, Man and Cybernetics, vol. 3, no. 6, pp. 610-621.
- [3] Mika S., Ratsch G., Weston J., Scholkopf B., and Mullers K., (1999), "Fisher discriminant analysis with kernels.," in Neural Networks for Signal Processing IX, Proceedings of the 1999 IEEE Signal Processing Society Workshop, pp. 41-48.
- [4] Zhang H., (2004), "The Optimality of Naive Bayes.," in Proceedings of the 17th International FLAIRS conference, vol. 1, no. 2.
- [5] Cortes C. and Vapnik V., (1995), "Support-vector networks.," Machine Learning, vol. 20, no. 3, pp. 273-297.

#### ABSTRACT

##### STATISTICAL QUALITY ASSESSMENT OF RETINAL IMAGES

In this study, we determined quality of retinal images by using automated quality assessment approaches based on statistical features and classified in three class, good, bad and outlier using four classifier. The accuracy of image quality assessment tests were %89,37 and %92,12 for LDA classifier, respectively.

**Key Words:** Statistical Image Features, Image Quality Assessment, LDA, SVM, k-NN, Bayes

#### ÇOKGENSEL ALANLARDA İKİ DEĞİŞKENLİ Kİ-KARE UYUM İYİLİĞİ TESTİ EXPRESSION DATA ANALYSIS

Orhan KESEMEN\*, Tuncay ULUYURT  
Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, Trabzon, Türkiye  
okesemen@gmail.com, tuncayuluyurt@hotmail.com

#### Giriş

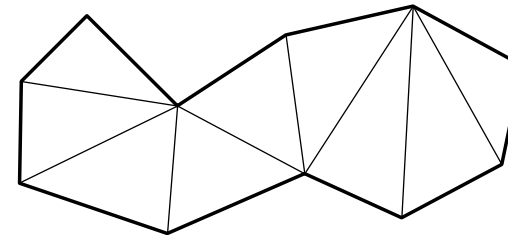
Çıkarımsal istatistik, modern istatistiğin temel konularından birisidir. Çıkarımsal istatistik, ileri sürülen bir hipotez eşliğinde, parametrik ve parametrik olmayan testler olmak üzere, iki biçimde incelenmektedir. Gerekli varsayımların geçerli olmadığı durumlarda, parametrik teknikler büyük ölçüde güvenilirliklerini kaybederler. Bu gibi durumlarda, parametrik olmayan teknikler devreye girer [1]. Araştırmacılar, deneysel veriler ile kuramsal veriler arasındaki uyum iyiliğini belirlemek için genellikle Ki-kare testi kullanırlar. Bu teste amaç beklenen frekans sayısı ile gözlenen frekans değerleri karşılaştırılır. Gözlenen değerler ya deneylerden ya da gözlem sonucunda elde edilirler. Kuramsal veya beklenen frekans sayısı ise ileri sürülen sıfır hipotezinde verilen dağılımından yararlanılarak hesaplanmaktadır.

#### Ki-Kare Uyum İyiliği Testi

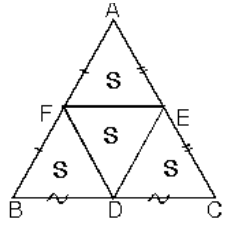
Ki-kare testi 1900 yılında Pearson tarafından ortaya atılmıştır. Bu testteki istatistik sapmaların karelerinin toplamının beklenen frekans sayısına bölünmesiyle elde edilmiştir [2]. Ki-kare uyum iyiliği testi birçok uygulamada gözlenen değerlerin tek değişkenli dağılımlara uyup uymadığını kontrol etmek için kullanılmıştır. Ancak iki boyutlu bir alanda elde edilen gözlem değerleri test edilmesi için dörtgenel bir alanla çerçevesi gerekmektedir. Ancak bilindiği gibi gerçek yaşamda elde edilen veriler her zaman dörtgen alan içinde olmazlar.

Keyfi bir alan içinde gözlenen verilerin bilindik bir geometrik şekil ile tanımlanmasında en uygun yöntem çokgenel tanımlamadır [3]. Çokgenel bir alan içinde tanımlanan olasılık yoğunluk değerinden beklenen frekans sayısını bulmak için integralinin alınması gerekir. Bu ise oldukça zor olmaktadır. Önerilen yöntemde ise çokgenel alan uygun üçgenlere bölünerek (Şekil 1) hem sınıf sayısı belirlenebilir hem de bu alandaki beklenen frekans sayısı hesaplanabilmektedir. Beklenen frekans sayısı belirlenirken üçgen üzerinden integral alınmasında üçgen tepe noktasından iki parçaya bölünerek hesaplanabilmektedir. Aynı şekilde üçgen içerisine düşen gözlenen frekans sayısı da algoritmik olarak hesaplanabilmektedir [3].

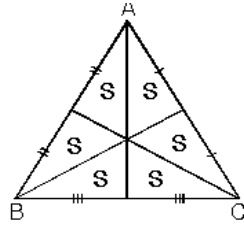
Eğer üçgen sayısı Ki-kare testi için yeterli değilse her üçgen kendi içerisinde kenarortayları yardımıyla dört parçaya (Şekil 2(a)) veya altı parçaya bölünerek işlemler gerçekleştirilmiştir.



Şekil 1. Çokgenel alanın üçgenlere bölünmesi



(a)



(b)

Şekil 2. Üçgenlerin kenar ortaylar kullanılarak yeni üçgenlere bölünmesi

#### KAYNAKLAR

- [1] H. Bircan, H., Karagöz Y. ve Kasapoğlu, Y., (2003) Ki-Kare Ve Kolmogorov Smirnov Uygunluk Testlerinin Simülasyon İle Elde Edilen Veriler Üzerinde Karşılaştırılması, C.Ü. İktisadi ve İdari Bilimler Dergisi, Cilt 4, Sayı 1, 200.
- [2] Gibbons, J. D. and Chakraborti, S., (2003) Nonparametric Statistical Inference, Fourth Edition, Marcel Dekker, Inc., Alabama, USA.
- [3] Kesemen, O., and Doğru, F. Z., (2011) Cumulative Distribution Functions of Two Variables In Polygonal Areas, 7. International Statistical Congress, 28 April – 01 May, Antalya, Turkey.

#### ABSTRACT

##### BIVARIATE CHI-SQUARE GOODNESS OF FIT TEST IN POLYGONAL AREAS

In this study, an arbitrary two-dimensional distribution of an arbitrary field the Chi-square test is described how to do. The two-dimensional polygon definition is used to define an arbitrary area. This is the expected frequency in the area to find the most appropriate given polygon is divided into triangles. The calculated value of the distribution of each triangle, and thus falling into the expected frequency values were determined for each triangle It is also observed in the frequency values that fall into each triangle by calculating the Chi-square test was also investigated how to do.

**Key Words:** Chi-square, two-variable distributions, goodness of fit test, polygonal distributions.

#### PARAMETRİK OLMAYAN REGRESYONDA OPTİMAL BANT GENİŞLİĞİ SEÇİM YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Aydın KARAKOCA<sup>1</sup>, Murat ERİŞOĞLU<sup>1</sup>, Alper SİNAN<sup>2</sup>  
<sup>1</sup>Necmettin Erbakan Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Konya

E-mail : akarakoca@konya.edu.tr , merisoglu@konya.edu.tr

<sup>2</sup>Sinop Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Sinop

E-mail : alpsin@sinop.edu.tr

#### Giriş

Genel olarak parametrik olmayan regresyon modeli  $f$  fonksiyonu özel olarak belirtilmeden,

$$y_i = f(x_i) + \varepsilon_i \quad (1)$$

eşitliği ile verilebilir. Parametrik olmayan regresyon modelinde modelin parametrelerinin tahmini yerine doğrudan  $f$  fonksiyonu tahmin edilir. Parametrik olmayan regresyonda çoğu metot  $f$  fonksiyonunun düzgün (smooth) ve sürekli bir fonksiyon olduğu ve hataların sıfır ortalamalı ve sabit varyanslı normal dağılım gösterdiği varsayımına dayanır. Parametrik olmayan regresyon modelinde açıklayıcı değişken sayısı arttıkça tahmin zorlaştığı için özel olarak tek değişkenli durumun göz önüne alındığı tek değişkenli basit parametrik olmayan regresyon modeli sıklıkla kullanılmaktadır.

#### Kernel Regresyon

Parametrik olmayan regresyonda bağımsız değişken  $X$  ve bağımlı değişken  $Y$  arasındaki fonksiyonel ilişkiyi tahmin etmek için kullanılacak yöntemlerden biri de Nadaraya-Watson ve yerel polinomial kernel regresyon yöntemidir. Bu yöntemde,

$$E(Y/X) = m(x) \quad (2)$$

koşullu beklenen değeri ile belirtilen  $m(\cdot)$  fonksiyonu tahmini ile ilgilenilmektedir. Basit parametrik olmayan regresyon modelinde  $(x_1, y_1), K, (x_n, y_n)$  bağımsız gözlemleri için Nadaraya-Watson tahmin edicisi (3) eşitliği ile verilmiştir.

$$\hat{m}_h(x) = \frac{\sum_{i=1}^N K\left(\frac{x_i - x}{h}\right) y_i}{\sum_{i=1}^N K\left(\frac{x_i - x}{h}\right)} \quad (3)$$

(3) ile verilen eşitlikte  $K$ , kernel fonksiyonu ve  $h$  bant genişliği parametresidir. Literatürde sıklıkla kullanılan kernel fonksiyonları Boxcar, Gaussian, Epanechnikov ve Tricube kernel fonksiyonlarıdır.

#### Bant genişliği Seçimi

Kernel regresyonda kernel fonksiyonunun seçimi ve bant genişliği ( $h$ ) parametresinin seçimi önemli bir problemdir. Farklı bant genişlikleri ve farklı kernel fonksiyonları için elde edilen tahminler değişmektedir.

Bant genişliği seçimi için ortalama hata kareler(ASE), Hata kareler toplamı (RSS) ve Çapraz geçerlilik (CV) kriterleri kullanılmaktadır.

#### Uygulama ve Sonuç

Çalışmada farklı tipteki veri setleri için farklı kernel fonksiyonları ve bant genişliği seçim kriterlerine göre elde edilen sonuçlar karşılaştırılmıştır. Kernel fonksiyonlarının ve seçim kriterlerinin birbirlerine göre üstünlükleri belirtilmiştir.

#### KAYNAKLAR

1. Hardle, W., 1994, "Applied Nonparametric Regression", Humboldt Univ., Berlin Germany.
2. Fox, J., 2002, "Nonparametric Regression", Thousand Oaks CA:Sage.
3. Silverman, B. W., 1990, "Density Estimation for Statistics and Data Analysis", Chapman and Hall, London.
4. Pagan, A.and Ullah, A., 1999, "Nonparametric Econometrics", Cambridge University Press,Cambridge, U.K.

#### ABSTRACT

#### COMPARISON OF BANDWIDTH SELECTION METHODS FOR NONPARAMETRIC REGRESSION

Kernel estimation method is a useful method for estimation when the assumptions of parametric model is not provided. Bandwidth and kernel function selection is crucial in kernel regression In this study different bandwidth selection methods for different kernels are applied several data sets and estimations are compared.

**Key Words:** nonparametric regression, kernel regression, bandwidth

## BİLDİRİ OTURUMLARI 5

### SESSION 5

#### *Aktüerya Bilimleri*



## TRAFİK SİGORTASINDA RİSK ÖLÇÜMÜ

Aslıhan ŞENTÜRK ACAR  
Hacettepe Üniversitesi, Fen Fakültesi, Aktüerya Bilimleri Bölümü, Beytepe, Ankara

aslihans@hacettepe.edu.tr

Uğur KARABEY  
Hacettepe Üniversitesi, Fen Fakültesi, Aktüerya Bilimleri Bölümü, Beytepe, Ankara

ukarabey@hacettepe.edu.tr

### Giriş

Çoklu doğrusal regresyon, istatistiksel analizlerde sıklıkla kullanılan bir tekniktir. Bu regresyon modellerinde hata teriminin ortalama etrafında sabit bir varyans ile normal dağıldığı varsayılır. Bu varsayım, aktüeryal uygulamalarda analiz edilen verinin modellenmesinde yetersiz kalabilmektedir. Bu nedenle, normal dağılım göstermeyen değişkenlerin de modellenmesini sağlayan Genelleştirilmiş Doğrusal Modeller (GDM) daha sık tercih edilmektedir.

GDM'ler klasik doğrusal regresyon modellerin genelleştirilmiş biçimi olarak tanımlanmaktadır. Bu genelleme şu şekilde özetlenebilir,

- Ortalamadan rastgele sapmalar, normal dağılımdan farklı olarak üstel yayılım ailesinden Poisson, binom, gamma gibi dağılımlara sahip olabilir,
- Klasik doğrusal modellerde rastlantı değişkeninin ortalaması, açıklayıcı değişkenlerin doğrusal bir fonksiyonu iken GDM'lerde ölçüğe göre doğrusal olabilmektedir [2].

Bu çalışmada, araç sigortası fiyatlandırılmasında kullanılan regresyon modelleri incelenmiş, özel bir sigorta şirketinden alınan karayolları motorlu araçlar zorunlu mali sorumluluk sigortası (trafik sigortası) veri kümesine, incelenen modellerden bazıları uydurulmuş ve sonuçlar karşılaştırılmıştır. Veriye en iyi uyumu gösteren model kullanılarak bir benzetim çalışması yapılmış ve elde edilen hasar dağılımının risk ölçümleri incelenmiştir.

### Genelleştirilmiş Doğrusal Modeller

Hem doğrusal modellerin hem de GDM'lerde amaç, bağımlı değişken ile açıklayıcı değişkenler arasındaki ilişkinin açıklanmasıdır. Doğrusal modellerde bağımlı değişken,  $Y = \mu + \epsilon$  biçiminde tanımlanır. Burada  $\mu$ ,  $Y$  bağımlı değişkeninin ortalamasıdır. Modelin vektörel gösterimi,

$$\underline{Y} = X \underline{\beta} + \underline{\epsilon} \quad (1)$$

biçimindedir. Burada,  $\underline{Y}$  bağımlı değişkeninin gözlenmiş değerlerini içeren sütun vektörü,  $X$  açıklayıcı değişkenlerin tasarım matrisi,  $\underline{\beta}$  parametreler vektörü ve  $\underline{\epsilon}$  artıklar vektörüdür. Modelin temel varsayımları, sabit varyanslılık, normallik ve ilişkisizliktir. Eş.(1)'de verilen modelde, sistematik bileşen olarak  $p$  adet açıklayıcı değişken,  $\underline{\eta} = X \underline{\beta}$  doğrusal tahmin edicide birleşir [3].

Doğrusal modellerde rassal ve sistematik bileşenler arasındaki ilişki bağ fonksiyonu ile sağlanır. Klasik doğrusal modelde bağ fonksiyonu birim fonksiyondur. Genelleştirilmiş doğrusal modellerde, bağımlı değişkenin üstel dağılım ailesinden bir dağılıma sahip olduğu ve bağ fonksiyonunun monoton diferansiyellenebilir bir fonksiyon olduğu varsayılır Bağımlı değişkenin üstel aileden bir dağılıma sahip olmasının sonucu olarak, bu değişken genellikle değişen varyansa sahiptir [1].  $g$  bir bağ fonksiyon olmak üzere

$E[\underline{Y}] = \underline{\mu} = \underline{\eta}$  eşitliği kullanılarak, GDM'de bağımlı değişkenin beklenen değeri aşağıdaki gibi tanımlanır,

$$E[\underline{Y}] = \underline{\mu} = g^{-1}(\underline{\eta}) \quad (2)$$

Literatürde kullanılan bazı önemli bağ fonksiyonları,

### Çizelge 1. Bağ fonksiyonları

Log	$\eta = \log(\mu)$
Lojıt	$\eta = \log(\mu / (1 - \mu))$
Tamamlayıcı log-log	$\eta = \log\{-\log(1 - \mu)\}$

biçiminde tanımlanır [3].

### KAYNAKLAR

- [1] De Jong, P., Heller G.Z., (2008), Generalized Linear Models for Insurance Data, Cambridge University Press, London.
- [2] Kaas, R., Goovaerts M., Dhaene J., Denuit M., (2008), Modern Actuarial Risk Theory: Using R, Springer Verlag Inc., Berlin.
- [3] McCullagh, P., Nelder, J.A., (1989), Generalized Linear Models, Chapman and Hall, London.

### ABSTRACT

#### RISK MEASUREMENT IN MOTOR THIRD PARTY LIABILITY INSURANCE

In this study, the models which is used for motor insurance pricing are examined. Some of these models are fitted to a motor compulsory third party liability insurance data set that is taken from a Turkish insurance company and the results are compared. With using the best model a case study is simulated. Commonly used risk measures are calculated for simulated claim distribution.

**Key Words:** generalized linear models, risk measures, motor third party liability insurance.

## HASARLAR VE YATIRIMLAR ARASINDAKİ BAĞIMLILIK İLE HAYAT DIŞI SİGORTA ŞİRKETLERİNDE DİNAMİK FİNANSAL ANALİZ – BİR UYGULAMA

Betül Zehra KARAGÜL\*, Murat BÜYÜKYAZICI  
Hacettepe Üniversitesi, Fen Fakültesi, Aktüerya Bilimleri Bölümü, 06800, Ankara, Türkiye  
\*betul.zehra@hacettepe.edu.tr  
muratby@hacettepe.edu.tr

### Dinamik Finansal Analiz Modeli

Finansal modelleme yapmak şirketlerin geleceği öngörmesi ve risklerini doğru yönetebilmesi için oldukça önemlidir. Etkin ve yaygın olarak kullanılan finansal modellerden bir tanesi 1990'ların sonlarında ortaya çıkmış ve oluşumundan bu yana akademik alanda üzerine pek çok araştırma ve uygulama yayınlanmış olan Dinamik Finansal Analiz (DFA)'dir.

Literatürde tanımlanmış tek bir DFA modeli yoktur ve şirketler kendi ihtiyaçlarına göre farklı DFA modelleri belirleyebilirler. Bu özelliği ile DFA modeli oldukça kullanışlıdır.

Bu çalışmada bir hayat dışı sigorta şirketi için gerekli olan temel bileşenleri içeren Dinamik Finansal Analiz modeli kullanılmaktadır. Bu bileşenler öz sermaye, yatırım kazancı, sigorta kazancı, vergi oranı, yatırımların getirileri, prim geliri, maliyetler ve hasarlardır.

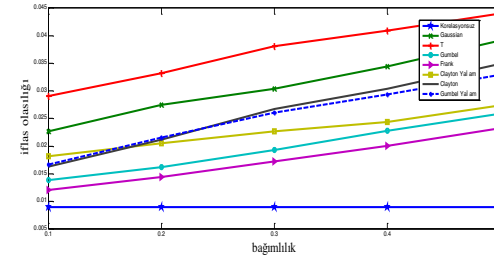
Hayat dışı sigorta şirketleri ağır kuyruklu ve çarpık risklere sahiptir. Oluşması durumunda sigortacıyı hem yatırım işinden hem de sigorta işinden zarara uğratan büyük hasarlar bu risklere örnektir. Kullanılan DFA modeline farklı varlık sınıfları (yüksek riskli yatırımlar ile düşük riskli yatırımlar) arasındaki bağımlılık, farklı tür yükümlülükler (katastrof hasarlar ile katastrof olmayan hasarlar) arasındaki bağımlılık ve yükümlülükler ile varlıklar arasındaki bağımlılık dahil edilmiştir. Bağımlılık yapısının genel şekli şöyledir;



Şekil 1. DFA modeli içerisindeki bağımlılık yapısı

### Uygulama

Çalışmada 5 yıllık süre için MATLAB yardımıyla 100.000 iterasyon ile DFA simülasyonunu yapılmıştır. Model parametreleri T.C Başbakanlık Hazine Müsteşarlığı'nın yayınlamış olduğu Sigortacılık ve Bireysel Emeklilik Faaliyetleri Hakkındaki Raporları incelenerek Türkiye hayatdışı sigorta şirketleri verisi yardımıyla elde edilmiştir. Kullanışlı olmaları nedeniyle bağımlılıklar kopyalar yardımıyla incelenmiştir. Farklı kopyaların etkilerini görmek için eliptik ve arşimedyan kopyalar ve bazı yaşam kopyaları kullanılmıştır. Farklı bağımlılık seviyeleri için de simülasyon çalışması yapılmış ve sonuçların iflas olasılığı, beklenen poliçeli açığı ve sharpe oranları üzerine etkileri incelenmiştir. Yatırımlar arasındaki bağımlılık seviyesinin iflas olasılığına etkileri Şekil 2.'de gösterilmiştir.



Şekil 2. Yatırımlar arasındaki bağımlılığın iflas olasılığına etkisi

Simülasyon sonucu yatırımlar ile hasarlar arasındaki bağımlılığın şirketin risk getiri ve performansı üzerinde önemli etkilere sahip olduğu görülmüştür. Bu durum düzenleyici ve denetleyici kurumlar ile sigorta şirketi hissedarları açısından oldukça önemlidir.

### KAYNAKLAR

- [1] Eling, M., Parnitzke T., Schmeiser H., Management Strategies and Dynamic Financial Analysis, Variance, 2(1): 52-70, **2008**
- [2] Eling, M., Toplek, D., Modeling and management of nonlinear dependencies –copulas in dynamic financial analysis, The Journal of Risk and Insurance, Vol. 76, No. 3, 651-681, **2009**
- [3] Nelsen, R.B., An Introduction to Copulas, second edition, Springer, **2006**
- [4] Kaufmann, R., Gadmer, A., Klett, R., Introduction to Dynamic Financial Analysis, Astin Bulletin, 31(1), 213-249, **2001**
- [5] Malevergne, Y., Sornette, D., Extreme Financial Risks: From Dependence to Risk Management, Springer, **2006**.

### ABSTRACT

#### DYNAMIC FINANCIAL ANALYSIS IN NONLIFE INSURANCE WITH THE DEPENDENCY BETWEEN INVESTMENTS AND CLAIMS -AN APPLICATION

In this study, with a Dynamic Financial Analysis model that is used parameters which is necessary for a non-life insurance company, a simulation study has been done. The dependency between investments and claims has been integrated into the model using the copulas. Different copulas have been used in DFA. The correlations between investments and between claims have been varied. In line with the simulation study, these dependencies' effects on the insurer's risk and return profile, the default risk and the ruin probability have been evaluated. Simulation was done with MATLAB programming language, 100.000 iterations.

**Key Words:** Dynamic Financial Analysis, non-linear dependency, copulas, nonlife insurance, simulation

## TÜRKİYE ÖLÜMLÜLÜK ORANLARININ DÜZLEŞTİRİLMESİ VE TAHMİN EDİLMESİ

Ezgi NEVRUZ\*

Hacettepe Üniversitesi  
Fen Fakültesi  
Aktüerya Bilimleri Bölümü  
06800-Çankaya, Ankara, Türkiye  
[ezginevruz@hacettepe.edu.tr](mailto:ezginevruz@hacettepe.edu.tr)

Sema TÜZEL

Hacettepe Üniversitesi  
Fen Fakültesi  
Aktüerya Bilimleri Bölümü  
06800-Çankaya, Ankara, Türkiye  
[sematurzel@hacettepe.edu.tr](mailto:sematurzel@hacettepe.edu.tr)

Erengül (ÖZKÖK) DODD

[erengulozkok@gmail.com](mailto:erengulozkok@gmail.com)

### Ölümlülük Oranlarının Modellenmesi

Son yüzyılda ölümlülük azalmaya eğilimli olduğundan, bu konu literatürde geniş olarak çalışılmaktadır. Gompertz-Makeham dağılımı en yaygın kullanılan ölümlülük kanunu iken ölümlülüğü düzleştirmek için eğri (spline) kullanımı son yıllarda artmıştır. Yaşlanma ve ölümlülük, ölümlülüğün yetişkin bireylerde yaşla birlikte ivme kazandığı Gompertz-Makeham dağılımı ile modellenilebilir.

Çalışmada öncelikle, uygulamada kullanılan Türkiye nüfus verisi anlatılmakta, daha sonra ölümlülüğün modellenmesinde kullanılan eğriler ve Gompertz-Makeham modeli teorik olarak tanımlanmaktadır. Son olarak, gelecek nüfusun tahmin edilebilirliği ve gerçek değerlerle karşılaştırılması tartışılmaktadır.

### Eğriler ile düzleştirme:

Son yıllarda B-eğrileri, P-eğrileri ve doğal eğriler düzleştirmede sıklıkla kullanılmaktadır. Genelleştirilmiş lineer modeller (GLM) yerine, genelleştirilmiş toplamsal modeller (GTM) kullanmak modellemedeki esnekliği artırmaktadır [4]. Bu çalışmada 0 yaşı modellemek için, uygun bir serbestlik derecesi seçilerek doğal kübik eğrilerle GTM kullanılacaktır.

$y_i$  yanıt değişkeni iken  $\mu_i = E[y_i]$  üstel dağılım ailesinden olmak üzere;

$$g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots \quad (1)$$

biçimindeki tam parametrik GLM model yapısı,

$$g(\mu_i) = \beta_0 + s_1(x_{1i}) + s_2(x_{2i}) + \dots \quad (2)$$

ile gösterilen GTM yapısı ile değiştirilmektedir. (1) ve (2) eşitliklerinde  $g$  bilinen monoton türevlenebilir bağ fonksiyonunu,  $s_i$ 'ler bilinmeyen düzleştirme fonksiyonlarını ve  $\beta_i$ 'ler ise tahmin edilen parametreleri göstermektedir.

Eğri fonksiyonlarının GTM'de düzleştirme fonksiyonları olarak kullanıldığı yapı ise

$$f(x) = b_0 + b_1 \phi_1(x) + b_2 \phi_2(x) + \dots + b_k \phi_k(x) \quad (3)$$

biçimindedir. Eşitlikte  $f(x)$  düzenli olarak birleşen  $k$  tane kübik eğrinin keyfi bir fonksiyonudur ve  $\phi_1(x), \phi_2(x), \dots, \phi_k(x)$  ise baz terimlerdir.

### Gompertz-Makeham ile modelleme:

Ölümlülüğün matematiksel bir formülle düzleştirilmesi bilinen bir yöntemdir. [2] tarafından önerilen  $(r,s)$  sıralaması ile Gompertz-Makeham modeli veya  $GM(r,s)$  dağılımı

$$GM_{\alpha}^{r,s}(x) = \sum_{i=1}^r \alpha_i x^{i-1} + \exp \left\{ \sum_{i=r+1}^{r+s} \alpha_i x^{i-r-1} \right\} \quad (4)$$

biçiminde ifade edilebilir. Eşitlik (4)'de  $r$  ve  $s$  negatif olmayan tamsayıdır ve ikisi birlikte sıfır değeri alamaz. Burada  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{r+s})$  katsayılar vektörünü,  $x$  ise yaşı göstermektedir. Yıl ve cinsiyet değişkenleri ile yaş-yıl, yaş-cinsiyet ve yıl-cinsiyet etkileşimleri bu modele [3]'te gösterildiği şekilde açıklayıcı değişken olarak eklenmiştir.

Bu çalışmada, [1]'de olduğu gibi ölüm sayılarının  $(D_x)$  her bir yaş ve takvim yılı için  $\lambda E_x^c$  parametresiyle Poisson dağılımlı olduğu varsayılmıştır. Burada  $\lambda$  ölümlülük oranı iken  $E_x^c$  merkezi riske maruz kişi sayıdır.

GM(r,s) modelde parametre tahminleri Poisson varsayımı altında en çok olabilirlik yöntemiyle elde edilmiş, her bir yaş grubu için uygun model seçiminde ise Bayes bilgi kriteri (BIC) kullanılmıştır.

### KAYNAKLAR

- [1] Brouhns N., Denuit M., Vermunt J. K. (2002), *A Poisson log-bilinear regression approach to the construction of projected life tables*, Insurance: Mathematics and Economics, Vol. 31, 373-393.
- [2] Forfar D. O., McCutcheon J. J., Wilkie A. D. (1988), *On graduation by mathematical formula*, Journal of the Institute of Actuaries, Vol. 115, 1-149.
- [3] Ozkok E., Streftaris G., Waters H. R., Wilkie A. D. (2012), *Modelling critical illness claim diagnosis rates I: methodology*, Scandinavian Actuarial Journal, DOI: 10.1080/03461238.2012.728537, 1-19.
- [4] Wood S.N. ve N.H. Augustin (2002), *GAMs with integrated model selection using penalized regression splines and applications to environmental modelling*, Ecological Modelling, Vol. 157, 157-177.

### ABSTRACT

#### SMOOTHING AND FORECASTING TURKISH MORTALITY RATES

*As human mortality has improved substantially over the last century, smoothing and forecasting mortality rates become a problem of fundamental importance for life insurance companies. Their future cash flows depend on future mortality rates, therefore smoothing the observed mortality rates and long periods of mortality forecasts are required.*

*In this study, we model Turkish mortality using the data between 1938 and 1995. The infant mortality is modelled employing splines while the mortality rates for ages 1 to 84 is modelled using a Gompertz-Makeham model. Under this model we assume the number of deaths have a Poisson distribution. Model parameters are estimated by the maximum likelihood method, model selection is performed and the projection results are discussed.*

**Key Words:** Forecasting, Gompertz-Makeham model, mortality, splines.

## ÖLÜMLÜLÜK GELİŞİM ORANLARININ MODELLENMESİ

Funda KUL<sup>a</sup>, Övgücan G. KARADAĞ<sup>b</sup>, Meral SUCU<sup>c</sup>

<sup>a</sup>fundakaraman@hacettepe.edu.tr, Hacettepe University, Department of Actuarial Science, Ankara, Turkey  
<sup>b</sup>oygucan@hacettepe.edu.tr, Hacettepe University, Department of Actuarial Science, Ankara, Turkey  
<sup>c</sup>msucu@hacettepe.edu.tr, Hacettepe University, Department of Actuarial Science, Ankara, Turkey

Bu çalışmada ölümlülük gelişim oranlarının normal dağılıma uyduğu varsayımı altında, kadın ve erkekler için 1980-2012 yılları arasındaki Türkiye ölümlülük verisi kullanılarak Lee-Carter ölümlülük modeli yardımıyla beklenen yaşam süresi için tahminler elde edilmiş ve bulunan sonuçlar yorumlanmıştır.

**Anahtar Kelimeler:** Lee-Carter, Ölümlülük Gelişim Oranı, Zaman Serisi, Ölümlülük Tahmini.

### 1. Giriş

Yaşa özel ölümlülük oranlarının modellenmesinde sıklıkla kullanılan Lee-Carter (1992) modeli,

$$\log(m_{x,t}) = \alpha_x + \beta_x K_t + \varepsilon_{x,t} \quad (1)$$

şekindedir [4]. Bu modelde yaşa özel ölüm hızının logaritması ( $\log m_{x,t}$ ); (x) yaşı, (t) takvim yılını,  $\alpha_x$  zamandan bağımsız yaşa özel bileşeni,  $\beta_x$  ölümlülük düzeyinin yaşa göre değişimini ve zamana bağlı ölümlülük düzeyini gösteren  $K_t$  bileşenleri ile modellenmektedir. Lee-Carter ölümlülük modelinde sabit terimli rastgele yürüyüş modeli kullanılarak geleceğe yönelik ölümlülük oranları tahmini yapılmaktadır. Lee-Carter modelinden sonra birçok ölümlülük modeli geliştirilmiştir [1].

Bu çalışmada, Lee-Carter modeli kullanılarak takvim yılı ve yaşlara göre Türkiye için ölümlülük gelişim oranları elde edilerek modellenmeye çalışılacak ve elde edilen sonuçlara göre gelecekte beklenen yaşam sürelerindeki değişim incelenecektir.

### 2. Yöntem

Bu çalışmada, 1980-2012 yılları arasında 20-85 yaş aralığında beşerli yaş gruplarında Türkiye İstatistik Kurumu (TÜİK) tarafından derlenen kadın ve erkekler için yaşayan ve ölen kişi sayıları kullanılacaktır.

Ölüm sayısı (d) ve ölüm riskine maruz kalan birim sayısı (e) ile gösterilmiştir. (w) ağırlık gösterimi ise (d) ve (e) verisinden herhangi birinin olmaması durumunda 0 ve her ikisinin bulunması durumunda 1 değerini almaktadır. Veri kümesi kullanılarak merkezi ölüm hızı,

$$\hat{m}_{x,t} = \frac{d_{x,t}}{e_{x,t}} \quad (2)$$

eşitliğinden elde edilecektir. Merkezi ölüm hızı kullanılarak ölümlülük gelişim oranları ise,

$$\hat{z}_{x,t} = 2 \frac{(1 - \hat{m}_{x,t} / \hat{m}_{x,t-1})}{(1 + \hat{m}_{x,t} / \hat{m}_{x,t-1})} \quad (3)$$

şeklinde bulunur [3]. Burada, ölümlülük gelişim oranlarında ilk yıl için tüm yaşlardaki ağırlık değerleri sıfır olarak alınmış ve buna bağlı olarak ölümlülük gelişim oranları da tüm yaşlar için sıfır olarak alınmıştır.

Ölümlülük gelişim oranlarını gösteren  $z_{x,t}$ 'nin değişken yayılımına sahip birbirinden bağımsız Normal dağılıma uyduğu varsayımı altında beklenen değer ve varyansı şu şekildedir [2].

$$E(z_{x,t}) = \eta_{x,t} = \alpha_x + \beta_x K_t \quad (4)$$

$$Var(z_{x,t}) = \frac{\phi_{x,t} \sigma^2}{w_{x,t}} \quad (5)$$

Yıldırım (2010) çalışmasında, Türkiye ölümlülük verisi kullanılarak Lee-Carter modelinde  $K_t$  için en uygun zaman serisi modelinin ARIMA(1,1,0) olduğu belirtilmiştir [5]. Bu durumda ölümlülük gelişim oranları için en uygun zaman serisi modeli ARMA (1,0) olduğu görülmektedir [3].  $K_t$  için zaman serisi modeli,

$$K_t - \mu = \varphi(K_{t-1} - \mu) + \varepsilon_t \quad (6)$$

biçimindedir. Burada artık süreci, sıfır ortalama ve  $\tau^2$  varyansı ile Normal dağılıma sahiptir. Bu durumda  $K_{t_n+j}$  tahminleri de şu şekilde elde edilir.

$$K_{t_n+j} = \mu + \varphi^j (K_{t_n} - \mu), \quad j=1,2,3,\dots \quad (7)$$

### 3. Sonuçlar

Ölümlülük gelişim hızları için oluşturulan model kullanılarak geleceğe ilişkin yaşa özel ölümlülük oranları tahmin edilerek, beklenen yaşam süreleri için öngörü değerleri elde edilecektir.

### KAYNAKLAR

- [1] Haberman, S. And Renshaw ,A.E, (2011), A comparative study of parametric mortality projection models, Insurance: Mathematic and Economics 45, 255-270.
- [2] Haberman, S. And Renshaw ,A.E, (2012), Parametric mortality improvement rate modeling and projecting, Insurance: Mathematic and Economics 50, 309-333.
- [3] Haberman, S. And Renshaw ,A.E, (2013), Modelling and projecting mortality improvement rates using a cohort perspective, Insurance: Mathematic and Economics 53, 150-168.
- [4] Lee R.D., Carter L., (1992), Modelling and forecasting the time series of US mortality, Journal of the American Association 87, 659-671.
- [5] Yıldırım F., (2010), Türkiye Ölümlülük Yapısının Lee-Carter ve Bulanık Lee-Carter İle Modellenmesi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü Yüksek Lisans Tezi.

### ABSTRACT

#### MODELLING OF MORTALITY IMPROVEMENT RATES

In this paper, the estimates of life expectancies are found and the results are interpreted by using Lee-Carter mortality model for Turkish male and female mortality data from 1980-2012 years, under the assumption of mortality improvement rates fits Normal distribution.

**Key Words:** Lee-Carter, Mortality Improvement Rate, Time Series, Mortality Forecasting.

YAŞAM ÇÖZÜMLEMESİNDE ZAYIFLIK MODELİ VE TÜRKİYE İÇİN BİREYSEL EMEKLİLİK  
VERİLERİYLE BİR UYGULAMA

Güven ŞİMŞEK  
Hacettepe Üniversitesi, Fen Fakültesi, Aktüerya Bilimleri Bölümü, 06800, Ankara, Türkiye  
guvensimsek@hacettepe.edu.tr

Durdu KARASOY  
Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800, Ankara, Türkiye  
durdurdu@hacettepe.edu.tr

1. Bireysel Emeklilik Sistemi

Sosyal güvenlik alanında yaşanan sorunlar ülkeler için sosyal güvenlik sistemine ilişkin reform çalışmalarının yapılmasını gündeme getirmiştir. Bu konuda en radikal değişiklik emeklilik aylığı sistemini, kamu kesiminden özel emeklilik fonlarına dayalı bir sisteme dönüştüren Şili'de gerçekleşmiştir [1]. Bu model özel emekliliğe geçiş konusunda birçok ülke için referans niteliği taşımaktadır.

Türkiye'de ise 2001 yılında 4632 sayılı Bireysel Emeklilik Tasarruf ve Yatırım Sistemi Kanunu ile bireysel emeklilik sistemi kurularak, sosyal güvenlik alanında önemli bir adım atılmıştır. Sistem, bireylerin emekliliğe yönelik tasarruflarının artırılmasını ve emeklilik döneminde refah düzeyinin yükseltilmesini amaçlamaktadır.

Bu çalışmada, bireysel emeklilik sisteminde kalma süresini etkileyen faktörlerin incelenmesi için Emeklilik Gözetim Merkezi'ne ait bireysel emeklilik verileri kullanılmıştır. Verilere, orantısız tehlikeler durumunda kullanılan zayıflık modeli uygulanmış ve sonuçlar yorumlanmıştır.

2. Zayıflık Modeli

Yaşam çözümlemesinde tehlike fonksiyonunu etkileyen açıklayıcı değişkenleri belirlemek ve birime ait tehlike fonksiyonunu elde etmek için kullanılan temel model orantılı tehlikeler modelidir. Cox [2] tarafından önerildiğinden Cox regresyon modeli olarak da adlandırılmaktadır. Orantılı tehlikeler varsayımının sağlanmaması durumunda genişletilmiş Cox regresyon modeli, parametrik regresyon modelleri ve zayıflık modeli kullanılmaktadır.

Benzer özelliklere sahip birimler arasında yaşam sürelerindeki farklılıkları açıklamak için zayıflık modeli kullanılmaktadır. Vaupel v.d. [3] tarafından yapılan mortalite çalışmalarında zayıflık modeli kullanılmıştır. Daha sonra Lancaster [4] işsizlik sürelerinin modellenmesi için zayıflık modelini incelemiştir.

Zayıflık modeli, paylaşılmış ve paylaşılmamış zayıflık modelleri olmak üzere ikiye ayrılmaktadır. Birimler arasındaki heterojenliği modellemek için paylaşılmamış zayıflık modeli kullanılırken, gruplar arasındaki heterojenliği modellemek için paylaşılmış zayıflık modeli kullanılmaktadır.

Paylaşılmamış zayıflık modeli için kullanılan tehlike fonksiyonu ölçülemeyen rastgele etkiyi (zayıflık terimini) içermektedir. Bu model için koşullu tehlike fonksiyonu,

$$h(t / \alpha) = \alpha h(t) \quad (1)$$

biçiminde gösterilmektedir. Zayıflık terimi ( $\alpha > 0$ ), birim ortalamaya ve  $\theta$  varyansa sahiptir. Paylaşılmamış zayıflık modeli için yaşam fonksiyonu,

$$S(t / \alpha) = [S(t)]^{\alpha} \quad (2)$$

biçimindedir. Gamma ya da ters Gaussian, zayıflık terimi için en yaygın kullanılan dağılımlardır.

Birimlerin aynı zayıflık terimini paylaşmalarına izin veren paylaşılmış zayıflık modelinde birimler arasında bağımlılık ortaya çıkmaktadır. n gruptan oluşan bir yaşam verisi için paylaşılmış zayıflık modeli,

$$h_{ij}(t / \alpha_i) = \alpha_i h_{ij}(t), \quad j = 1, 2, \dots, n_i \quad (3)$$

biçimindedir.  $n_i$ , i. grupta bulunan birim sayısıdır.  $h_{ij}(t) = h(t/x_{ij})$  biçiminde olmak üzere, i gruptaki herhangi bir birim için, standart tehlike fonksiyonu paylaşılmış zayıflık  $\alpha_i$  ile çarpılmaktadır. Yaşam fonksiyonu ise aşağıdaki gibidir [5]:

$$S_{ij}(t / \alpha_i) = [S_{ij}(t)]^{\alpha_i} \quad (4)$$

KAYNAKLAR

- [1] Gillion, C., Bonilla, A. (1992), Analysis of a national private pension scheme: The case of Chile, International Labour Review, Sayı 2, Cilt 131, 171-195.
- [2] Cox, D.R. (1972), Regression models and life-tables, Journal of the Royal Statistical Society, Series B, 34, 187-220.
- [3] Vaupel, J.W., Manton, K., Stallard, E. (1979), The impact of heterogeneity in individual frailty on the dynamics of mortality, Demography, 16, 439-454.
- [4] Lancaster, T. (1979), Econometric methods for the duration of unemployment, Econometrica, 47, 939-956.
- [5] Ata, N., Karasoy, D. (2011), Sağlık Çözümü için Zayıflık Modeli ve Mide Kanseri Hastalarına İlişkin Verilerle Bir Uygulama, Çankaya University Journal of Science and Engineering, 8, 2, 225-235.

ABSTRACT

FRAILTY MODEL IN SURVIVAL ANALYSIS, WITH AN APPLICATION TO INDIVIDUAL PENSION DATA FOR TURKEY

In Turkey in late 1990s, problems emerged in social security field like the disruption of balance of income and expenditure, decrease in asset-liability ratio. As a solution to these problems, individual pension system has been put into practice as a supplementary to the existing social insurance system.

In this study, We used the Emeklilik Gözetim Merkezi's data on the existing individual pension data in researching the factors that influence the time spent within the individual pension system. Frailty models are applied for the data and the attained results are discussed.

**Key Words:** Social security, individual pension, Cox regression model, parametric regression model, frailty model

LINKAGE BETWEEN EDUCATION, ECONOMIC GROWTH AND CO<sub>2</sub> EMISSION: A  
MULTIVARIATE ANALYSIS ON TURKEY

K. Demirberk ÜNLÜ<sup>1</sup>, A. Sevtap SELCUK-KESTEL<sup>2</sup>  
<sup>1-2</sup>Middle East Technical University  
Institute of Applied Mathematics  
06800 ANKARA  
<sup>1-2</sup>demirberg@gmail.com, skestel@metuedu.tr,

BİLDİRİ OTURUMLARI 5

SESSION 5

Applied Statistics 2

EXTENDED SUMMARY

Kyoto Protocol is an international agreement, which aims to fight against global warming and climate changes. The main reason of the global warming can be shown as greenhouse gases (GHG). GHG consist of 4 major gases such as Carbon dioxide (CO<sub>2</sub>), Methane (CH<sub>4</sub>), Nitrous oxide (N<sub>2</sub>O) and Fluorinated gases (F- gases). According to the United States Environmental Protection Agency (EPA), the economic activities that release GHG are: energy supply (26%), industry (19%), land use, land-use change and forestry (17%), agriculture (14%), transportation (13%), commercial and residential building (8%) and waste and wastewater (3%). Participants of this protocol must fulfill their target GHG emission. By 2013, 192 countries have signed the agreement. One of the participants of this agreement is Turkey who signed the agreement in 2009. In recent years Turkey has made significant changes in her education system. Compulsory education has been increased to 12 years. Also Turkey has the highest rate of young population among the European countries so our aim in this study is to investigate the casual relationship between education, carbon emission and economic growth.

In his study [1] found a unidirectional causality between income and CO<sub>2</sub> emission while [4] cannot find any evidence of causality between these two variables. The relationship between education and income in Turkey was explored by [3]. Since education causes a higher level of income and this income may cause carbon emission, we employed the Johansen-Juselius [2] cointegration method to investigate the causality between education, gross domestic product and carbon emission. For the period between 1971 and 2010, all in natural logarithms, primary school enrolment (PRM), secondary school enrolment (SEC), tertiary enrolment (TER), gross domestic product (GDP) and CO<sub>2</sub> emission (CO<sub>2</sub>) are used. The result of the cointegration test, which is given in Table 1, indicates two cointegration vectors.

Table 1. Cointegration Test Results (lag = 3)

Null Hypot.	Trace Stat	5%	Max-Eigen Stat	5%	Eigenvalues
None	103.5069*	68.52	46.2340*	33.46	0.7231
At most 1	57.2730*	47.21	29.0521*	27.07	0.5538
At most 2	28.2208	29.68	15.5320	20.97	0.3504
At most 3	12.6889	15.41	10.0779	14.07	0.2442
At most 4	2.6110	3.76	2.6610	3.76	0.0699

The empirical results show that GDP does not cause CO<sub>2</sub> emission but in the short-run and in the long-run SEC and TER cause CO<sub>2</sub>. Significant and negative signed error correction term (ECT) indicates increase in SEC or increase in TER causes increase in CO<sub>2</sub>. Also all level of education provides a positive contribution to GDP.

REFERENCES

[1] Halicioglu F. (2009), An Econometric study of CO<sub>2</sub> emissions, energy consumption, income and foreign trade in Turkey, Energy Policy, 37, 1156-1164.

[2] Johansen, S. and Juselius, K. Maximum likelihood estimation and inference on cointegration - with applications to the demand for money, Oxford Bulletin of Economics and Statistics, 52, 169-210

[3] Sari, R. and Soytas, U. (2006), Income and education in Turkey: A multivariate analysis, Education Economics, 2, 181-196.

[4] Soytas, U. and Sari, R. (2009), Energy consumption, economic growth and carbon emissions: Challenges faced by an EU candidate member, Ecological Economics, 68, 1667-1675.

#### ABSTRACT

#### LINKAGE BETWEEN EDUCATION, ECONOMIC GROWTH AND CO<sub>2</sub> EMISSION: A MULTIVARIATE ANALYSIS ON TURKEY

This study investigates the long-run relationship between carbon emission, economic growth and education in Turkey for the period between 1971 and 2010 by using Johansen- Juselius method and Vector Error Correction model. As a second step in this study Impulse-Response function and Generalized Forecast Error Variance Decomposition test are utilized.

**Key Word:** CO<sub>2</sub> emission, economic growth, education, Turkey, vector error correction model.

#### FRACTAL DIMENSION OF CEMENT PASTES MICROSTRUCTURE BASED ON STATISTICAL POWER OF REGRESSION LINES COMPARISON

Ömer Utku Erzengin<sup>1</sup>, Siddika Gamze Erzengin<sup>2</sup>

<sup>1</sup>Süleyman Demirel University, Department of Statistics, Civilized East Campus, Isparta, Türkiye, omererzengin@sdu.edu.tr

<sup>2</sup>Süleyman Demirel University, Chemical Engineering Department, Wild West Campus, Isparta, Türkiye, gamzeerzengin@sdu.edu.tr

Analysing the geometric microstructural elements in hydrated cement systems carry a great role in daily functionality. In order to reveal these systems formation, Scanning Electron Microscopy (SEM) is one of the important method of characterization of cement paste microstructure according to its surface. Surface of the cement paste is made up of many elements, pores are one of the fundamental element in the microstructure. Gel phase of the cement composes pores whose formation could be expressed by different mathematical and statistical methodology. Chemical admixtures which are polymelamine sulphonate, modified sodium lauryl sulphate, urea formaldehyde and triethanolamine used in this study also play important role in cement pores formation.

Gel phase of cement paste could be explained by Euclid geometry but in order to interpret entire processes classical Euclid geometry is not enough. One of the alternative method is fractal analysis assessing fractal characteristics of data<sup>1</sup>. Fractal analysis aim is to find fractal dimension and other fractal characteristics from a dataset that is extracted from phenomena including natural geometric objects.

Fractal dimension of cement paste surface geometry "D" could be estimated. Fractal dimension "D" is between  $0 < D < 3$  as distinct from Euclidian dimension. Fractal dimension does not have to be an integer. A fractal dimension is a ratio providing a statistical index of complexity comparing how detail in a pattern changes with the scale at which it is measured.<sup>2</sup>

Generally as term fractal is associated with geometrical objects with two topic: self-similarity and fractional dimensionality. Self-similarity means that an object is composed of sub-units and sub-sub-units on multiple levels that (statistically) take after the structure of the main object. When sub-units, those are independent of scale, are magnified they resemble the main object.

Revealing fractal dimension of cement paste pores, SEM image analysis, based on basically box counting algorithm is used. The basic procedure is systematically lay a series of grids of decreasing the boxes over an image and count the data for each successive box. Several grids of decreasing box size are placed over an SEM image and the number of boxes that contain pixels is counted for each grid. Data are gathered for each box of every grid, fractal dimension can be expressed as ordinary linear regression slope.

In this study, fractal dimension of cement paste pores (with different chemical admixtures) observed by SEM were estimated. The aim of the study is to compare fractal dimension of cement paste pores in the existence of different chemical admixtures. After applying ANOVA, no significant statistical difference was found between fractal dimensions ( $p > 0.7$  comparison regression lines slopes<sup>3</sup>). Cement gel pores' fractal dimensions are polymelamine sulphonate=1.771, polymelamine sulphonate and modified sodium lauryl sulphate =1.823, urea formaldehyde=1.847. In the presentation also statistical power of the fractal dimensions (regression slopes) comparison will be mentioned.<sup>5</sup>

**Key Words:** Fractal Dimension, Image Analysis, Regression, Power

#### References

[1] Wang Y. S. (2001), A fractal study of the fracture surfaces of cement pastes and mortars using a stereoscopic SEM method, Diamond Cem. Concr. Res. 31, 1385-1392

[2] Mandelbrot B. B. (1984), Les objects fractal, Paris, Flammarion.

[3] Korpa A. and Trettin R. (2006) The influence of different drying methods on cement paste microstructures as reflected by gas adsorption: comparison between freeze-drying (F-drying), D-drying, P-drying and oven-drying methods. Cem. Concr. Res. 36, 634-649

[4] Clogg C. C., Petkova E., Haritou A. (1995) Statistical Methods for Comparing Regression Coefficients Between Models. University of Chicago Press. AJS 100-5, 1261-1293.

[5] Zar J. H. (2010). Biostatistical Analysis. 5th Edition. Upper Saddle River - NJ, Pearson Prentice-Hall.

## A NEW SCALARIZATION METHOD IN NONLINEAR AND NONCONVEX MULTI-OBJECTIVE OPTIMIZATION

Refail KASIMBEYLİ

Department of Industrial Engineering, Anadolu University, İki Eylül Campus, 26555 Eskişehir  
rkasimbeyli@anadolu.edu.tr

### Introduction

In general, scalarization means the replacement of a multi-objective optimization problem by a suitable scalar optimization problem which is an optimization problem with a real valued objective functional. Since the scalar optimization theory is widely developed, scalarization turns out to be of great importance for the multi-objective optimization theory.

One way of constructing a single-objective optimization problem is through scalarizing functions involving possibly some parameters or additional constraints. Multi-objective optimization methods utilize different scalarizing functions in different ways.

In a large variety of methods, norms have usually been used to identify the minimal (or nondominated) elements that are closest to some (reference) point. In particular, the family of  $l_p$  norms has been studied extensively by many researchers [1].

It would be impractical to supply a complete bibliography of studies in the literature on multi-objective optimization, multicriteria analysis and scalarization methods; instead we refer the reader to the books [1,2].

By summarizing the properties of different scalarization techniques, the following questions related to the quality or performance of the technique can be drawn:

- Does this method always yield minimal (or efficient) elements?
- If so, can all minimal (or efficient) elements be detected in this way?
- Will the preference and reference point information of the decision maker be taken into consideration by this technique?

The simple investigation of different techniques may show that not all methods can positively answer all the questions raised here.

The conic scalarization method presented in this paper is shown to answer positively all these questions.

In this paper, a new kind of properly efficient points has been introduced, and a full characterization of the conic scalarization method is presented. It is shown that the zero sublevel set of every function from the suggested class of monotone sublinear functions is a convex closed and pointed cone containing the negative ordering cone. We introduce the notion of a separable cone and show that two closed cones (one of them is separable) having only the vertex in common can be separated by a zero sublevel set of some function from this class. It is shown that the scalar optimization problem constructed by using the monotonically increasing sublinear functions, enables to completely characterize the whole set of efficient and properly efficient solutions of multi-objective problems without convexity and boundedness conditions.

### Conic Scalarization Method

Consider a multi-objective optimization problem:

$$\min_{x \in X} [f_1(x), \dots, f_n(x)]$$

Choose preference parameters:

- Weight vector  $w = (w_1, \dots, w_n)$ , where  $w_i > 0$  denotes the preference degree of  $i$  th objective function for decision maker.
- Reference point  $r = (r_1, \dots, r_n)$ . Such a point may be identified by a decision maker in cases when he or she desires to calculate minimal elements that are close to some point. The conic scalarization method does not impose any restrictions on the ways for determining reference points. These points can be chosen arbitrarily.



Choose an augmentation parameter  $\alpha$  such that  $\alpha < \min\{w_1, \dots, w_n\}$ . Then, the scalar problem for the given parameters  $(w, \alpha)$  and  $r$  is:

$$\min_{x \in X} [w_i(f_i(x) - r_i) + \alpha \sum_{i=1}^n |f_i(x) - r_i|]$$

By solving this scalar problem for "all" possible values of the augmentation parameter  $\alpha$  between zero and  $\min\{w_1, \dots, w_n\}$ , one can calculate all the efficient solutions corresponding to the decision maker's preferences (the weighting vector  $w = (w_1, \dots, w_n)$  and the reference point  $r$ ) [3,4,5].

#### REFERENCES

- [1] Ehrgott, M. (2005), Multicriteria Optimization. Springer, Berlin.  
 [2] Eichfelder, G. (2008), Adaptive Scalarization Methods in Multi-Objective Optimization. Springer, Berlin.  
 [3] Gasimov, R.N. (2001), Characterization of the Benson proper efficiency and scalarization in nonconvex vector optimization, In: Koksalan, M., Zions, S. (eds.) Multiple Criteria Decision Making in the New Millennium, Book Series: Lecture Notes in Economics and Mathematical Systems, vol. 507, pp. 189–198.  
 [4] Kasimbeyli, R. (2010), A nonlinear cone separation theorem and scalarization in nonconvex vector optimization, SIAM J. Optim. 20, pp.1591–1619.  
 [5] Kasimbeyli, R. (2013), A conic scalarization method in multi-objective optimization, Journal of Global Optimization, 56 (2), pp. 279-297.

#### ABSTRACT

#### A NEW SCALARIZATION METHOD IN NONLINEAR AND NONCONVEX MULTI-OBJECTIVE OPTIMIZATION

This paper presents the conic scalarization method for scalarization of nonlinear multi-objective optimization problems. By this method, choosing a suitable scalarizing parameter set consisting of a weighting vector, an augmentation parameter, and a reference point, decision maker may guarantee a most preferred efficient or properly efficient solution.

**Key Words:** Conic Scalarization Method, Multi-Objective Optimization, Proper Efficiency.

#### TURKISH ELECTRICITY PRICE BEHAVIOR BASED ON SOME ECONOMIC INDICATORS: AN ECONOMETRIC APPROACH

Aseña ÖZDEMİR<sup>1</sup>, Demirberk ÜNLÜ<sup>2</sup>, A. Sevtaş SELCUK-KESTEL<sup>3</sup>, Kasirga YILDIRAK<sup>4</sup>

<sup>1-3</sup>Middle East Technical University  
Institute of Applied Mathematics  
06800 ANKARA

<sup>4</sup>Trakya University  
Department of Economics  
EDİRNE

<sup>1-3</sup>Aseña.ozdemir@metu.edu.tr, demirberg@gmail.com, skestel@metuedu.tr, <sup>4</sup>kasirga@metu.edu.tr

#### EXTENDED SUMMARY

As a consequence of social and economic development in Turkey in last decades, energy demand and especially, demand for electricity has increased rapidly. In order to meet this increasing demand it is important to analyze and control electricity supply. This study aims to investigate the cointegrated behavior between annual electricity consumption, production in Turkey and macroeconomic indicators based on the studies in the literature [2,3]. It examines the stochastic behavior of changes in these parameters in time domain by employing the techniques from Time Series Analysis. The cointegration analysis, the autoregressive distributed lag approach (ARDL) [1], error correction method, impulse response functions and variance decomposition analyses are performed to understand the association and impact of these variables on the electricity demand in the Turkish electricity market. The main features of this study are the implementation of such wide range of economic indicators to determine the electricity demand and supply over period 1970-2010 which have witnessed many new changes in the regulations and many economic crises. Data set contains annually collected electricity consumption (CONS), Gross domestic product (GDP), urban population (URB), the electricity power transmission and distribution losses (LOS), the industry value added (IND), the average annual electricity prices (PRC), Turkish Population (POP), total annual electricity production (PRD), labor supply in electricity sector (LBR), gross profit for Turkish electricity market (PRF), investments in Turkish electricity sector (INV) and average annual electricity prices (PRC) collected from Turkish Electricity Transmission (TEIAS) and World Bank. are the variables used in the proposed analysis. All these are modified to per capita by adjusting each data set to population except the prices and gross profit.

The ARDL bounds testing procedure is employed to determine the cointegration variables (Table 1) to determine long run relationship between them through cointegration for electricity demand and supply. The tests results for demand is reduced to 4 cointegrating relationships when dependent variables are selected as; CONS, GDP, IND and mean adjusted PRC. However, there exist 6 cointegrating relationship for the supply when dependent variables are selected as; PROD, GDP, INV, LOS, LBR and PRF whereas; no cointegrating relationship exists if the dependent variable is mean adjusted annual PRC. Based on the significancy with respect to Bounds-Test, cointegrated relation of the annual electricity production are expressed as follows:

$$\widehat{CONS} = -627.1865 + 0.0020823GDP - 0.50248IND + 1.9291LOS - 0.17769PRC + 1977.1URB$$

$$\widehat{PRD} = 237.672 + 0.7032GDP - 0.041482INV + 5.1507LOS - 67.5671PRC + 384860.9LBR + 109.1725PRF$$

Table 1. Bounds-Testing procedure results

Demand		Supply	
Cointegration hypothesis	F-statistics	Cointegration hypothesis	F-stat.
F(CON GDP,IND,LOS,PRICE,URB)	3.1012**	F(PRD GDP,INV,LOS,PRC,LBR,PRF)	3.5164*

F(GDP CON,IND,LOS,PRICE,URB)	6.3478*	F(GDP PRD,INV,LOS,PRC,LBR,PRF)	3.8908**
F(IND CON,GDP,LOS,PRC,URB)	7.2093*	F(INV GDP,PRD,LOS,PRC,LBR,PRF)	6.0427*
F(LOS CON,GDP,IND,PRICE,URB)	18.595	F(LOS GDP,PRD,INV,PRC,LBR,PRF)	2.7166***
F(PRC CON,GDP,IND,LOS,URB)	5.5008*	F(PRC GDP,PRD,INV,LOS,LBR,PRF)	11.212
F(URB CON,GDP,IND,LOS,PRC)	0.88845	F(LBR GDP,PRD,INV,LOS,PRC,PRF)	7.4945*
* significance at 1%, ** at 2.5% levels [1]		F(PRF GDP,PRD,INV,LOS,PRC,LBR)	5.8336*

The empirical findings indicate that there exist cointegration between electricity consumption and annual average electricity losses, GDP, urbanization ratio, industry value added and electricity transmission and distribution losses. Based on the selection method criteria electricity consumption, industry value added, annual average electricity prices and GDP are chosen as the dependent variables. Moreover, the direction and sign of the relationships determined by the error correction models yield results as expected indicating the appropriateness of the model. In addition, impulse response graphs also catch the impact of variables selected to the electricity consumption on the time horizon. Furthermore, variance decomposition implies electricity consumption mainly affected by its own shock and secondly, it is affected mainly by shocks in GDP and industry value added. Moreover, for the comparison of equilibrium prices and actual prices it can be said that, they are slightly different because still in Turkey electricity prices are determined through government intervention rather than equilibrium prices formed at market equilibrium points.

## REFERENCES

- [1] Pesaran, M.H., Shin, Y., Smith, R.J.,2001. Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, 16, 289-326.
- [2] Soytaş, U., Sari, R., 2003. Energy consumption and GDP: Causality relationship in G-7 countries and emerging markets. *Energy Economics*, 25, 33-37.
- [3] Sari, R., Ewing B.T., Soytaş, U., 2007. The relationship between disaggregate energy consumption and industrial production in the United States: An ARDL approach, *Energy Economics*, 30, 2302-2313.

## ABSTRACT

### TRKISH ELECTRICITY PRICE BEHAVIOR BASED ON SOME ECONOMIC INDICATORS: AN ECONOMETRIC APPROACH

This paper investigates the association between main macroeconomic and sectorial indicators with electricity consumption and production. For estimating electricity demand, annual electricity consumption, Gross Domestic Product (GDP), electricity transmission and distribution losses, industry value added, annual average electricity prices and urbanization ratio are used as variables. On the other hand, for estimating electricity supply variables annual electricity production, gross profit and labor supply in electricity sector, electricity transmission and distribution losses, annual average electricity prices, GDP and investments in electricity sector are used. Both electricity consumption and production are estimated through the autoregressive distributed lag approach (ARDL) by using annual data for the period from 1970 to 2010 to determine the equilibrium prices in electricity market in Turkey.

**Key Words:** Turkish Electricity Market, ARDL, Demand, Supply, Equilibrium price.

## A COMPARATIVE SIMULATION STUDY FOR MODIFIED RANK SET SAMPLING METHODS

Bekir ÇETİNTAV\* Selma GÜRLER

Dokuz Eylül University, Faculty of Science, Department of Statistics., Tinaztepe 35160 Buca, Izmir, TURKEY

E-mail: bekir.cetintav@deu.edu.tr , Selma.erdogan@deu.edu.tr

Ranked Set Sampling (RSS) is first proposed by McIntyre in 1952 to effectively estimate the yield of pasture in Australia. The basic premise is a population under study and the assumption is that a set of sampling units drawn from the population can be ranked by certain means rather cheaply without the actual measurement of the variable of interest which is costly and / or time-consuming (Chen, 2003).

The RSS has an easy-to-use sampling procedure. First, the size of the sets ( $m$ ), number of cycles ( $n$ ) and sample size ( $N$ ) should be decided. These  $m \times n$  observations are called a standard ranked set sample.  $X_{(h)j}$  means  $h$ th unit in a specific set in the  $j$ th cycle and  $X_{[h]j}$  means the  $h$ th ranked unit from the  $j$ th cycle for  $h = 1, 2, \dots, m$  and  $j = 1, 2, \dots, n$ . The following illustration shows the procedure (for  $m = 3$ ), briefly.

Cycle 1

$$\begin{aligned} X_{(1)1} &\leq X_{(2)1} \leq X_{(3)1} \rightarrow X_{[1]1} \\ X_{(1)1} &\leq X_{(2)1} \leq X_{(3)1} \rightarrow X_{[2]1} \\ X_{(1)1} &\leq X_{(2)1} \leq X_{(3)1} \rightarrow X_{[3]1} \end{aligned}$$

...

Cycle  $n$

$$\begin{aligned} X_{(1)n} &\leq X_{(2)n} \leq X_{(3)n} \rightarrow X_{[1]n} \\ X_{(1)n} &\leq X_{(2)n} \leq X_{(3)n} \rightarrow X_{[2]n} \\ X_{(1)n} &\leq X_{(2)n} \leq X_{(3)n} \rightarrow X_{[3]n} \end{aligned}$$

Figure 1. Illustration of The RSS procedure

For the chosen ranked set sample, the estimator of  $\mu$  is

$$\hat{\mu} = \frac{1}{m} \sum_{h=1}^m \frac{1}{n} \sum_{j=1}^n X_{[h]j} = \frac{1}{m} \sum_{h=1}^m \bar{X}_{[h]}. \quad (1)$$

Takahasi et al.(1968) and Dell et al.(1972) studied the necessary mathematical background. They showed that the sample mean of the RSS is an unbiased estimator of the population mean and its variance is smaller than the sample mean of a simple random sample (SRS) even if the ranking is not perfect.

Several researchers, including Yanagawa and Shirahata (1976), Ridout and Cobby (1987), and Muttlak and McDonald (1990), proposed some modifications for the RSS procedure. The common point of all modified methods is using not every order statistic in each cycle; instead the researcher chooses which ones would be most useful. The errors in ranking process could decrease variance of RSS estimator and hence reducing its efficiency. Some of most using modified methods are introduced.

Extreme RSS method, could be called the first modified RSS method, was introduced by Samawi et al. (1996) to estimate population mean. The starting point is to be more practical method than Standard RSS ranked set sampling by using only the first and/or the last ordered unit (or the median for odd sample sizes). Al-Nasser and Mustafa (2009) developed Robust Extreme RSS to combine robust approach and Extreme RSS procedure. The Median RSS procedure was first proposed by Muttlak (1997). It can be described by selecting random sets of size  $m$  from the population and rank the units within each set as in RSS mechanism. Ibrahim, Syam and Al-Omari (2010) modified it for stratified case. Al-Saleh and Al-Kadiri (2000) first considered double ranked set sampling (DRSS). They investigated a two-stage ranking procedure for mean estimation. Al-Saleh and Al-Omari (2002) extended and generalized the method for two or more stages and called it Multistage RSS. Al-Omari and Jaber (2008) combine it with a modified method Percentile RSS. Muttlak (2003) investigated another modification of RSS, namely, Quartile RSS. Similar to his median RSS method, he suggested using the lower and upper quartiles for estimation of population mean instead of using all ordered units of sets. Similar to this method, Al-Omari and Jaber (2008) proposed Percentile RSS and combine it with Double RSS, as mentioned above.

In this study, we consider to compare the methods of Double RSS, Median RSS, Quartile RSS and Extreme RSS. They are all compared with RSS and SRS but are not compared with each other in the literature. In this way we can see more-appropriate methods for different cases such as; i) different set sizes ii) different kind of symmetric and asymmetric distributions iii) different sample sizes. Thus it would be a remarkable study for RSS researchers.

#### A COMPARATIVE SIMULATION STUDY FOR MODIFIED RANK SET SAMPLING METHODS

The ranked set sampling (RSS) is efficient and advantageous method where sets are randomly selected from a population, units in the sets are ranked by “sense of proportion” and this easier and cheaper information is used for inference. In this study, we consider RSS method and some modified versions of it for estimating the population mean. A simulation study is conducted for comparison the methods of Double RSS, Median RSS, Quartile RSS and Extreme RSS. The results give useful information about the features of the methods.

**Key Words:** sampling, ranked set sampling, modified ranked set sampling methods

#### REFERENCES

- [1] McIntyre G.A., A method of unbiased selective sampling using ranked sets, *Aust. J. Agric. Res.* 3, pp. 385–390 1952.
- [2] Takahasi K. and Wakimoto K. On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, 20:1–31, 1968.
- [3] Chen Z., Bai Z., Sinha B. K., *Ranked Set Sampling: Theory and Application*, Springer, 2003
- [4] Dell T.R., Clutter J. L. Ranked set sampling theory with order statistics background. *Biometrika*, 28:545–555, 1972.
- [5] Al-Nasser A. D., Mustafa A. B., Robust extreme ranked set sampling, *Journal of Statistical Computation and Simulation*, 79:7, 859-867, 2009

#### AN ALGORITHM FOR THE DETERMINATION OF ECONOMIC DESIGN OF X-BAR CONTROL CHARTS

Aysun Sagbas<sup>1</sup>, Adnan MAZMANOGLU<sup>2</sup>

<sup>1</sup> Department of Industrial Engineering, University of Namik Kemal, Corlu, Tekirdag/ Turkey  
asagbas@nku.edu.tr

<sup>2</sup> Department of Statistics, University of Istanbul Aydin, Florya Campus, Istanbul/Turkey  
adnanmazmanoglu@aydin.edu.tr

#### 1. Introduction

Control charts may also be used to estimate the parameters of a production process and to determine process capability. They are on-line process control technique widely used for this purpose. The control charts may also provide information useful in improving the process. A survey by Saniga [1] indicated that the averages control chart, X-bar chart, dominates the use of any other control chart techniques if quality is measured on a continuous scale. Various models for choosing control chart design parameters according to economic considerations, specifically expected total cost minimization, can be found in the quality control literature. Duncan [2] proposed the first model for determining the sample size, the interval between successive samples, and the control limits of an X-bar control chart. Since then, considerable attention has been devoted to the optimal economic determination of these three parameters [3-6]. Montgomery [7] gave a thorough review of the literature for the economic design and applications of various control charts. Ho and Case [8] also provided a literature review of the related topics covering the period 1981–1991. Lorenzen and Vance [9] presented a unifying perspective on the economic design of different types of control charts, including a discussion of the implicit underlying assumptions and the relative lack of sensitivity of the cost performance of these assumptions. Chung [10] proposed a simplified procedure for determining the values of  $n$ ,  $h$ , and  $k$  based on Duncan's model. This procedure consists only in solving an explicit equation for  $h$  in terms of  $n$  and  $k$ . Woodall [11] criticized the economic approach for, among other reasons, its possibility of providing a plan that generates an excessive number of false alarms. In this paper, an approach proposed by Lorenzen and Vance [9] and Duncan [2] is described and a simple procedure is developed to solve the problem of selecting the appropriate sample size, the width of the control limits, and the interval between samples. It considers appropriate revenues and costs, and provides economically optimum values of  $n$ ,  $h$ , and  $k$  upon the shift, time and cost parameters involved. An example from the petrol industry is presented, and the results are discussed.

#### 2. Methodology

In this study, the first economic model developed by Duncan [2] and a general model proposed by Lorenzen and Vance [9] was performed to minimize the expected total costs per time unit as a function of the design parameters. A simple procedure is described to determine the design parameters of an X-bar control chart. Fortran programming is used to compute optimum sampling size, sampling period, and control limit width. An example selected in the petrol industry is presented to illustrate the solution procedure for the economical design of the X-bar chart. For a certain combination of sampling size, sampling period, and control limit, the program also calculates the corresponding  $\alpha$  risk and power  $(1-\beta)$ . After the solution, the results of the economic models performed are compared.

#### References

- [1] Yu FJ, Low C (2005). An algorithm for the determination of optimal design parameters of X-bar control charts. *International J. of Advance Manufacturing Technology*, 26, 86–89.
- [2] Ho C, Case KE (1994). Economic design of control charts: a literature review for 1981–1991, *Journal of Quality Technology*, 26, 1–78.
- [3] Lorenzen TJ, Vance LC (1986). The economic design of control charts: A Unified approach. *Technometrics*, 28(1), 3-10.
- [4] Chung KJ (1990). A simplified procedure for the economic design of control charts. *International Journal of Production Research*, 28, 1239-1246.
- [5] Woodall WH (1986). Weaknesses of the economic design of control charts. *Technometrics*, 28, 408-410.

#### ABSTRACT

Control charts are the primary quality improvement tools used to establish and maintain statistical control of manufacturing processes. The effective use of control charts is largely dependent upon their design, that is, selection of the decision variables such as sample size (n), sampling period (h), and control limit (k) based on some subjective and/or objective criteria. Traditionally, control charts have been designed with respect to statistical criteria only. The design of control chart has economic consequence in that the costs of sampling and testing, costs associated with investigating out-of control signals and possibly correcting assignable causes, and costs of allowing non-conforming units to reach the consumer. For this reason an economic design is necessary before the control chart is used. Economic control charts should be designed in order to achieve minimum quality control costs, not consider the statistical properties. In this paper, an evolutionary algorithm based on Fortran programming is presented to determine to optimal design parameters of an X- bar control charts in Lorenzen and Vance economic model which applies to all control charts and Duncan model, regardless of the statistic used, such that they minimize the expected total costs per time unit. A numerical example and industrial case study are provided to illustrate the solution procedure.

**Key Words:** Statistical process control, X-bar control chart, economic design

## BİLDİRİ OTURUMLARI 5

### SESSION 5

#### *Kalkınma Bakanlığı Özel Oturumu 2*

## DOĞU MARMARA 2014-2023 BÖLGE PLANI KAPSAMINDA HAZIRLANAN MİKRO BÖLGELEME ÇALIŞMASI

Fatma AVŞAR\*; Hale Mahzeminli GÜLBAZ; Candan Umut ÖZDEN; Cem BAYRAK; Hülya YORULMAZ  
Adres: Doğu Marmara Kalkınma Ajansı Yenişehir Mahallesi Demokrasi Bulvarı No: 72/A İzmit/ Kocaeli  
E-mail(sırasıyla); fatmaavsar@marka.org.tr; hale.m@marka.org.tr; c.umutozden@marka.org.tr;  
cembayrak@marka.org.tr; hulyayorulmaz@marka.org.tr

### 1. Doğu Marmara 2014-2023 Bölge Planı'nın ve Hazırlayan Kurumun Tanımlanması

Ülkemizde İmar Kanunu'nun 8. maddesi gereği Kalkınma Bakanlığı (mülga Devlet Planlama Teşkilatı) bölgesel planlamadan sorumlu kurum olarak belirlenmiştir. Kalkınma Bakanlığı ise günümüzde plan onama hakkını saklı tutmak üzere; kuruluş süreçlerini tamamlamış olan kalkınma ajanslarının kendi bölgeleri için bölge planı hazırlamalarını uygun bulmuştur. Bu çerçevede, 5449 sayılı kanun uyarınca ve 2009/15236 sayılı Bakanlar Kurulu Kararıyla kurulan Doğu Marmara Kalkınma Ajansı (MARKA) Kocaeli, Sakarya, Düzce, Bolu ve Yalova illerini kapsayan TR42 düzey-2 bölgesini kapsayan bölge planının hazırlanmasından mesul tutulmuştur. Bu göreve dayanarak, 10 yıllık bir bakış açısıyla Doğu Marmara 2014-2023 Bölge Planı MARKA tarafından hazırlanmıştır.

Doğu Marmara 2014-2023 Bölge Planı'nın amacı; bölgenin sosyo-ekonomik kalkınmasını mekânsal olarak tanımlayabilmek ve kalkınma sürecini stratejik olarak programlamaktır. Kalkınma Planının yanı sıra, ulusal ölçekte kabul edilen diğer strateji ve politikaları Doğu Marmara Bölgesi'ne uyarlamak üzere hazırlanmış olan Bölge Planı; planlama hiyerarşisinde yer alan planların çerçevesini tanımlama özelliğini de taşımaktadır. Bu özelliği ile Bölge Planı; sosyo-ekonomik kalkınma ilkelerini belirliyor olması dolayısı ile bölgedeki sosyo-ekonomik kararları, mekânsal gelişmeyi tanımlıyor olması dolayısı ile de bölgedeki mekân odaklı kararları bağlayıcıdır.

### 2. Bölge Planı Kapsamında Hazırlanan Mikro Bölgeleme Çalışmasının Tanımlanması

5 ili ve 51 ilçeyi kapsayan Doğu Marmara Bölgesi gelişmişlik açısından ülke ortalamasının oldukça üzerinde yer alsa da, gelişmişliğin bölge geneline homojen yayılımı söz konusu değildir ve pek çok gelişmişlik göstergesi alt bölge birimlerinde çeşitlilik göstermektedir. Heterojen yapının söz konusu olduğu bölgede, bütüncül planlama ilkelerinin belirlenmesi ve safi bölge genelini kapsayan programların tanımlanması, kalkınma çabasının beklenen hızda ve verimde gerçekleşmesini engellemektedir. Bu noktadan hareketle, Doğu Marmara Bölgesi'nde önümüzdeki 10 yılın gelişme ilkeleri tanımlanırken, bazı stratejiler tüm düzey-2 bölgesini kapsamaktayken, bazı stratejiler için bölge altı birimlerin odak seçilmesi ihtiyacı ortaya çıkmıştır. Bu çerçevede, bölge altı birimlerin belirlenmesi adına Bölge Planı hazırlıkları kapsamında mikro bölgeleme çalışması yapılmış ve plan kurgusu mikro bölgeleme çalışması ile belirlenen alt bölge birimleri üzerine şekillendirilmiştir. Bu özelliği ile mikro bölgeleme çalışması, bölge planının uygulanabilirliğine temel teşkil etmektedir.

### 3. Mikro Bölgeleme Çalışmasında Başvurulan İstatistikî Yöntemler

Mikro bölgeleme çalışmasında bölge ilçe veritabanı oluşturulması; verilerin işlenmesi; veri analizi ve temel bileşenler analizi, çekim (gravity) modeli, ilçe yöneticileri ile yarı yapılandırılmış mülakat ve kümeleme çeşitli istatistiksel analizler gerçekleştirilmiştir. Saha çalışması, analiz çalışmaları ve uzman görüşleri birleştirilerek belirlenen mikro bölgeleme çalışmasının sonucunda Doğu Marmara Bölgesi'nde 6 gelişmişlik seviyesi, birbiriyle çalışan 11 ilçe kümesi ve benzer gelişim karakteri sergileyen 3 alt bölge birimi oluşturulmuştur.

İlçe Veritabanı Oluşturulması-Verilerin İşlenmesi: Paydaşlardan ikincil veri temini kapsamında 67 kurum ile gerekli veriler için talep yazıları hazırlanmış ve kurum ilgilileri ile istişare edilerek ilçe veritabanı oluşturulmuştur. Eksik ve tutarsız veriler kurum ilgilileri ile görüşülerek düzeltilmiştir.

Analiz Çalışmaları: Oluşturulan ilçe veri tabanı ile ilçelerin sosyal ve ekonomik özellikleri arasında bir ilişkilendirmeye olanak sağlayacak 21 değişken ve 4 alt ana başlık altında (demografi, sağlık, eğitim ve mali-ekonomik) sosyo-ekonomik gelişmişlik endeksi (SEGE) oluşturulmuş ve endeks genel sıralamasına göre bölge 6 gelişmişlik sınıfına ayrılmıştır. Temel bileşenler analizi, çekim (gravity) model, değişen pay analizi (shift share), yığınlaşma analizi, üç yıldız analizi, coğrafi bilgi sistemleri temelli analizler (tematik haritalama, ulaşım süresi üzerinden ağ analizi, doğal eşik analizi, sayısal yükselti modellenmesi, hizmet alanı analizi ve erişilebilirlik analizi, trafik yoğunluğu analizi başlıca kullanılan analiz teknikleridir.

Saha ve Anket Çalışmaları: Çalışmaları: Bu kapsamda 51 ilçenin yerel ve idari amirleri ile yapılandırılmış mülakat gerçekleştirilmiştir. Bu çalışmaların sonuçları 51 ilçe durum raporu olarak derlenmiş ve yayımlanmıştır. Bu çalışmalar çerçevesinde Kalkınma Kurulu Bölgesel Gelişme Ulusal Stratejisi ve Bölge Planı çalıştayında il bazlı sorun tespiti ve 5 ilde ise tespit edilen sorunlara yönelik proje ve uygulama önerileri üzerine ajans uzmanları moderatörlüğünde istişare toplantıları düzenlenmiştir. Ayrıca çalıştay katılımcılarına Bölgesel Gelişme Anketleri ve tüm bölge paydaşlarına bölge planı internet sitesi (<http://www.dogumarmarabolgeplani.gov.tr/>) üzerinden yenilikçilik, yaşanabilirlik ve rekabetçilik temalarında anketler düzenlenmiştir.

Sonuç itibarıyla analiz, saha çalışmaları ve uzman görüşleri harmanlanarak Doğu Marmara Bölgesi ilçeleri global, dinamik ve çevre alt bölgeler olmak üzere üç alt bölgeye ve İzmit, Gebze, Adapazarı, Güney Körfez, Doğu Merkez, Hendek, Karadeniz Kıyısı, Sakarya Güney, Bolu Güney, bolu Doğu ve Termal olmak üzere 11 ilçe kümesi olarak mikro bölgelere ayrılmış ve bu kapsamda 10 yıllık vizyon ile strateji ve politika önerileri geliştirilmiştir.

### MICRO-ZONATION STUDY PREPARED FOR EAST MARMARA 2014-2023 REGIONAL PLAN

In this paper, the statistical process micro-zonation applied in East Marmara 2014-2023 Region Plan that is prepared by East Marmara Development Agency is summarized. To do this, how sub-regions of East Marmara Region are obtained is explained. The main steps of micro-zonation are preparing datasets of 51 districts of the region; processing of data set; statistical analysis studies such as Economic and Social Development Index of East Marmara Districts, gravity models, shift share analysis, 3-star analysis of regional clusters, analyses based on Geographic Information Systems based; field studies, surveys and expert views. As a result of the all tools, 3 main sub-regions and 11 district clusters are determined in East Marmara Region.

**AJANS MALİ DESTEKLERİNDE SONUÇ ODAKLI İZLEME DEĞERLENDİRME MODELİ –  
ETKİLİLİK ANALİZİ ÖRNEĞİ**

Osman TOKSİN  
İpekyolu Kalkınma Ajansı

İncilipınar Mh. Muammer Aksoy Blv. Vakıflar Güven İş Mrk. Kat 2 Şehitkamil/GAZİANTEP  
osman.toksin@ika.org.tr

**Giriş**

Devlet teşviklerinin program hedefleri üzerindeki etkililiğinin araştırılması önemli bir konudur. Bölgesel Kalkınma Ajanslarının temel amaçlarından birisi de çeşitli mali destek araçları kullanarak bölgesinde sürdürülebilir kalkınmaya katkıda bulunmaktır. Kalkınma Ajansları sorumlu oldukları bölgede daha fazla ama aynı zamanda daha etkili mali destek mekanizmaları yürütmeyi hedeflemektedirler. Ajanslar, gerçekleşen mali desteklerin bölge ekonomisi üzerindeki sonuçlarının gereğince değerlendirilmesi ve analiz edilmesi ile daha etkili mali destek mekanizmaları geliştirebilecektir. Mali desteklerin etkililiği, Kalkınma Ajanslarının performanslarının ölçümü için de önemli bir indikatör olarak karşımıza çıkmaktadır. Etkililiğin ölçülmesi ancak etkin bir izleme değerlendirme sistemi ile mümkün olabilir. Sağlıklı bir bölgesel plan hazırlanması için mali desteklerin uygulama aşamasını müteakip İzleme Değerlendirme sürecinin bir sonraki planlama döngüsüne geri bildirimini yeterince sağlamalıdır. Geri bildirim sürecinde özellikle proje-program-plan arasındaki ilişki ortaya çıkarılmalıdır. Bu bakımdan devlet müdahalelerinin sonuçlarının gelecekteki planlama çalışmalarına ışık tutması çok kritik bir husus olarak karşımıza çıkmaktadır.

**Metodoloji**

Bu çalışmada uluslararası literatürde kabul edilen değerlendirme kriterleri (ilgili, etkinlik, etkililik, etki ve sürdürülebilirlik) çerçevesinde sonuç odaklı bir izleme değerlendirme modeli anlatılarak etkililik analizi üzerine bir uygulama gerçekleştirilmiştir. İpekyolu Kalkınma Ajansı'nda son iki yılda sonuçlanan mali destek programlarının KOBİ bileşeni kapsamında etkililik analizi yapılmıştır.

Ekonominin büyüklüğünü gösteren en önemli indikatörlerden bir tanesi imalat sanayi kapasitesidir. İmalat sanayii kapasite artışları bazen yeni yatırımlar bazen de yenileme yatırımları ile gerçekleşmektedir. Mevcut üretim hattı yenilenme yatırımları, genellikle yeni teknoloji ile donatılmış üretim hattının eskisinin yerini alması şeklinde ortaya çıkmaktadır. Yeni yatırımlar firmaların verimliliğine de önemli katkılar sağlamaktadır. Aynı zamanda üretkenliğin artmasını ve üretim artıklarının-enerji maliyetlerinin azalmasını sağlamaktadır. Hammaddede ve enerji maliyetlerindeki azalma ile ürün kalitesindeki artış ise firmaların sektördeki rekabet gücünü önemli ölçüde artırmaktadır. Rekabet gücü artan firmaların buna paralel olarak siparişlerinde, üretim miktarında ve satışlarında artış gerçekleşmektedir. Böylece firmada doğrudan ya da dolaylı olarak yeni yatırımla ilişkili istihdam edilen personel sayısı artmaktadır. Kar amacı güden işletmelerdeki gelir ve istihdam artışları bölgesel kalkınmanın en önemli göstergelerden ikisi olarak karşımıza çıkmaktadır. Bunun yanı sıra ölçek ekonomisinden kaynaklanan birim maliyetlerindeki azalma firmanın karlılığını artırmakta ve firmaya rakiplerine göre önemli derecede rekabet avantajı sağlamaktadır.

Bu kapsamda, Kapasite Artışı (%), İstihdam Artışı (Kişi), Satışlardaki Değişim (%), İhracat Artışı (%) ve Karlılık Oranındaki Değişim (%) başlıklarında 2010 ve 2011 Mali destek programlarında desteklenen 59 projenin, NACE Rev.2 sektör bölümlerine göre sınıflandırılmış 18 farklı sektörde program hedefleri doğrultusunda etkililiği araştırılmıştır. Çalışmada mali desteklerdeki dara kaybı, ikame etkisi ve çarpan etkisi dikkate alınarak Ajans desteklerinin net katkısallıkları da hesaplanmıştır. Çalışma kapsamında yapılan analizlerde, firmaları etkileyen mali destekler dışındaki diğer etkenlerin (artan/azalan üretkenlik, diğer yatırımlar, artan/azalan talep ve rekabet koşulları vb.) göstergeler üzerindeki etkisi sabit olarak kabul edilmiştir.

**Bulgular**

Araştırmada, proje ve programların sektörel performansını ortaya koyması bakımından kritik öneme sahip veriler elde edilmiştir. Kapasite oranı, istihdam, satış değişimi, ihracat ve karlılık başlıklarında Ajans desteklerinin etkililiği analiz edilmiştir. Ayrıca NACE Rev.2 'ye göre sınıflandırılmış 18 farklı sektörde Ajansın mali desteklerdeki brüt ve net katkısallıkları hesaplanmıştır. Dara kaybı ve çarpan etkisi hesaplamalara dahil edilerek Ajansın net katkısallığı hesaplanmıştır. Böylece gelecekte çıkılması muhtemel proje teklif çağrılarının bölge ekonomisi için olası etkileri konusunda fikir edinilebilmektedir. Sonuç odaklı izleme

değerlendirme modelinin mali desteklerin etkililiğinin ölçülmesi ve planlamaya geri bildirim sağlaması bakımından önemli bir uygulama aracı olduğu tespit edilmiştir.

**Çizelge1. Ajans Katkısallığı (Özet Bulgular)**

(Dara Kaybından Arındırılmış, Çarpan Etkisi İlave Edilmiş)		
	Brüt Katkısallık	Net Katkısallık
<b>Proje Sayısı</b>		<b>59</b>
<b>Ajans Desteği Sayesinde Yapılan Toplam Yatırım (TL)</b>	<b>23.054.309</b>	<b>18.575.566</b>
<b>Gerçekleşen Mali Destek Tutarı (TL)</b>	<b>10.046.640</b>	<b>8.111.537</b>
<b>Kapasite Artış Oranı (Ortalama %)</b>	<b>21</b>	<b>18</b>
<b>Net Satışlardaki Artış Tutarı (Toplam TL)</b>	<b>22.829.510</b>	<b>19.386.608</b>
<b>Doğrudan/Dolaylı Ek İstihdam Sayısı (Kişi Sayısı)</b>	<b>435</b>	<b>350</b>
<b>Her Ek İstihdam için Gerçekleşen Mali Destek Tutarı (TL)</b>		<b>23.161</b>

**KAYNAKLAR**

- [1] EC HANDBOOK For Results-Oriented Monitoring Of Ec External Assistance, 2008, EC EuropeAid Co-operation Office.
- [2] European Commission, Measuring Structural Funds Employment Effects, Working Document No. 6, March 2007
- [3] ERSAYIN Z. Mart 2012 Kamu Müdahalelerinde Katkısallığın Değerlendirilmesi: Çukurova E İzmir Kalkınma Ajansları Kobi Destekleri Örneği (Uzmanlık Tezi) BÖLGESEL GELİŞME VE YAPISAL UYUM GENEL MÜDÜRLÜĞÜ ISBN 978-975-19-5256-1
- [4] KUSEK, J. Z. and RIST. R. C. , 2004, Ten Steps to a Results-Based Monitoring and Evaluation System, THE WORLD BANK Washington, D.C..
- [5] KHAN, M. A. , 2001, "A Guidebook on Results Based Monitoring and Evaluation: Key Concepts, Issues and Applications." Monitoring and Progress Review Division, Ministry of Plan Implementation, Government of Sri Lanka. Colombo, Sri Lanka.

### TR41 BÖLGESİ İLÇE KALKINMA ENDEKSİ

Emine ARSLAN PAULI\*, Yasin DALGIÇ

Bursa Eskişehir Bilecik Kalkınma Ajansı (BEBKA), Yeni Yalova Yolu 4.Km. BUTTİM Plaza Kat .6 16250  
Osmangazi/ BURSA  
emine.arslan@bebka.org.tr, yasin.dalgic@bebka.org.tr

#### Giriş

Bölgesel kalkınma alanında, kalkınmanın nasıl ölçüleceği, bunun için kıyaslanabilir bir ölçünün nasıl oluşturulacağı, kalkınma gibi çok değişkenli bir kavramın nasıl sayısal olarak somutlaştırılabileceği başta gelen sorulardır. Türkiye’de kalkınma ajanslarının da bölge içi gelişmişlik farklılıklarının giderilmesinde görev almasıyla, bölge içi gelişmişlik farklılıkların somut olarak ortaya konma gereği daha da önem kazanmıştır. Bölge içi gelişmiş farklılıkların ölçülmesi ve bu farklılıklara yönelik plan, program ve politikalar üretilebilmesi amacıyla TR41 Bölgesi Bursa, Eskişehir ve Bilecik ilçeleri için kalkınmanın farklı yönlerinin ilçe düzeyinde ele alındığı ‘TR41 Bölgesi İlçe Kalkınma Endeksi’ oluşturulmuştur. Oluşturulan endeks, TR41 Bölgesi’nin Bursa iline bağlı 17, Eskişehir iline bağlı 14 ve Bilecik iline bağlı 8 ilçe olmak üzere toplam 39 ilçesi düzeyinde, 2010-2012 yılları dönemi için ulaşılabilen güncel veriler kullanılarak hazırlanmış 24 göstergeden oluşmaktadır. Endeks oluşturulması için çok değişkenli istatistik tekniklerinden Temel Bileşenler Analizi (TBA) kullanılmıştır. Hesaplanan endeks değerleri kullanılarak bölgenin 39 ilçesi kalkınma açısından sıralanmış, kullanılan göstergeler bazında da irdelenmiştir.

#### Metodoloji

İlçe düzeyinde erişilebilen istatistiklerden yola çıkarak ve kalkınmayı etkileyen nüfus, eğitim, sağlık, enerji ve çevre konularında hesaplanan 24 göstergeye uygulanan Temel Bileşenler Analizinde, Kaiser kuralı ile 1’in üzerinde özdeğere sahip olan faktörler dikkate alınmış, bu özdeğerlerin lineer bileşeni olarak ‘faktör skorları’ hesaplanmıştır. Çalışmada faktör skorları doğrudan ‘endeks değeri’ olarak alınmıştır. Analizin uygunluğunu belirleyen Kaiser-Meyer-Olkin Measure of Sampling Adequacy testi sonucu gösterge setinin yeterliliğine işaret etmekte, özdeğeri 1’in üzerinde olan altı faktör varyansı yaklaşık %84’ünü açıklamaktadır.

#### Çizelge1. TR41 İlçe Kalkınma Endeksi Gösterge Listesi

Gösterge	Yılı	Gösterge	Yılı
Nüfusun Toplam Bölge Nüfusu İçindeki Payı (%)	2012	Yüz Bin Kişi Başına Hekim Sayısı	2012
Nüfus Artış Hızı (binde)	2012	Yüz Bin Kişi Başına Hemşire Sayısı	2012
İl ve İlçe Merkezleri Nüfus Artış Hızı (%0)	2012	Okuma Yazma Bilen Oranı (%) (6 yaş ve üzeri)	2012
Belde ve Köyler Nüfus Artış Hızı (%0)	2012	Okuma Yazma Bilen Kadın Oranı (6 yaş ve üzeri)	2012
Şehirleşme Oranı (İl ve ilçe merkezleri nüfusunun toplam nüfus içindeki oranı) (%)	2012	Kişi Başına En Az Yüksekökol/Üniversite Mezunu Sayısı (15 yaş ve üzeri nüfus başına) (%)	2012
Genç Bağımlılık Oranı (0-14 yaş)	2012	İlköğretimde Derslik Başına Düşen Öğrenci Sayısı	2010
Yaşlı Bağımlılık Oranı (65+ yaş)	2012	Ortaöğretimde Derslik Başına Düşen Öğrenci Sayısı	2010
Banka Şubesi Sayısı	2012	İlköğretimde Öğretmen Başına Düşen Öğrenci Sayısı	2010
Kişi Başına Elektrik Tüketimi (kWh)	2011	Ortaöğretimde Öğretmen Başına Düşen Öğrenci Sayısı	2010
4a Sigortalı Sayısının 15+ Yaş Nüfusa Oranı (%)	2011	Düzenli Depolama İle Bertaraf Edilen Atık Miktarı Oranı (%)	2010

Bin Kişi Başına Hastane Yatak Sayısı	2012	Kişi Başına Belediye Tarafından Dağıtılan Su Miktarı (m3/yıl)	2010
Yüz Bin Kişi Başına Uzman Hekim Sayısı	2012	Kişi Başına Arıtma Tesislerinde Arttırılan Su Miktarı (m3/yıl)	2010

#### Bulgular

Endeks değerlerine göre TR41 Bölgesi, Bursa, Eskişehir ve Bilecik illerinin ilçeleri sıralanmış, endeks değerlerine göre sıralanıp altı gruba ayrılmıştır.



Şekil 1. TR41 Bölgesi İlçe Kalkınma Endeksi

Şekil 2. İlk 16 İlçenin Endeks Değeri

Temel bileşenler analiziyle oluşturulan yük matrisine göre, TR41 Bölgesi İlçe Kalkınma Endeksinde en yüksek yüke pozitif yönde Şehirleşme Oranı ve Kişi Başına Uzman Hekim Sayısı sahiptir. Bu göstergelerin endekste pozitif yükünün olması, gelişmişlikle bu gösterge arasındaki doğru orantıyı ortaya koymaktadır. Diğer yandan, büyüklük açısından üçüncü yüksek yüke sahip gösterge Yaşlı Bağımlılık Oranı için yük matrisinden ortaya çıkan negatif ilişki, yaşlı bağımlılık oranı arttıkça gelişmişliğin olumsuz etkilendiğini ortaya koymaktadır..

#### KAYNAKLAR

- [1] Kaiser, H.F. (1960), ‘The application of electronic computers to factor analysis.’ Educational and Psychological Measurement 20, 141-151.
- [2] Dokuzuncu Kalkınma Planı 2007-2013, Sosyal Güvenlik Özel İhtisas Komisyonu, Yayın No: 2729 ÖİK:681, Ankara 2007.

#### ABSTRACT

#### DISTRICT DEVELOPMENT INDEX OF TR41 REGION

In the area of development, how to measure development and objectify the concept of development analytically are the vital in regional development. By the development agencies, of which mission is decreasing measure the discrepancies within the region, it became more important to portray the discrepancies within the region. For TR41 Bursa Eskişehir Bilecik Region (one of the 26 NUTS2 level regions).For the districts in the region, which is 39 districts in total, based on 24 indicators and taking the 2010-2012 years as reference period, District Development Index of TR41 Region has been developed. For development of the index the index, the method of Principal Component Analysis (PCA) used and the districts were ranked and examined on the basis of the indicators used.

**Key Words:** regional development index, TR41 Bursa Eskişehir Bilecik, principal component analysis

**TÜRKİYE'DE KALKINMA AJANSLARI KAPSAMINDAKİ İLLERİN BAZI SOSYO EKONOMİK  
GÖSTERGELER BAKIMINDAN ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZLERLE  
DEĞERLENDİRİLMESİ**

Hasan BULUT\*, Yüksel ÖNER, Taner TUNÇ

Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 55139, Samsun, TÜRKİYE

hasan.bulut@omu.edu.tr

**Özet**

Bölgeler arası gelişmişlik farklarının azaltılmasına yönelik bölgelerin sosyoekonomik analizlerinin yapılması ve Avrupa Birliği (AB) ile karşılaştırılabilir veriler üretilmesi amacıyla, ülkemiz AB bölgesel sınıflandırması olan NUTS kriterlerine göre 2002 yılında alınan Bakanlar Kurulu kararıyla ekonomik, sosyal, kültürel ve coğrafi yönlerden benzer illerin nüfus büyüklükleri de dikkate alınarak 26 istatistikî bölge birimine ayrılmıştır.

2006 yılında AB'ye uyum yasaları çerçevesinde Devlet Planlama Teşkilatı (DPT) bünyesinde ülkemizde ki bölgesel kalkınmayı hızlandırmak için 26 istatistikî bölgenin her birinde bir tane olmak üzere kalkınma ajansları kurulmuştur.

Bu çalışmanın asıl amacı, Türkiye'de illerin sosyoekonomik gelişmişlik düzeyini belirlemek için dönem dönem yapılan çalışmaları bu kalkınma ajanslarına göre güncellemektir. İllerin bireysel olarak incelenmesi yerine kalkınma ajanslarının tek bir idari bölge olarak düşünülüp ele alınmasının bölgeler arası gelişmişlik seviyelerini daha iyi açıklayacağı düşünülmektedir. Böylece aynı bölgedeki bir ilin gelişmişlik seviyesi artsa bile diğer illerde böyle bir gelişme söz konusu değilse, bölgenin gelişmekte olduğu ve kalkınma ajansının doğru politikalar izlediği yönündeki iddiaların doğruluğu tartışılabilir olacaktır.

Bu amaçla kalkınma ajansları kapsamında yer alan illere ait bazı sosyoekonomik göstergelerden yararlanarak her bir kalkınma ajansı bölgesi için söz konusu göstergelere ait toplam değerler bulunduktan sonra, elde edilen çok değişkenli veri yapısı Faktör analizi ve Çok değişkenli varyans analizi (MANOVA) gibi Çok Değişkenli İstatistiksel analizler kullanılarak kalkınma ajansı bölgeleri değerlendirilecektir.

**KAYNAKLAR**

- [1] Rencher A. C. (2002), Methods of Multivariate Analysis, New York, United States of America.
- [2] Erilli N. A., Tunç T., Öner Y. ve Yolcu U. (2008), İllerin Sosyoekonomik Verilere Dayanarak Bulanık Kümeleme Analizi İle Sınıflandırılması, e-Journal of New World Sciences Academy.
- [3] Yıldız E. B., Sivri U. ve Berber M. (2012), Türkiye'de İllerin Sosyoekonomik Gelişmişlik Sıralaması, Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Kayseri.
- [4] Albayrak A. S. (2005), Türkiye'de İllerin Sosyoekonomik Gelişmişlik Düzeylerinin Çok Değişkenli İstatistik Yöntemlerle İncelenmesi, ZKÜ Sosyal Bilimler Dergisi, Zonguldak.

**ABSTRACT**

**INTERPRETING SOME CITIES, IN THE SCOPE OF DEVELOPMENT AGENCIES, OF TURKEY  
IN THE MANNER OF SOCIAL-ECONOMIC INDICATORS BY SOME MULTIVARIATE  
STATISTICAL ANALYZES**

In the year 2006 development agencies were established in each 26 statistical regions of Turkey within Devlet Planlama Teşkilatı (DPT) aiming to promote rapid regional development in the manner of compliance laws to European Union (EU).

The actual aim of this paper is to update the periodic studies on defining social-economic development levels of cities in Turkey according to these established development agencies. It is believed that considering the development agencies as a one administrative authority would define levels of developments of regions better than considering the cities one by one as an individual. For doing this total values of development agencies of considered regions are found in the manner of their leading social-economical indicators and then development agencies regions will be interpreted by using Multivariate Statistical Analysis Methods such as Factor Analysis and MANOVA.

**Key Words:** Development of Social-Economic, Factor Analysis, MANOVA



**ÇOK DEĞİŞKENLİ İSTATİSTİKSEL YÖNTEMLERİN TRC2 BÖLGESİ İLÇELERİ ÜZERİNE  
UYGULAMALARI**

Erhan DEMİRCAN

2014-2023 TRC2 Bölgesi Bölge Planı çalışmaları kapsamında TRC2 Bölgesi ilçeleri çok değişkenli istatistiksel yöntemler kullanılarak sıralanmış ve sınıflandırılmıştır. Çalışmanın ana amacı, TRC2 Bölgesi ilçelerinin sosyo-ekonomik gelişmişlik düzeylerini çok değişkenli bir yaklaşımla incelemektir. Araştırmanın ilk bölümünde, 25 ilçeye ait 31 adet sosyal ve ekonomik değişken kullanılarak temel bileşenler analizi ile gelişmişlik endeksleri elde edilmiştir. Çalışmanın ikinci bölümünde, TRC2 Bölgesi ilçeleri çok değişkenli istatistiksel yöntemlerden kümeleme ve diskriminant analizi kullanılarak sınıflandırılmıştır.

**Anahtar Kelimeler:** Sosyo-ekonomik Gelişme, Temel Bileşenler Analizi, Kümeleme Analizi, Diskriminant Analizi,

**POSTER BİLDİRİLER**

## SAĞLIK YÖNETİMİ ENFORMASYON SİSTEMLERİNİN ERKEN TEŞHİS VE TANI İÇİN STOKASTİK MODELLER İLE GELİŞTİRİLMESİ

Adem DOĞANER, Sinan ÇALIK  
Fırat Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 23119, Elazığ/TÜRKİYE  
adoganer@firat.edu.tr scalik@firat.edu.tr

### Sağlık Yönetimi Enformasyon Sistemleri

Bilişim sistemleri, gelişen bilgisayar teknolojileri vasıtasıyla hizmet sektörlerinde büyük bir ihtiyacı karşılayarak ve birçok işlem konusunda zaman kazanımı ve pratiklik sağlamaktadır. Bilişim sistemleri, çok karmaşık bir yapıya sahip olan sağlık sistemlerinde de önemli çözümler ve pratik uygulamalar kazandırmaktadır. Sağlık bilişim sistemleri temelde uygulama farklılıkları bakımından klinik bilgi sistemleri ve teşhis tedavi sistemleri olarak iki grupta incelenebilmektedir. Sağlık bilişim sistemleri, sağlık alanında elde edilen veriler üzerinde çalışarak bu alanların geliştirilmesi için pratik uygulamalar üzerine bilgisayar teknolojilerinden yararlanarak çözümler üretmektedir. Klinik bilgi sistemleri, hastane yönetimi, hasta kayıt ve takibi, tıbbi görüntüleme ve laboratuvar analiz sonuçları kayıtları, tedaviye yönelik karar destek gibi alanlarda bilgisayar destekli sistemler geliştiren bilişim sistemleri olarak ifade edilebilmektedir. Teşhis ve tedavi sistemleri, tedavi ve teşhise destek sağlayan bilgisayar destekli sistemlerden oluşmaktadır. Sağlık bilgi sistemleri, sağlık alanındaki yapılan hareketlerin ortak bir havuzda birleşerek belirli bir veritabanı oluşturulmasına olanak sağlamıştır. Veritabanından elde edilen bilgiler doğrultusunda gerek hastane yönetimi aşamasında, gerekse de tıp çalışanlarının hasta teşhis tedavi konusunda destek aldığı önemli bir sistem haline almıştır.

Çalışmada sağlık enformasyon sistemlerinin geliştirilerek teşhis ve tanı ve ileriye yönelik meydana gelebilecek hastalıkların tahminlerinde başarılı sonuçlar sağlanması amaçlanmıştır. İleriye yönelik tahminlemelerde istatistiksel yöntemler sıklıkla uygulanmaktadır. Stokastik modeller, belirli olasılıklar ve koşullar altında olasılık teoremlerine dayanarak tahminler için kullanılabilirlerdir.

Mevcut enformasyon sistemlerinin pek çoğu belirli algoritmalar, veritabanları ve varlık ilişki diyagramları ile kayıtları tutma ve düzenli bilgiler oluşturma işlemini üstlenmektedir. Çalışmada enformasyon sistemlerinin stokastik modeller ile ilişkilendirilerek ileriye yönelik tahminler üretebilme yeteneğinin kazandırılması hedeflenmiştir. Bilgi sistemleri ile ilişkilendirilecek yöntem saklı Markov modelleridir.

### 2. Saklı Markov Modeli

Saklı Markov Modeli, Markov zincirindeki durumlar arası geçiş olasılıklarının yanı sıra durumlara bağlı olarak gelişen gözlem olasılıkları olmak üzere iki stokastik süreçten oluşmaktadır. Saklı Markov modelinde bilinen ve bilinmeyen parametreler mevcuttur. Durumlar arası geçişler Saklı Markov modelinde gizlidir. Fakat bu durumlara bağlı olarak gelişen gözlemler mevcuttur ve gözlem olasılıkları bilinmektedir. Gözlem olasılıkları aracılığı ile sistemin hangi durumda olduğu ve durumlar arası geçişler tahmin edilmek istenmektedir. Saklı Markov modelinde geçiş olasılıkları dağılımı  $A$ , her bir durum için gözlem sembolleri olasılıkları  $B$  ve başlangıç durum dağılımı  $\pi$  olmak üzere kesikli parametrelili saklı Markov Modeli;

$$\lambda = (A, B, \pi)$$

olarak ifade edilir.

Saklı Markov modeli sistemin elverişliliği bakımından olasılıksal incelemeler gerçekleştirilebilir. Bu incelemeler, saklı Markov modelinin ilgilendiği üç temel problem üzerinedir ve her bir problem çözümü için geliştirilmiş algoritmalar mevcuttur.

Saklı Markov modeli, bilgisayar bilimleri başta olmak üzere birçok alanda yaygın bir uygulama sahası bulmuştur.

Çalışmada, farklı yaş ve gruptan gelen hastaların hastalık belirtileri gözlemler olarak bilgi sistemine dahil edilmiştir. Belirtiler sonucu oluşan hastalıklar ise durumlar olarak nitelendirilmiştir. Belirtilenden yola çıkarak hastanın hangi hastalığa yakalanacağı olasılıksal olarak tahmin edilmiştir. Düzenlenen saklı Markov modeli, enformasyon sistemi ile ilişkilendirilerek hastaların geçmişte geçirdikleri hastalıklar ve tıbbi değerler dikkate alınarak gelecekte hangi hastalıklarla karşı karşıya kalabilecekleri tahmin edilmiştir.

### KAYNAKLAR

- [1] Rabiner, L., (1989). *A tutorial on hidden Markov models and selected applications in speech recognition*. Proceeding of IEEE. **77(2)**:257-286.
- [2] Baum, L.E., (1972). *An Inequality and Associated Maximization Technique in Statistical Estimation of Probabilistic Functions of a Markov Process*. Inequalities **3**: 1-8.
- [3] Dynkin, E.B., (1961), *Theory of Markov processes*. Pergamon Press, London.
- [4] Ömürbek, N. and Altın, F.G., (2009). *Sağlık Bilişim Sistemlerinin Uygulanmasına İlişkin bir Araştırma*, Sıparta, SDÜ Fen Edebiyat Fakültesi Sosyal Bilimler Dergisi, 19:211-232.
- [5] Güleş, H.K. and Özata M., (2005). *Sağlık Bilişim Sistemleri*, 1. Basım, Ankara, Nobel Yayın.

### ABSTRACT

#### HEALTH MANAGEMENT INFORMATION SYSTEMS DEVELOPMENT WITH STOCHASTIC MODELS FOR EARLY DETECTION AND DIAGNOSIS

*In this study, Health information systems for early detection and diagnosis were developed using Hidden Markov models. Symptoms and diseases were defined as observations and states for hidden Markov model, respectively.*

**Key Words:** Information Systems, Symptoms, Diseases, Hidden Markov Model.

## AÇIMLAYICI FAKTÖR ANALİZİ İLE KAHVEHANELERDE HİZMET KALİTESİ BOYUTLARININ BELİRLENMESİ

Algin OKURSOY\*

Alper Turan DEVLİ

Muhsin ÖZDEMİR

ADÜ Söke İşletme Fakültesi  
Yönetim Bilişim Sistemleri Bölümü  
09200 Aydın / Türkiye  
aokursoy@adu.edu.tr

ADÜ Sultanhisar MYO  
09650 Aydın / Türkiye  
alperdevli@gmail.com

ADÜ Nazilli İİBF İşletme  
Bölümü  
09800 Aydın / Türkiye  
ozdemmm@gmail.com

### 1. Giriş

Alan yazında hizmet kalitesinin ölçülmesi için yapılan çalışmaların amacı hizmet kalitesi ve müşteri memnuniyeti arasındaki ilişkiyi bir standart bir model çerçevesinde açıklamaktır. Müşteri memnuniyeti, en genel anlamı ile bir müşterinin bir ürünü veya hizmeti kullandıktan sonra deneyimin değerlendirilmesi sonucunda geliştirdiği olumlu tutum olarak tanımlanabilir. Müşteri memnuniyeti temel olarak, işletmenin performansı sayesinde müşteri beklentilerinin karşılanması veya aşılmasıdır (Çabuk ve Yağcı, 2007:164). Müşteri memnuniyeti, müşterinin ürün veya hizmetin arzu, beklenti ve ihtiyaçları karşılama yetisinden kaynaklanan genel memnuniyet derecesi, yapılan ürün veya hizmet tercihinin rakiplerine göre yarattığı iç huzur ve tüketicideki rahatlık hissidir (Hellier vd., 2003:1765). Bu durumda, kaliteli mal ve hizmetler müşteri memnuniyetini sağlamak için işletmelerin ortaya koyması gerekli çıktılardır.

Bu çalışmanın amacı kahvehanelerde hizmet kalitesini belirleyen boyutların belirlenmesi amaçlanmıştır. Öncelikle alan yazında daha önce yapılmış çalışmalar yardımıyla bu boyutlar açıklanmaya çalışılacaktır. Belirlenene boyutları ölçülebilmesi için alan yazında önerilen algılanan hizmet kalitesinin ölçülmesine yönelik bir ölçek seçilerek, kahvehaneden hizmet alan kişilere uygulanacaktır. Daha önce kahvehanelerde hizmet kalitesi üzerine herhangi bir çalışma yapılmadığından dolayı sonuçlar alan yazına katkı sağlayacaktır.

### Hizmet Kalitesi

Sasser, Olsen ve Wyckaff (1978) tarafından önerilen hizmet kalitesi modeline göre, personel, tesis ve materyal olmak üzere hizmet kalitesinin üç boyutu olduğu ileri öne sürülmektedir (Parasuraman vd, 1985: 42-43). Grönross (1984) tarafından ortaya atılan modele göre, hizmet kalitesinin teknik, fonksiyonel ve imaj olmak üzere üç boyutta değerlendirilmesini önermektedir (Kang ve James, 2004, 268).

Parasuraman, Zeithaml ve Berry (PZB) tarafından 1985 yılında geliştirilen ve fark analizi adı verilen modelde ise, hizmeti veren işletmelerin anlayışı ve uygulamaları ile hizmeti kullanan müşterilerin hizmetten beledikleri ve hizmeti tükettikten sonra ortaya çıkan gerçek hizmet arasındaki farklar ve bu farkların kaynağı yer almaktadır (Saat, 1999, 108-109). Bu modeli temel alarak PZB SERVQUAL ölçeğini geliştirmişlerdir. Buna göre, müşteri, beklentileri fazlasıyla karşılandıktan sonra kaliteyi pozitif olarak algılamaktadır. SERVQUAL ölçeğinde temel olarak beş boyut bulunmaktadır. Bunlar, Fiziksel görünüm, güvenilirlik, yanıt verebilirlik, güvence ve empati olarak isimlendirilmiştir (Bülbül ve Demirel, 2008, 182).

SERVQUAL ölçeği, restoran ve toplu yemek hizmeti veren işletmelerde hizmetin fiziksel yönlerini değerlendirmede yetersiz kalabilmektedir (Johns, 1996). Benzer şekilde SERVQUAL modeli, müşterinin gıda kalitesi üzerindeki algılarını ölçmede de yetersiz kalabilmektedir (Oberoi ve Hales, 1990). Bunlara ek olarak SERVQUAL ölçeğinde bulunan beş faktör, restoranlar ve toplu yemek hizmeti veren işletmelerde olduğu gibi fiziksel ürünlerin öne çıktığı işletmelerde yeniden elde edilmeyebilmektedir (Kivela vd., 1999: 211). SERVQUAL ölçeğindeki beş boyut evrensel değildir (Buttle, 1996). Ayrıca, SERVQUAL ölçeği beklenti ve algılar arasındaki farkı ölçerken, Teas (1993), müşteri memnuniyetinin hizmeti aldıktan sonra oluşan bir durum olduğunu belirtmiştir.

Hizmet kalitesinin ölçümüne ilişkin bir diğer model ise Cronin ve Taylor (1992) tarafından önerilen SERPERF'dir. Bu çalışmada, SERVQUAL modelinin hizmet kalitesini ölçmede yeterli olmadığını savunulmuş ve alternatif bir model olarak SERPERF ölçeği önerilmiştir (Bülbül ve Demirel, 2008, 183). Her iki ölçeğinde aynı maddeleri kullanmalarına rağmen, SERPERF modelinin SERVQUAL modeli gibi beş boyuttan değil; sadece tek boyutta hizmet kalitesini ölçebildiği savunulmuştur (Cronin ve Taylor, 1992: 61). Ayrıca bu model, kalite değerlendirmesi için beklenti ve algı arasındaki fark yerine, müşteri tarafından algılanan hizmet kalitesinin (hizmetin performansının) değerlendirilmesinin yeterli olacağı belirtilmektedir.

### KAYNAKLAR

- [1] PARASURAMAN, A., ZEITHAML, V.A. AND BERRY, L.L. (1985). A conceptual model of service quality and implications for future research. *Journal of Marketing*. Vol. 49, Fall, 41-50.ss.
- [2] CRONIN, J.J. JR., TAYLOR, S.A. (1992). Measuring Services Quality: A Reexamination and Extension. *Journal of Marketing*. 56(3), 55-68.ss.
- [3] TEAS, R., K. (1993). Expectations, Performans Evaluation, and Consumers' Perception of Quality. *Journal of Marketing*. 57(4), 18-34.ss.
- [4] HELLIER, P.K., GEURSEN, G. M., CARR, R.A., RICKARD, J. A. (2003). Customer Repurchase Intention: A General Structural Equation Model. *European Journal of Marketin.*, 37 (11/12), 1762-800.ss.
- [5] KIVELA, J., INBAKARAN, R., REECE, J. (1999). Consumer research in the restaurant environment, Part 1: A conceptual model of dining satisfaction and return patronage. *International Journal of Contemporary Hospitality Management*. 11(5), 205-222.ss.

### ABSTRACT

#### DETERMINATION OF SERVICE QUALITY DIMENSIONS IN COFFEEHOUSE BY AN EXPLORATORY FACTOR ANALYSIS

**Key Words:** Exploratory factor analysis, Coffeehouse, Service Quality.

There are numerous factors effecting customers' satisfaction perceptions of service quality in coffeehouse. Among these factors, price, service quality, food quality and physical environment are the most significant factors influencing service quality perceptions. In this study, we administered a survey to better understand the possible effects of mentioned four factors on service quality perceptions of customers at coffeehouse. The underlying dimensions of service quality have been investigated by using Exploratory Factor Analysis (EFA) technique.

## STATISTICAL PROCESSING OF WIND SPEED DATA FOR ENERGY POTENTIAL IN CORLU-TEKIRDAG, TURKEY

Aysun SAGBAS<sup>1</sup> and Adnan MAZMANOGLU<sup>2\*</sup>

<sup>1</sup> University of Namik Kemal, Department of Industrial Engineering, Tekirdag/Turkey, asagbas@nku.edu.tr

<sup>2\*</sup> University of Istanbul Aydin, Department of Statistics, Istanbul/Turkey, adnanmazmanoglu@aydin.edu.tr

### INTRODUCTION

Wind energy has become very popular in the last few years due to the environmental concerns, the increase in the price of fossil fuels and the decrease of the cost of wind power both in the world and in Turkey. Wind power has a positive effect on the quality of the air that we breathe, and the combustion of fossil fuels also produces the gasses sulphur dioxide and nitrogen oxide, both serious sources of air pollution. In the last decade, many researchers have shown interest in the estimation of the wind energy potential of several sites over the world. Chang et al., [1] analyzed the wind characteristic and energy potential at Taiwan presenting the Weibull characteristics parameters for the wind of this region and also simulation of wind generation with an wind turbine of 1,5MW in 3 different heights. Zaharim et al., [2] studied Weibull and Lognormal distributions to fit the wind speed data. There have been studies in recent years assessing the wind energy potential in Turkey. The yearly mean wind speed and the corresponding wind power density values for cities and regions were studied by various researchers. Akpinar and Akpinar [3] discussed Weibull and Rayleigh distributions to describe the wind energy potential of Nurdagi/Gaziantep district on the basis of 5 years hourly time series wind speed data. They concluded that the Weibull distribution provides better fit to probability distributions as compared to Rayleigh model. The wind energy potential in the eastern Mediterranean region using hourly wind data taken from seven stations during 1992–2001 periods investigated by Sahin et al., [4]. In the present study, wind energy potential in Corlu-Tekirdag is investigated using statistical methods to analyze the time series of monthly average wind speed in the period between 2010 and 2012 measured by using the wind data recorded at Tekirdag meteorological station for a period of 36 months.

### STATISTICAL ANALYSIS

Accurate estimation of long term wind speed probability distribution is a fundamental and challenging task in wind energy planning. Several continuous mathematical functions called

probability density functions can be used to model the wind speed frequency curve by fitting long time series measured data. Extensive literature search indicates that various parametric distribution models have been presented to estimate wind speed probability distributions, which are used in a variety of applications such as wind farm planning, long-term strategy of wind generators, and reliability evaluation of wind resources. Therefore, in the present study, these two PDFs have been preferred for the evaluation of measuremental data to determine the wind potential and available wind energy. The Weibull probability distribution function that is a special case of generalized gamma distribution for wind speed is expressed with Equation (1).

$$f(v) = \frac{k}{c} \left(\frac{v}{c}\right)^{k-1} \exp\left[-\left(\frac{v}{c}\right)^k\right] \quad (1)$$

where  $v$  is the wind speed,  $c$  is a Weibull scale parameter,  $k$  is a dimensionless Weibull shape parameter,  $f(v)$  is the probability of observing wind speed.

One of the important steps in the evaluation of different functions is the interpretation of different statistical parameters, namely, slope ( $s$ ), coefficient of determination ( $R^2$ ), mean bias error ( $MBE$ ), and root mean square error ( $RMSE$ ), as were used in the present article. Not any single statistical parameter can adequately be an indication of the goodness of a model. It is difficult to independently interpret the individual parameters; therefore, in the present article, as a more robust indication as the goodness of different functions, the accuracy

score proposed by Celik [5] is applied to the statistical parameters. A composite evaluation index, including some statistical metrics is defined in Equation (2).

$$\psi = 3 \left[ 1 - |1 - s| \right] + R^2 + \left[ 1 - \frac{RMSE}{RMSE_{max}} \right] + \left[ 1 - \frac{|MBE|}{|MBE_{max}|} \right] \quad (2)$$

### REFERENCES

- [1] Akpinar E.K. and Akpinar S., 2004. Determination of the wind energy potential for Maden Elazig, Turkey. *Energy Conversion and Management* 45, 2901–14.
- [2] Celik A.N., 2003. A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey. *Renewable Energy* 29, 593–604.
- [3] Chang T.J., Wu Y.T., Hsu H.Y., Chu C.R. and Liao C.M., 2003. Assessment of wind characteristics and wind turbine characteristics in Taiwan. *Renew. Energy* 28, 851–871.
- [4] Sahin B., Bilgili M. and Akilli H., 2005. The wind power potential of the eastern Mediterranean region of Turkey. *Journal of Wind Engineering and Industrial Aerodynamics* 93 (2), 171–183.
- [5] Zaharim A., Razali A.M., Abidin R.Z. and Sopian K., 2009. Fitting of statistical distributions to wind speed data in Malaysia. *European Journal of Scientific Research* 26 (1), 6–12.

### ABSTRACT

The rapidly increasing demand for energy and the high restriction on pollution levels have led to an increasing interest in large-scale utilization of renewable energies like many countries including Turkey. As wind energy is becoming the world's fastest growing source of renewable and environmental friendly energy, the predictability of wind power generation is essential for the integration of wind energy into the power system. In the present study, a comprehensive evaluation on probability density functions for the wind speed data in Corlu -Tekirdag, Turkey is introduced. The probability density distributions have been derived from long term wind speed data and the distributional parameters are identified. The wind energy potential of the location has been studied based on the Weibull and Rayleigh models which is commonly used in the literature. A time series of hourly monthly average wind speed and the wind power potential at a height of 10 m above ground level in the period between 2003 and 2007 provided by the National Meteorology are assessed for the efficiency of electricity production. To evaluate the performance of the considered distributions, the interpretation of the statistical parameters, namely, slope ( $s$ ),  $R^2$ , mean bias error ( $MBE$ ), and root mean squared error ( $RMSE$ ) is an important step. In the present article, a novel statistical tool using these four statistical parameters is used to evaluate the overall accuracy score of the considered models.

**Keywords:** Wind energy, wind speed, probability distributions, Turkey

## BULANIK MANTIK İLİŞKİLERE DAYALI ARMA(1,1) TİPİ BULANIK ZAMAN SERİSİ ÖNGÖRÜ YÖNTEMİNİN BİR UYGULAMASI

Cem KOÇAK\*, Erol EĞRİOĞLU

\*Hitit Üniversitesi, Sağlık Yüksekokulu, 19000, Çorum, TÜRKİYE, cemkocak@hotmail.com  
Orduokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 5540, Samsun, TÜRKİYE, erole@omu.tr

### 1.Giriş

Literatürde geliştirilen bulanık zaman serisi yöntemlerinin tümü, klasik zaman serisi kuramının sadece AR modeli üzerinde yoğunlaşmış, MA ve ARMA modellerini kullanarak çok daha az sayıda çalışma yapılmıştır. Literatürde önerilen bulanık zaman serisi modellerinin çoğunlukla AR değişkenlerini içeren modeller olması bir çok çalışmada model belirleme hatasının ortaya çıkabileceğini göstermektedir. Model belirleme hatasını gidermek adına, Kocak [1]'nin çalışmasında bulanık mantık grup ilişkilerine dayalı olarak bulanık ilişkilerin belirlendiği yeni bir birinci dereceden bulanık ARMA(1,1) zaman serisi öngörü modeli için bir çözüm algoritması geliştirilmiştir. Kocak [1]'nin çözüm algoritmasında (1)' de belirtilen bulanık mantık ilişki kullanılmaktadır.

$$F(t-1), \varepsilon(t-1) \rightarrow F(t) \quad (1)$$

(1)' deki ifade bulanık zaman serisinin bir önceki değerinden ve bir önceki hata değerinden etkilendiğini göstermektedir ve birinci dereceden bulanık otoregresif hareketli ortalamalar (ARMA(1,1)) zaman serisi öngörü modeli olarak adlandırılır. Kocak [1]'nin (1)'de verilen ifadeyi çözümleyen algoritmasında, Chen [2] yöntemindeki MA (Hareketli Ortalamalar) değişkenlerini kullanılmaktan kaynaklanan olumsuzlukları gidermek amaçlanmaktadır.

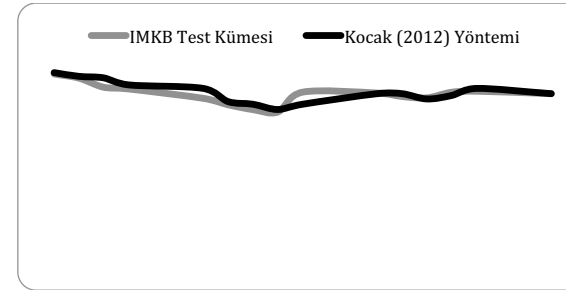
Bu çalışmada, Kocak [1]'in yöntemini, literatürdeki diğer bulanık zaman serisi yöntemleri ile karşılaştırmak amacıyla 95 gözlemlilik 20.05.2008 -29.09.2008 dönemindeki IMKB zaman serisi kullanılmıştır. Verinin son 15 gözlemlilik test kümesi olarak alınan çözümleme sonuçlarına göre, yöntemlerin elde edilen HKOK (Hata Kareler Ortalamasının Karekökü), OMYH (Ortalama Mutlak Yüzdalık Hata) ve YD (Yön Doğruluğu) Çizelge 1'de verilmiştir.

Çizelge 1. Yöntemlerin en iyi sonuçları için IMKB zaman serisinin 15 gözlemlilik test kümesi için elde edilen performansları

Performans	Song ve Chissom [4]	Chen [2]	Chen [3]	Aladağ vd. [5]	Kocak [1]*
HKOK	1155.27	1266,64	1143.56	1145,92	1115,49
OMYH	0.02682	0.03015	0.0261	0,02549	0,02357
YD	0.42857	0,42857	0,57143	0,64286	0,64286

\*En İyi Sonuç

Çizelge 1. incelendiğinde, Kocak [1] yönteminin 1115.49 minimum HKOK, % 2,357 minimum OMYH ve % 64,28 yön doğruluğu değerleri ile diğer yöntemlere göre en iyi öngörü performansına sahip olduğu görülmektedir. Önerilen yöntemden sonra en iyi performansa sahip olan Chen [3] yöntemi ile hesaplanan 1143.56 HKOK değerinin bile önerilen yöntemin 1115.49 HKOK değerinden oldukça büyük bir değer olması, önerilen yöntemin performansının literatürdeki yöntemlerden önemli ölçüde yüksek bir performansa sahip olduğunu saptanmıştır. Şekil 1' de ise test kümesi ile önerilen Kocak [1] yönteminin öngörülerinin birbirine oldukça yakın sonuçlar olduğu görülmektedir.



Şekil 1. 3. IMKB zaman serisinin 15 gözlemlilik test kümesi ve Kocak [1]'in öngörülerinin grafiği

### KAYNAKLAR

- [1] Kocak C. (2013), ARMA(1,1) type first order fuzzy time series forecast method based on fuzzy logic relation, *Mathematical Problems in Engineering* (Submitted Article).
- [2] Chen S.M. (1996), Forecasting enrollments based on fuzzy time-series, *Fuzzy Sets and Systems*, vol.81, pp.311-319.
- [3] Chen S.M. (2002), Forecasting enrollments based on high order fuzzy time series, *Cybernetics and Systems*, vol.33, pp.1-16.
- [4] Song Q. and Chissom B.S. (1993), Fuzzy time Series and its models, *Fuzzy Sets and Systems*, vol.54, pp.269-277.
- [5] Aladağ C.H., Basaran M.A., Egrioglu E., Yolcu U, and Uslu V.R. (2009), Forecasting in high order fuzzy time series by using neural networks to define fuzzy relations, *Expert Systems with Applications*, vol.36, pp.4228-4231.

### ABSTRACT

#### AN APPLICATION OF ARMA(1,1) TYPE FUZZY TIME SERIES FORECAST METHOD BASED ON FUZZY LOGIC RELATIONS

The Most of fuzzy time series models developed in literature utilize only autoregressive (AR) variables. However, it is necessary to have many of daily life time series be explained with Autoregressive Moving Averages (ARMA) models. For this reason, a new first order fuzzy ARMA(1,1) time series forecasting method solution algorithm based on fuzzy logic group relation tables has been developed by Kocak [1]. In this study, Kocak [1]'s method and some methods of literature have been applied on Istanbul Stock Exchange national 100 index (IMKB) and obtained results have comprised in regards to forecasting performance.

**Key Words:** Fuzzy Time Series, Autoregressive Moving Average, ARMA Type, First Order

A STUDY FOR VISUALIZING CATEGORICAL DATA USING MOSAIC PLOT

E. Çiğdem KASPAR\*, Özlem ERGÜT, Yaşar KÜÇÜKARDALI

Assistant Professor, Yeditepe University, Faculty of Medicine, Department of Biostatistics, 34755, İstanbul, TURKEY, ecaltunok@yeditepe.edu.tr  
 Research Assistant, Marmara University, Faculty of Economics, Department of Econometrics, 34730, İstanbul, TURKEY, ozlem.ergut@marmara.edu.tr  
 Professor, Yeditepe University, Faculty of Medicine, Department of Internal Medicine, 34755, İstanbul, TURKEY, yasarkardali@yahoo.com

OBJECTIVE

The researcher should know the importance of fully exploring the features of data before statistical summarization and statistical tests. If the researcher can not be able to see the nature of the data which is inappropriate for statistical tests that she or he has chosen, research will result with mistaken conclusions. The goal of exploratory data analysis is to enable the investigator to assimilate and understand the information embodied within the data. The primary tools of exploratory data analysis are graphics and summary statistics that convert a confusing amount of numbers into pictures and few descriptive numbers that are easily assimilated and understood [1]. There are many types of graphs, but the basic idea is to provide a sketch that quickly conveys general trends in the data to the reader. Mosaic Plot which is attributed to Hartigan and Kleiner (1981), is one of the most powerful tool to visualize the categorical data. Mosaic plot is a graphical method to show the values (cell frequencies) in a contingency table cross-classified by three way and higher way tables. Enhanced mosaic display Friendly [2] achieves greater visual impact by using color and shading to reflect the size of the residual from independence and by reordering rows and columns to make the pattern of association more coherent. The resulting display shows both the observed frequencies and the pattern of deviations from a specified model [3]. In this study, it is aimed to present the Mosaic Plot as an alternative and powerful tool to explore the categorical data with an application.

MATERIALS and METHODS

For the Mosaic Plot, the area of each tile is proportional to the cell frequency,  $n_{ij}$ , variables were independent,  $n_{ij}$  would be proportional to the product of the row and column marginal totals,  $n_{i.}n_{.j}$ , so the tiles in each column would align horizontally. The fact that they do not align reveals an association between these two variables.

In this study, R software was used to illustrate the mosaic plot and the frequency table. The data were taken from a nursing home research which gender, malignancy, bmi, follow up time, number of medicines, number of diseases, dependance and deaths are the variables.

RESULTS

After application, Mosaic Plot and frequency distribution table was obtained (Figure 1) (Table 1). In Figure 1, for a two-way and tree-way tables, shows that the relations among the categories of BMI, malignancy and gender in a sample of individuals.

Table 1. Frequency Table

Variables	FREQUENCY	PERCENT	Variables	FREQUENCY	PERCENT
	N=599	(%)			(%)
<b>GENDER</b>			<b>DEPENDANCE</b>		

1 (Male)	247	41,2	0 (No dependency)	139	23,2
2 (Female)	352	58,8	1 (Semi-dependency)	382	63,8
<b>NUMBER OF MEDICINES</b>			2 (Complete)	78	13
1 (1<3)	40	6,7	<b>DEATHS</b>		
2 (3<6)	318	53,1	0 (Died)	536	89,5
3 (6<9)	197	32,9	1 (Alive)	63	10,5
4 (10+)	44	7,3	<b>BMI</b>		
<b>FOLLOW UP TIME</b>			1 (Thin)	152	25,4
1 (1-6)	287	47,9	2 (Normal)	233	38,9
2 (7-12)	76	12,7	3 (Overweight)	147	24,5
3 (13-24)	82	13,7	4 (Obese)	67	11,2
4 (25-60)	99	16,5	<b>NUMBER OF DISEASES</b>		
5 (60+ )	55	9,2	1 (1-3)	154	25,7
<b>MALIGNANCY</b>			2 (4-6)	276	46,1
0 (Absent)	536	89,5	3 (6+)	169	28,2
1 (Present)	63	10,5			

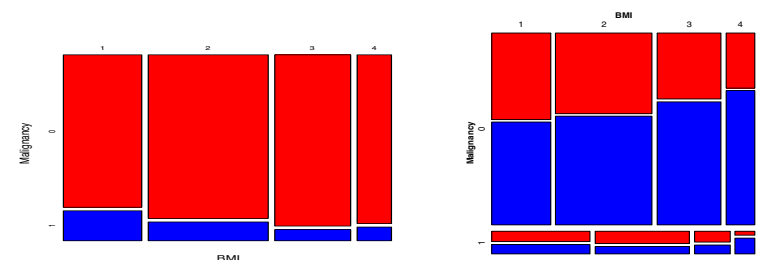


Figure 1. Mosaic Plot

4. CONCLUSION

We regard the mosaic plot as a natural and direct graphic adjunct to multi-way contingency tables. Mosaic plot reveal how variables are related. The representation of the mosaic plot is direct, because the size of each tile

reflects cell frequency. The use of color and shading to represent sign and magnitude of residuals from a model is a natural way to portray the pattern of departure from the model and rep-represents visually the information experienced analysts usually look for in tables of numbers.

#### KAYNAKLAR

- [1] LeBlanc, D. (2004), *Statistics: Concepts and Application for Sciences*, Canada, Jones & Bartlett .
- [2] Friendly, M. (1994), "Mosaic Displays for Multi-way Contingency Tables", *Journal of the American Statistical Association*, Vol.89, No.425, pp.190-200.
- [3] Friendly, M. (2000), *Visualizing Categorical Data*, NC, SAS Institute.
- [4] Hartigan, J.A . and Kleiner B. (1984), "A Mosaic of Television Ratings", *The American Statistics*, Vol.38, No.1, pp.32-35.

**Key Words:** *Mosaic Plot, Visualizing Data, Categorical Data*

#### RASGELE SAYILARI TEST ETME YÖNTEMLERİ

Erdinç AVAROĞLU<sup>1\*</sup> A.Bedri ÖZER<sup>2</sup> Mustafa TÜRK<sup>3</sup>

<sup>1</sup>İnönü Üniversitesi, Bilgi İşlem Daire Başkanlığı, Malatya, erdinç.avaroglu@inonu.edu.tr

<sup>2</sup>Fırat Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Elazığ, bozer@firat.edu.tr

<sup>3</sup>Fırat Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği, Elazığ, mturk@firat.edu.tr

#### Özet

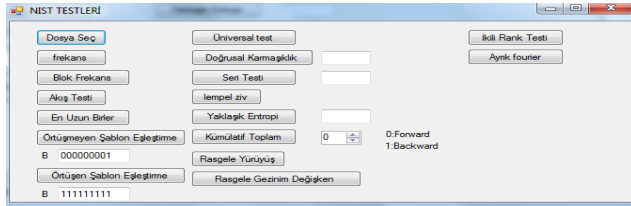
Rasgele sayılar uzun yıllar boyunca geniş kullanım alanına sahip olmuşlardır ve rasgele sayı üretme yöntemleri kullanılarak elde edilmiştir. Elde edilen rasgele sayılar ile rasgele sayı dizileri oluşturulmuştur. Çünkü tek başına bir rasgele sayıyı test etmek mümkün değildir. Oluşturulan rasgele sayı dizilerinin rasgeleliğini test etmek amacıyla FIPS140, NIST, Diehard, Crypt-X gibi çeşitli test paketleri geliştirilmiştir. Rasgele sayı dizilerinin test edilmesinde genel olarak FIPS140 ve NIST testleri tercih edilmiştir. Özellikle bir çok kullanıcı anlaşılması ve uygulanmasının kolay olması sebebiyle FIPS140 testini tercih etmiştir. Ancak daha geçerli olan NIST testlerinin dokümanı ve kodları yayınlanmasına rağmen anlaşılması zor olmuştur. Bu durumun giderilmesi amacıyla bu çalışmamızda NIST testleri incelenmiş olup kullanıcıların ürettikleri rasgele sayı dizilerini kolayca test edebilecekleri program C# dilinde yazılmıştır. Böylece bir çok kullanıcı için problem olan NIST testlerinin kullanımı için kolaylık sağlanmış olacaktır.

**ANAHTAR KELİMELER:** *Rasgelelik, Rasgele sayı üretme, İstatistik test yöntemleri*

#### Giriş

Rasgelelik en basit anlamda kesin olarak bilinmemektir. İstatistikte örnekler arasında ilişki bulunmaması anlamına gelmektedir. Rasgele sayılar ise, belirli bir aralık için tanımlı, oluşma olasılıkları birbirine eşit ve bu sayılar arasında belirli bir ilişki olmayan sayılar olarak tanımlanmıştır. Rasgele sayılar uzun yıllar boyunca simülasyon, şans oyunları, örnekleme gibi çok geniş kullanım alanına sahip olmuşlardır. Örneğin fizikte molekül hareketleri, istatistikte rasgele örneklerin seçimi, hava tahmini, simülasyon, şans oyunları ve kriptografide rasgele sayılara ihtiyaç duyulmuştur. Rasgele sayıların elde edilebilmesi amacıyla çeşitli rasgele sayı üreticileri (RSÜ) geliştirilmiştir. Bu rasgele sayı üreticileri gerçek rasgele sayı üreticileri (GRSÜ), sözde rasgele sayı üreticileri (SRSÜ) ve hibrit rasgele sayı üreticileri olmak üzere sınıflandırılmıştır. Sözde rasgele sayı üreticileri herhangi bir başlangıç (tohum) değeri olmadan başlayamaz. Tohum rasgele seçilmiş olmalıdır. Belirlenen tohum değeri belirli bir algoritmaya tabi tutularak uzun rasgele sayı dizileri üretilmiştir. SRSÜ'nün avantajlı yanı diğer uygulamalara oranla ucuz, kolay gerçekleştirilebilir ve hızlı olması ile donanım ihtiyacına gerek duymamasıdır. Ancak SRSÜ'ler ile üretilen sayılar tohum değeri tespit edildiğinde veya sistemde kullanılan fonksiyonlar yeterince karmaşık olmadığı takdirde tahmin edilebilmiştir. Gerçek rasgele sayı üreticileri gürültü kaynağı olarak kontrol edilemeyen ve kestirilemeyen gerçek fiziksel süreçleri kullanarak rasgele sayı üretmiştir. GRSÜ tarafından üretilen rasgele sayıların özellikleri ve rasgeleliliği fiziksel süreçlerin rasgeleliliğine bağlıdır. Kontrol edilemeyen fiziksel süreçler olduğu takdirde üretilen sayılarda kestirilemez ve kontrol edilemez. GRSU yavaş, maliyetli ve donanıma bağımlı olması dezavantajlı taraftır.

Rasgele sayı üreticilerinden elde edilen rasgele sayı dizilerinin rasgeleliliğini test etmek amacıyla FIPS140, NIST, Diehard, Crypt-X gibi çeşitli test paketleri geliştirilmiştir. Rasgele sayı dizilerinin test edilmesinde genel olarak FIPS140 ve NIST testleri tercih edilmiştir. Özellikle bir çok kullanıcı anlaşılması ve uygulanmasının kolay olması sebebiyle FIPS140 testini tercih etmiştir ve halen bu testin kullanımı devam etmektedir. Ancak daha geçerli olan ve büyük rasgele sayı dizilerini test etmemizi sağlayan NIST testlerinin dokümanı ve kodları yayınlanmasına rağmen anlaşılması zor olmuştur ve bir çok kullanıcı tarafından tercih edilmemiştir. Bu durumun giderilmesi amacıyla NIST testlerinin dokümanı incelenerek testlerin C# programlama dilinde yazılımı yapılmıştır. Yazılan programda matlab ortamında kaotik çekiciler kullanılarak elde edilen rasgele bit dizisi teste tabi tutulmuştur. Test sonucu başarılı olmuştur.



### Şekil 1. Program Kullanıcı Ara Yüzü

Kullanıcı ara yüzü anlaşılır bir şekilde tasarlanmış olup bir çok kullanıcı rasgele sayı dizilerinin rasgeleliliğini test edebilecektir. Bu amaçla yapılan program web ortamında kullanıcıların rasgele bit dizilerini test edebilmeleri için hazır hale getirilecektir. Ayrıca web ortamında 16 adet NIST testlerinin NIST dokümanlarına göre anlaşılabilir ayrıntılı açıklamaları da yapılacaktır.

### Kaynaklar

- [1] Koç, C.K., "Cryptographic Engineering", Springer, 2009
- [2] A statistical test suite for random and pseudo random number generators for cryptographic applications; 2010 : April[NIST800-22Rev1a]
- [3] Deng and Lin, Random Number Generation for the New Century. The American Statistician, May 2000; 54, 2

### ABSTRACT

#### TESTING METHODS OF RANDOM NUMBERS

Random numbers have had a wide range of usage for many years and obtained using the method of generating the random number. With Obtained random numbers, random number sequences created. Because it is not possible to test a random number alone. In order to test the randomness of random number sequences, various test suites have been developed, such as FIPS140, NIST, Diehard, Crypt-X. Generally, FIPS 140 and NIST tests are preferred testing of random number sequences. Since its simplicity to understand and implement, many users preferred FIPS140 test. However, in spite of the publication of the NIST test document and the code, which are more valid, it has been difficult to understand. In this study, in order to resolve this situation, NIST tests are examined, so users can easily produce a random number sequences to test the program written in C#. Thus, the problem for many users the convenience to use the NIST tests will be provided

**Key Words:** Randomness, Random number generation, Statistical testing methods

## KİSMİ M-REGRESYON ALGORİTMASINDAKİ FAİR AĞIRLIK FONKSİYONUN SEÇENEK İRLS AĞIRLIK FONKSİYONLARI İLE DEĞİŞTİRİLMESİ: BİR BENZETİM ÇALIŞMASI

Esra POLAT\*, Süleyman GÜNAY

Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800, Beytepe, Ankara, TÜRKİYE

espolat@hacettepe.edu.tr, sgunay@hacettepe.edu.tr

### 1. Giriş

Kısmi En Küçük Kareler Regresyon'da (Partial Least Squares Regression/PLSR) gizli değişkenlerin (Latent Variables/LV) kestirilmesinden sonra, bu bileşenler üzerinden bir ya da birden fazla bağımlı değişken kestirilir. Bu nedenle PLS'yi sağlamlaştırmak için doğru bir yaklaşım, sağlam regresyon kestiricisinin kısmi bir biçimini oluşturmaktır. Serneels vd. (2005)'de M regresyon kestiricilerinin 'kısmi' bir biçimi önerilmiş ve bu kestirici, kısmi sağlam M (partial robust M/PRM) regresyon kestiricisi olarak adlandırılmıştır. Aykırı değerlerin ağırlığı azaltılarak oluşturulan PRM regresyonunda, kaldıraç gözlemleri ve dikey aykırı değerler olarak bilinen iki aykırı değer türünün de etkisini azaltmak için yinelemeli bir şemada sıfır ile bir arasında değişen sürekli ağırlıklar hesaplanır.

### 2. Kısmi Sağlam M-Kestiricileri

Özellikle bağımsız değişken sayısı gözlem sayısına göre büyük olduğunda ( $p > n$ ) kullanılan LV'ler regresyon modelinde, bağımlı değişken sınırlı sayıda  $k$  tane LV ile modellenir. LV'ler, satırları  $t_i, 1 \leq i \leq n$  vektörlerinden oluşan bileşen matrisi  $T_{n \times k}$ 'ye koyulur. LV regresyon modeli,  $y_i = t_i \gamma + \epsilon_i$  şeklinde verilir.  $\gamma$ 'nin boyutu ( $k$ ) küçük olduğu için, sağlam M kestiricisi kullanılarak bağımlı değişkeni LV'ler ile açıklayarak kestirilir. Klasik regresyonda kullanılan M kestiricilerinden temel farkı, dikey aykırı değerlerin ağırlığını azaltan  $w_i^T$  ağırlıkları  $r_i = y_i - t_i \gamma$  artıklarından ve kaldıraç gözlemlerinin ağırlığını azaltan  $w_i^X$  ağırlıkları  $t_i$  bileşenlerinden hesaplanır. Sonuç olarak ağırlıklar,  $w_i = w_i^T w_i^X$  şeklinde bulunur. Böylece elde edilen kestirici, PRM olarak adlandırılır.

Regresyon için M kestiricilerinin hesaplanması, bir Yinelemeli Olarak Yeniden Ağırlıklandırılmış En Küçük Kareler (Iteratively Reweighted Least squares/IRLS) algoritması ile tamamlanır. Bu nedenle, PRM regresyonu kestiricisini hesaplamak için Cummins ve Andrews (1995)'te önerilen IRPLS yöntemine benzer, bir yinelemeli olarak ağırlıklandırılmış PLS algoritması kullanılır. Ancak PRM için kullanılacak algoritmada, sağlam başlangıç değerleri ve hem artık hem de bileşen uzaylarındaki uzaklıklara dayanan ağırlıklar kullanılır. Serneels vd. (2005)'te kullanılan ağırlıklar özgün IRPLS makalesinde verilen birkaç ağırlık fonksiyonundan biri olan  $1/\sqrt{1 + |z/c|}$  şeklinde verilen Fair fonksiyonu kullanılarak bulunur. Buradaki  $c$ , ayarlama sabitidir ve  $c=4$  olarak alınmıştır. Serneels vd. (2005)'de Fair ağırlık fonksiyonundan başka ağırlık fonksiyonlarının da kullanılabileceği ve seçilen ağırlık fonksiyonu için herhangi bir en iyi olma özelliği iddia edilmemiştir. IRLS'de en sık kullanılan ağırlık fonksiyonları Çizelge 1'de verilmiştir.



**Çizelge 1. IRLS Ağırlık Fonksiyonları.**

Yöntem	Ağırlık Fonksiyonu	Varsayılan Sabit
Bisquare	$\begin{cases} 1 - (z/c)^2, &  z  \leq c \\ 0, &  z  > c \end{cases}$	c=4.685
Cauchy	$1/[1 + (z/c)^2]$	c=2.385
Huber	$\begin{cases} 1, &  z  < c \\ c/ z , &  z  \geq c \end{cases}$	c=1.345
Talworth	$\begin{cases} 1, &  z  \leq c \\ 0, &  z  > c \end{cases}$	c=2.795

**3. Benzetim Çalışması**

Sağlam regresyonda, aykırı değerlerin ağırlığını azaltma yöntemi önemlidir. Filzmoser (2010), sağlam regresyonda yumuşak ağırlıklandırma (soft weighting) olarak bilinen [0,1] aralığındaki sürekli ağırlıkların kullanılmasının ya da sert ağırlıklandırma (hard weighting) olarak da bilinen 0/1 ağırlıklarının kullanılmasının regresyon kestiricisinin etkinliğini etkileyeceğini ve özellikle, p>n olan büyük boyutlu veri kümelerinde yumuşak ağırlıklandırmanın daha etkin sonuçlar verdiğini ifade etmiştir. Bu çalışmanın amacı, PRM algoritmasında kullanılan Fair ağırlık fonksiyonunu Çizelge 1’de verilen seçenek dört ağırlık fonksiyonunu kullanarak değiştirmek, böylece değiştirilmiş ve özgün PRM algoritmalarını klasik PLSR algoritması SIMPLS ile farklı hata terimlerine sahip modeller için etkinlik açısından karşılaştırmaktır. Ayrıca, González vd. (2008)’de verilen simülasyon düzenini kullanarak etkinliğe ek olarak, yumuşak ve sağlam ağırlıklandırmanın PRM’de veriyeye uyum ve kestirim açısından etkilerini, normal dağılım için farklı aykırı değer türleri yaratarak karşılaştırmaktır.

**KAYNAKLAR**

- [1] Cummins D. J. and Andrews C. W. (1995), Iteratively Reweighted Partial Least Squares: A Performance Analysis By Monte Carlo Simulation, Journal of Chemometrics, 9, 489-507.
- [2] Filzmoser P., Semeels S., Maronna R., Van Espen P. J. (2009), Robust multivariate methods in chemometrics, Comprehensive Chemometrics, 681-722.
- [3] González J., Peña D., Romera R. (2008), A robust partial least squares regression method with applications, Journal of Chemometrics, 23, 78–90.
- [4] Semeels S., Croux C., Filzmoser P., Van Espen P. J. (2005), Partial Robust M-regression, Chemometrics and Intelligent Laboratory Systems, 79, 55-64.

**ABSTRACT**

**THE MODIFICATION OF FAIR WEIGHT FUNCTION IN PARTIAL ROBUST M-REGRESSION WITH ALTERNATIVE IRLS WEIGHT FUNCTIONS: A SIMULATION STUDY**

The aim of this study is to see the effects of modification of Fair weight function used in original robust PLSR method PRM by alternative weight functions, furthermore, is to examine the effects of soft and hard weightings in this algorithm in terms of efficiency, goodness of fit and prediction.

**Key Words:** Partial Robust M-Regression, IRLS weight functions, hard and soft weighting, efficiency, goodness of fit, prediction, robust.

**KLASİK VE SAĞLAM YANLI REGRESYON YÖNTEMLERİNİN KARŞILAŞTIRILMASI: TÜRKİYE’DEKİ İŞSİZLİK ORANI ÖRNEĞİ**

Esra POLAT\*, Semra TÜRKAN, Süleyman GÜNAY

Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800, Beytepe, Ankara, TÜRKİYE

sturkan@hacettepe.edu.tr, espolat@hacettepe.edu.tr, sgunay@hacettepe.edu.tr

**1. Giriş**

Veri kümesinde çoklu bağlantı olduğunda klasik Çoklu Doğrusal Regresyon (Multiple Linear Regression/MLR) kestiricileri çok büyük bir varyansa sahip olur ve En Küçük Kareler (Least Squares/LS) yöntemi ile elde edilen parametre tahminleri güvenilir değildir. Bu durumda Ridge Regresyon (Ridge Regression/RR), Temel Bileşenler Regresyonu (Principal Component Regression/PCR) ve Kısmi En Küçük Kareler Regresyonu (Partial Least Squares Regression/PLSR) gibi yanlı kestirim yöntemleri kullanılır. PLSR yöntemi için kestirimler yapılırken en sık kullanılan PLSR algoritması SIMPLS’dir. Ancak, her üç klasik yöntem için de sonuçlar aykırı gözlemlerden etkilenir. Bu nedenle, literatürde bu yöntemlerin çeşitli sağlam karşılıkları önerilmiştir. Bu çalışmanın amacı, yanlı klasik ve sağlam regresyon yöntemlerini kullanarak çoklu bağlantı ve aykırı değer varlığında Türkiye’deki işsizlik oranını tahmin etmek ve bu yöntemleri karşılaştırmaktır.

**2. Klasik ve Sağlam Ridge Regresyon Yöntemleri**

Hoerl ve Kennard (1970) tarafından önerilen RR tahmin edicisi,  $\hat{\beta}_{RR} = (\mathbf{X}^T \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}$  biçimindedir. Burada  $\mathbf{I}$ , p x p boyutlu birim matrisi ve k, yanlılık sabitini gösterir. Çoklu bağlantı olduğunda RR iyi sonuç vermesine rağmen, normallik sağlanmadığında ve aykırı değerler varlığında sağlam bir yöntem değildir. Bu nedenle, RR yöntemi ile bazı sağlam kestirim yöntemlerinin birleştirilmesi gerekir. Marona (2011), RR için sağlam MM kestiricisini önermiştir.  $\hat{\beta}_{ini}$  başlangıç kestiricisi,  $\mathbf{r} = \mathbf{r}(\hat{\beta}_{ini})$  ve  $\hat{\sigma}_{ini}^2$ , r’nin M –ölçek kestirimi

olmak üzere  $\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{r_i}{\hat{\sigma}_{ini}} \right) = \delta$  fonksiyonundan elde edilir. Burada  $\rho_0$  sınırlandırılmış  $\rho$ -fonksiyonu ve  $\delta$

seçilebilen sabit bir değerdir. Ridge MM kestiricisi  $L(\mathbf{X}, \mathbf{y}, \beta) = \hat{\sigma}_{ini}^2 \sum_{i=1}^n \rho \left( \frac{r_i(\beta)}{\hat{\sigma}_{ini}} \right) + k \|\beta_1\|^2$  fonksiyonu

minimum yapılarak elde edilir. Burada  $\mathbf{r} = (r_1, \dots, r_n)'$  ve  $\hat{\beta}_0 \mathbf{I}_n - \mathbf{X} \hat{\beta}_1$  artık vektörünü,  $\rho$  sınırlandırılmış başka bir  $\rho$ -fonksiyonunu ( $\rho \leq \rho_0$ ) göstermektedir.

**3. Klasik ve Sağlam Temel Bileşenler Regresyon Yöntemleri**

PCR ve PLSR, k<p olmak üzere k boyutlu t<sub>i</sub> bileşenlerini kullanarak, p boyutlu x-bağımsız değişkenler kümesi ve q boyutlu y-bağımlı değişkenler kümesi arasındaki ilişkinin varlığını açıklayan, Eş. (1)’de gösterilen iki doğrusal modele dayanır. Burada P<sub>p,k</sub>, x yükleri matrisini ve A<sub>k,q</sub>, y<sub>i</sub>’nin t<sub>i</sub> üzerine regresyonundan elde edilen eğim matrisini gösterir. f<sub>i</sub> ve g<sub>i</sub>, hata terimlerini gösterir. İki doğrusal model, özgün bağımsız değişkenler türünden y<sub>i</sub> = β<sub>0</sub> + B’<sub>q,p</sub>x<sub>i</sub> + e<sub>i</sub> şeklinde yazılır. Burada, B<sub>p,q</sub> = P<sub>p,k</sub>A<sub>k,q</sub> ve β<sub>0</sub> =  $\bar{y} - B'_{q,p}\bar{x}$ ’dir.

$$\begin{aligned} x_i &= \bar{x} + P_{p,k} t_i + f_i \\ y_i &= \bar{y} + A'_{q,k} t_i + g_i \end{aligned} \quad (1)$$

PCR yöntemi, Temel Bileşenler Analizi (Principal Component Analysis/PCA) ile LS regresyonu birleştirdiğinden sonuçlar aykırı gözlemlerden etkilenir. Bu nedenle, Hubert ve Vanden Branden’de (2003) önerilen sağlam PCR (robust PCR/RPCR) yönteminde ilk olarak, bağımsız değişkenler üzerine sağlam PCA yöntemi olan ROBPCA uygulanır. x<sub>i</sub>’nin kovaryans matrisinin sağlam kestiricisi olarak p<n olduğunda, En Küçük Kovaryans Determinantı (Minimum Covariance Determinant/MCD) kestiricisi kullanılır. MCD

kestiricilerin  $p > n$  olan yüksek boyutlu veri kümelerine uygulanabilmesi için ilk önce, ROBPCA yöntemi ile bu veri kümeleri indirgenir. RPCR yöntemi, projeksiyon izleme tekniklerini küçük boyutlardaki sağlam kovaryans kestirimi ile birleştirir. Daha sonra, sağlam regresyon yöntemi uygulanır. Eğer tek bir bağımlı değişken var ise, yeniden ağırlıklandırılmış LTS regresyon tercih edilir, aksi durumda MCD regresyon yapılır.

### 3. Klasik ve Sağlam Kısmi En Küçük Kareler Regresyon Yöntemleri

SIMPLS algoritmasının ilk aşamasında,  $t_i$  skorları varyans kovaryans matrisleri  $S_{xy}$  ve  $S_x$ 'e dayalı olarak hesaplanır ve ikinci aşamasında, bu  $k$  boyutlu  $t_i$  skorlarından MLR ile  $y_i$ 'ler kestirilir. Bu nedenle, algoritmanın her iki aşaması da aykırı değerlere karşı dirençli değildir. Hubert ve Vanden Branden (2003), SIMPLS algoritmasındaki  $S_{xy}$  ve  $S_x$ 'i, sağlamlaştırılmış kestirimleri ile yer değiştirerek, değişkenler üzerine ROBPCA ve MLR yerine sağlam regresyon yöntemi olan ROBPCA regresyonunu uygulayarak sağlam RSIMPLS algoritmasını önermiştir.

### KAYNAKLAR

- [1] Engelen S., Hubert M., Vanden Branden K., Verboven S. (2004), Robust PCR and Robust PLSR: A comparative study, Theory and Applications of Recent Robust Methods, Statistics for Industry and Technology, 105–117.
- [2] Göktaş A. ve İşçi Ö. (2010), Türkiye’de İşsizlik Oranının Temel Bileşenli Regresyon Analizi ile Belirlenmesi, SÜ İİBF Sosyal ve Ekonomik Araştırmalar Dergisi, 14(20), 279-294.
- [3] Hubert M. and Vanden Branden K. (2003), Robust methods for Partial Least Squares Regression, Journal of Chemometrics, 17, 537-549.
- [4] Maronna R.A. (2011), Robust ridge regression for high-dimensional data, Technometrics, 53(1): 44-53

### ABSTRACT

#### THE COMPARISON OF CLASSICAL AND ROBUST BIASED REGRESSION METHODS: AN EXAMPLE OF THE RATE OF UNEMPLOYMENT IN TURKEY

In case of multicollinearity, the classical Multiple Linear Regression estimates are unreliable. Hence, in this case biased regression methods such as Ridge Regression, Principal Component Regression and Partial Least Squares Regression are used. The aim of this study is to forecast the unemployment in Turkey in existence of multicollinearity and outliers by using three biased classic regression methods and their robust counterparts and is to compare these methods.

**Key Words:** Partial Least Squares Regression, Principal Component Regression, Ridge Regression, Robust, Unemployment.

## ORTAÖĞRETİM ÖĞRENCİLERİNİN MATEMATİK DERSİ TUTUM VE İLGİLERİNİN BELİRLENMESİ: BİGADIÇ ÖRNEĞİ

Fatih ÇEMREK\*, Özkan ÖZCAN\*

\*Eskişehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi İstatistik ESKİŞEHİR

fcmrek@gmail.com, ozkanozcan1903@hotmail.com

### GİRİŞ

Öğrenmeyi etkileyen en önemli duyuşsal özelliklerden biri olarak nitelendirilen tutum, bireylerin öğrenmelerini olumlu ya da olumsuz yönde etkileme gücüne sahiptir. Matematik dersi, bireylerin problem çözme, analitik düşünme, yorumlama becerilerini artıran temel bir ders olarak ilköğretim ve ortaöğretim programlarında başlangıçtan beri yer almaktadır. Matematiğe yönelik ilgi ve merak uyandırılması ve bu konulara ilişkin olumlu tutumların geliştirilmesi, öğrencilerin bilişsel yeterliklerinin gelişmesine katkı sağlar. Bu araştırmanın amacı, Duatepe (1999) tarafından geliştirilmiş olan matematik dersi tutum ölçeği kullanılarak ortaöğretim öğrencilerinin matematik dersine karşı olan tutum ve davranışlarını belirlemeye çalışmaktır. Ölçeğin geçerlik ve güvenilirlik çalışması, Balıkesir ili Bigadiç Cumhuriyet Lisesi’nde okuyan 180 öğrenci üzerinde gerçekleştirilmiştir. Yapı geçerliği için başvurulan KMO Barlett katsayısı 0,925 bulunmuştur. Faktör analizi sonucunda ölçeğin 3 faktörde toplandığı gözlenmiştir. Faktörler alan yazına dayalı olarak, zevk alma, ilgi-sevgi ve korku olarak adlandırılmıştır. Üç faktör varyansın toplamda %53,876’sını açıklamaktadır

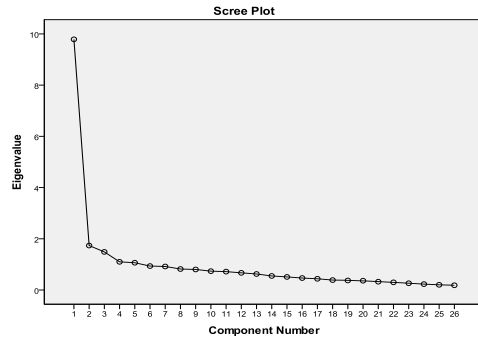
### BULGULAR

Ölçeğin yapı geçerliliğini belirlemek için yapılan faktör analizi ile ölçekte yer alan maddelerin matematik dersi tutumu ile ilgili hangi faktörleri ölçtüğü ortaya çıkarılmıştır. Matematik Dersi Tutum Ölçeği faktör analizi çalışması, temel bileşenler analizi (Principle Component Analysis) tekniği uygulanarak yapılmıştır. Bu çalışmayı desteklemek ve faktör sayısına sağlıklı karar verebilmek amacıyla faktörlerin öz değerlerine dayanan Scree sınaması grafiği de incelenmiştir.

Scree sınaması sonrasında ölçekten maddelerin azaltılma işlemine geçilmiştir. Tek ve iki maddelik faktör oluşturan maddelerin ölçekten çıkarılması sonucunda 3 faktörlü 23 maddelik tutum ölçeği belirlenmiştir. Dolayısıyla 3 madde ölçekten atılmıştır

Faktör	Öz değer	Varyans Yüzdesi	Toplam Varyans Yüzdesi
1	9,207	40,029	40,029
2	1,703	7,404	47,434
3	1,482	6,442	53,876

Tablo-1. Matematik Dersine Yönelik Tutum Ölçeğinin Faktör Yapıları (Matematik Dersine Yönelik Tutum Ölçeğinin Faktör Yapısına İlişkin Döndürülmemiş Varyans Değerleri)



Şekil-1. Scree Sınaması Grafiği

#### Tartışma Ve Sonuç

Araştırma kapsamında kullanılan ölçeğin yapı geçerliğini kontrol etmek amacıyla gerçekleştirilen faktör analizi sonucu ölçekte kalmasına karar verilen 23 maddenin belirli bir yapıyı ölçebileceği sonucuna ulaşılmıştır. Faktör analizi sonuçları ölçeğin üç alt boyuttan oluşan bir yapıya sahip olduğunu ortaya koymaktadır. Bu faktörler, alan yazına dayalı olarak; zevk alma, ilgi-sevgi ve korku biçiminde adlandırılmıştır.

#### KAYNAKLAR

- [1] Aşkar, P. (1986). "Matematik Dersine Yönelik Tutumu Ölçen Likert-Tipi Bir Ölçeğin Geliştirilmesi". Eğitim ve Bilim. Cilt:11, sayı:62. (31-36).
- [2] Avcı, E., Coşkuntuncel, O.ve İnandı Y.,(2011), "Ortaöğretim On İkinci Sınıf Öğrencilerinin Matematik Dersine Karşı Tutumları" Mersin Üniversitesi Eğitim Fakültesi Dergisi, Cilt 7, Sayı: 1, Haziran 2011, ss.50-58.
- [3] Baloğlu, M. (2001). "Matematik Korkusunu Yenmek". Kuram ve Uygulamada Eğitim Bilimleri Dergisi. Cilt:1, sayı:1. (59-76).
- [4] Büyüköztürk, Ş. (2002). "Sosyal Bilimler için Veri Analizi El Kitabı". Ankara: Pegem Yayıncılık (2.Baskı).

#### DETERMINATION OF SECONDARY SCHOOL STUDENTS 'MATHEMATICS COURSE ATTITUDE AND INTERESTS: BİGADİÇ EXAMPLE

Regarded as one of the important factors that affect learning attitudes, individuals have the power to affect learning positively or negatively. Awakening of curiosity and interest towards mathematics and the development of positive attitudes on these issues, contribute to the development of cognitive competencies of students among. The purpose of this study, Duatpe (1999), which was developed by secondary school students using mathematics to mathematics attitude scale is to determine the attitudes and behaviors.

**Key Words:** (mathematics, math lesson, math attitude, math interest)

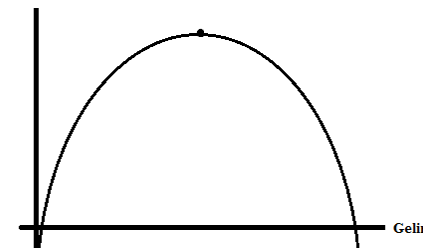
#### ÇEVREYE UYARLANMIŞ KUZNETS EĞRİSİ: PORTEKİZ ÜZERİNE BİR UYGULAMA

Mehmet MERT Hakan BOZDAĞ  
Akdeniz Üniversitesi İ.İ.B.F. Ekonometri Bölümü, Süleyman Demirel Üniversitesi İ.İ.B.F. Ekonometri BölümüMail: mmert@akdeniz.edu.tr, hakanbozdog@sdu.edu.tr

#### Çevreye Uyarlanmış Kuznet Eğrisi

Simon Kuznets (1955), iktisadi büyüme ve kalkınma ile birlikte gelir dağılımının önce bozulacağını, ancak gelir artışının devam etmesi ile birlikte gelir dağılımındaki adaletsizliğin azalacağını ileri sürmüştür. Literatüründe Kuznets Eğrisi adı ile anılan bu görüş, 1990'lı yıllarla birlikte iktisadi büyüme ile çevre kirliliği ve çevre tahribatı için de kullanılır hale gelmiştir. İktisadi büyüme ile birlikte çevre kirliliğinin ve/veya çevre tahribatının artacağı, belli bir gelir düzeyinden sonra azalacağına ilişkin görüşe literatürde "Çevreye Uyarlanmış Kuznets Eğrisi" adı verilmektedir.

#### Çevre Kirliliği



#### Şekil 1. Kuznets Eğrisi

Şekilde 1' de gelir düzeyi yükselince çevre kirliliği artmakta, söz konusu düzeyden itibaren azalan bir seyir izlemektedir. Gelir düzeyi ile çevre kirliliği arasındaki ilişkinin neden ters U biçimli bir seyir izlediğinin teorik düzeyde açıklanmasında çeşitli faktörlerin etkili olduğu ileri sürülmektedir.

Çevreye Uyarlanmış Kuznets Eğrisi ilişkisinin açıklanmasında genel olarak ölçek, kompozisyon ve teknoloji etkileri belirtilmektedir.

Çevreye Uyarlanmış Kuznets Eğrisi'nin araştırılmasına yönelik çalışmalarda genel olarak aşağıdaki modelin tahmin edildiği görülmektedir:

$$y_{it} = \alpha_i + \beta_1 x_{it} + \beta_2 x_{it}^2 + \beta_3 x_{it}^3 + \beta_4 z_{it} + \epsilon_{it} \quad (1)$$

1 nolu denklemde, bağımlı değişken olarak kişi başı CO2 emisimi, bağımsız değişkenler olarak ise kişi başına milli gelir, kişi başına milli gelirin karesi ve küpü kullanılmaktadır.  $\beta$  katsayısının işaretine göre kullanılan çevre göstergesi ile milli gelir arasındaki ilişkinin şekli de değişmekte, kimi zaman "U" ve ters "U" elde edilirken, kimi zaman ise "N", ters "N" ve ters "J" ilişkileri de elde edilmektedir.

Bu çalışmada Portekiz için 1961-2009 yılları arasında tahmin edilen VAR modelinde, bağımlı değişken CO2 emisyonu, bağımsız değişkenler ise kişi başına milli gelir (GDP), kişi başına milli gelirin karesi ve kişi başına milli gelirin küpü verilerinden yararlanılmıştır. Veri seti Dünya Bankası internet sitesinden temin edilmiştir.

VAR modelinde kullanılacak serilerin ilk olarak birim kök içerip içermediği kontrol edilmiş; tüm serilerin ikinci dereceden durağan olduğu tespit edilmiştir.

Değişkenler arasındaki uzun dönem ilişkisinin varlığı Johansen Eşbütünleşme testi ile belirlenmiştir. Oluşturulan VAR modelinin gecikme seviyesi AIC ve SIC kriterine göre 3 olarak belirlenmiştir. VAR modeli yardımı ile değişkenler arasındaki Hata düzeltme modeli kullanılarak CO2 emisimi ile GDP arasındaki Granger nedensellik analizi yapılmıştır.

Sonuç olarak 1961-2009 yılları arasında Portekiz için Çevresel Kuznet hipotezinin geçerli olup olmadığı tespit edilmeye çalışılmıştır.

#### KAYNAKLAR

- [1] Fodha M. and Zaghoud O. (2010), Economic growth and pollutant emissions in Tunisia: An empirical analysis of the environmental Kuznets curve, Energy Policy, 38 11301156
- [2] Grossman, G.M., Krueger, A.B., (1995). Economic growth and the environment. The Quarterly Journal of Economics 110 (2), 353-377.
- [3] Huang, W.M., Lee, G.W.M. & Wu, C.C., (2008). GHG emissions, GDP growth and the Kyoto Protocol: A revisit of Environmental Kuznets Curve hypothesis. Energy Policy, 36, 239-247.
- [4] Kijima, M., Nishibe K. & Ohya A., (2010). Economic modeling for the environmental Kuznets curve: A survey. Journal of Economic Dynamics & Control, 34, 1187-1201.
- [5] Kuznets, S., (1955). Economic growth and income inequality. American Economic Review 45 (1), 1–28.

#### ABSTRACT

##### ENVIRONMENTAL KUZNETS CURVE: AN EMPIRICAL ANALYSIS FOR PORTUGAL

The relationship between environmental quality and economic growth has been empirically modeled through emissions–income relationship so far, and the outcome of most of these studies has been formulated by the so called environmental Kuznets curve (EKC) hypothesis. According to the environmental Kuznets curve hypothesis, the relationship between per-capita GDP and per capita pollutant emissions has a U shape, an inverted-U shape, a N shape and an inverted-N shape. This implies that, economic growth may be profitable for environmental quality. This paper investigates the validity of Environmental Kuznets Curve Hypothesis for Portugal, using time series data for 1961-2009 period. Firstly, EKC relationship for CO2 emissions for Bosnia & Herzegovina, over a time period of 1992–2009, has been tested and then tried to determine relationship between CO2 emissions–income.

**Key Words:** Economic Growth, Environmental Kuznets Curve, VAR Analysis.

##### 2012 VEFAT VERİLERİNE GÖRE TEK YAŞLAR İÇİN TÜRKİYE YAŞAM TABLOSUNUN GÜNCELLENMESİ

Hanife TAYLAN ve Gökkan YAPAR  
Dokuz Eylül Üniversitesi, Fen Fakültesi, İstatistik Bölümü  
e-mail: hanife.taylan@deu.edu.tr ve guckan.yapar@deu.edu.tr

##### 2012 Türkiye Yaşam Tablosu

Aktüerya biliminin temelini oluşturan yaşam tabloları bir nüfusun ölüm ve yaşam olasılıklarını hesaplayarak gelecekte beklenen ömürlerini elde eder. Yaşam tablolarının barındırdığı en önemli değişken olan gelecekte beklenen yaşam süresi, demografik olarak bir gelişmişlik göstergesi olmasının yanında sigortacılık sektörü içinde önemli bir bilgidir. Gelecekte insanların ortalama ne kadar yaşayacağını bilmesi sosyal güvenlik sistemi ve özel sigortacılık sektöründe prim, rezerv, teminat hesaplamaları, emeklilik sürelerinin belirlenmesi ya da ürün oluşturma gibi finansal yönetimin temelini oluşturan alanlarda kullanılır. Günümüzde yaşam koşullarının iyileşmesi, tıp ve teknolojik alanda yaşanan gelişmeler demografik yapının değişmesine yol açmıştır. Bu gelişmeler ise uzun ömürlülüğe ve yaşlı nüfusun artmasına neden olmuştur. Değişen bu yapıya uyum sağlamak, sektörün yanlış hesaplamalardan korunması ve gelecekte karşılaşılabileceği riski yönetebilmesi açısından toplumun demografik yapısını yansıtan yaşam tablolarının sektörde kullanılması büyük önem taşımaktadır.

Yaşam tabloları dönem ve kuşak yaşam tabloları olmak üzere iki tipte oluşturulur (Preston vd., 2001). Kuşak yaşam tablosu aynı zaman aralığında doğan tüm bireylerin her bir üyesinin ölümüne kadar geçen sürenin modellenmesi ile oluşturulan tablolara denir. Dönem yaşam tabloları ise belirli bir zaman aralığında yaşayan mevcut nüfusun yaşam sürelerinin modellenmesiyle oluşturulur. Kuşak yaşam tabloları bir kuşağın ölüm davranışlarını incelediğinde yaklaşık yüz yıllık ölüm ve yaşam verisine ihtiyaç duyar ve bu çoğu ülke için henüz mevcut olmayan bir veri setidir. Ülkemizde ise ölüm ve yaşam verileri TÜİK tarafından yayımlanmaktadır. Bu verilerin doğru ve tüm nüfusu yansıtacak şekilde yayımlanması gerekir. 2009 yılı itibarıyla ölüm verileri Türkiye geneli için yayımlanmaya başlanmıştır ve kayıt sistemindeki bu değişiklik ülkemiz için dönem yaşam tablolarının doğru bir şekilde oluşturulmasına olanak tanımıştır. Bu çalışmada Türkiye geneli ölüm ve yaşam verileri kullanılarak ülkemizin kendi demografik yapısını yansıtan daha önceki çalışmalarımızda oluşturduğumuz cinsiyet ve yaş grupları bazında Özetlenmiş Dönem Yaşam Tablosu 2012 ölüm verileri kullanılarak güncellenmiştir. Ayrıca daha önce yaş grupları ve cinsiyet bazında oluşturulan yaşam tablosu bu çalışmada tek yaşlar için genişletilmiştir.

Bir ülkenin belirli bir zaman aralığındaki mevcut demografik yapısını yansıtan ve bu oranı baz alarak oluşturulan yaşam tablosuna dönem yaşam tablosu denir. Öncelikle demografi biliminde oran herhangi bir olayın (ölüm, doğum, göç vb.) gerçekleşme sayısının bu riske maruz kalan kişi yıl sayısına bölümüyle elde edilir. Tablonun oluşturulması için tanımlanacak ilk değişken yaşa özgü ölüm oranı olarak adlandırılan,  ${}_nM_x$ , belirli bir zaman aralığında  $x$  ve  $x+n$  yaşları arasında ölen kişi sayısının yine aynı yaş aralığında yaşayan kişi yıl sayısına bölünmesiyle elde edilir (Arias, 2010).

$${}_nM_x = \frac{{}_nD_x}{{}_nN_x} \approx {}_n m_x$$

Ortalama yaşayan kişi yıl sayısı  ${}_n a_x$  ile gösterilir ve elde edilmesinde kullanılan 4 ayrı yöntem mevcuttur (Preston vd., 2001). Bu çalışmada gerçek gözlemlerden faydalanılmıştır. Yaşa özgü ölüm olasılığı (ölüm hızı),  $x$  yaşında bir kimsenin  $n$  yıl içindeki ölüm olasılığını ifade eder ve yaşa özgü ölüm oranı ve yaşayan ortalama kişi yıl sayısı değişkenleri kullanılarak elde edilir. Bu dönüşüm gerçek kuşak davranışlarını baz alır ve bu dönüşüm Greville (1943) and Chiang (1968) tarafından tanımlanmıştır.

$${}_nL_x = n \cdot l_{x+n} + {}_n a_x \cdot {}_n d_x \rightarrow {}_n q_x = \frac{{}_n d_x}{l_x} = \frac{{}_n m_x}{1 + (1 - {}_n a_x) {}_n m_x}$$

ve sadece bir parametre gerektirir ve  $o$  parametre ise  $q_x$ 'dir. Ortalama yaşayan kişi yıl sayısı dönem yaşam tablolarının oluşturulmasında önemli bir yere sahiptir ve  ${}_n m_x \rightarrow {}_n q_x$  dönüşümü ile yaşa özgü ölüm olasılıklarının gözlenen verilerden elde edilmesine olarak sağlar (Preston vd., 2001). Yaşa özgü ölüm olasılıklarının elde edilmesi aşağıdaki dönüşüm kullanılır.

$${}_n q_x = \frac{n \cdot m_x}{1 + (1 - a_x) \cdot n \cdot m_x} \quad q_\infty = 1$$

Daha sonra sırasıyla yaşam tablosu fonksiyonları olan yaşama olasılıkları,  ${}_n p_x$ , yaşayan kişi yıl sayısı,  ${}_n L_x$ , toplam kişi yıl sayısı,  $T_x$ , elde edilmiştir (Selvin, 2008). Yaşam tablosu analizinden faydalanılarak x yaşında bir kimsenin beklenen yaşam ömrü aşağıdaki gibi hesaplanır.

$$e_x = \frac{T_x}{l_x}$$

#### KAYNAKLAR

- [1] Arias E., (2010). United States life tables by Hispanic origin. National Center for Health Statistics. Vital Health Stat 2(152).
- [2] Bell, F., C., Miller, M., L., (2005), Life Tables for the United States Social Security Area 1900-2100. Social Security Administration. Actuarial Study, 120.
- [3] Chiang, C.L. (1972). On constructing current life tables. Journal of the American Statistical Association 67, 538-541.
- [4] Preston, S. H., Heuveline, P., Guillot, M., 2001. Demography Measuring and Modeling Population Process. Blackwell, INC., Publications, USA.
- [5] Shryock, H.S., Siegel, J.S., and Associates, 1971. The methods and materials of demography. U.S. Government Printing Office, Washington, D.C.

#### ABSTRACT

##### UPDATING TURKEY LIFE TABLE FOR A SINGLE AGE BY USING 2012 DATE DATA

Life tables are widely used in many areas. In particular, these tables are crucial in actuary and demography. In actuarial science and demography, a life table is a table which shows, for each age, what the probability is that a person of that age will die before his or her next birthday. Developments countries use a life table which reflects demographic structure of their countries. The purpose of this study is updated a complete period life table by gender which is demonstrated demographic features of Turkey with Turkey Death Data Set.

**Key Words:** Life expectancy, complete period life table, age specific death probability, life/mortality table.

#### TÜRKİYE'DE RİSK DEĞERLENDİRMESİ YÖNTEM VE UYGULAMALARINA GENEL BİR BAKIŞ

Burak BİRGÖREN, Kezban BULUT

Kırıkkale Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü,71541, Yahşıhan/KIRIKKALE  
bbirgoren@yahoo.com, kezbanbulut@yahoo.com

#### Giriş

Türkiye'de 2012 yılında çıkarılan 6331 sayılı İş Sağlığı ve Güvenliği Kanununa göre işyerinde var olan ya da dışarıdan gelebilecek tehlikelerin belirlenmesi, bu tehlikelerin riske dönüşmesine yol açan faktörler ile tehlikelerden kaynaklanan risklerin analiz edilerek derecelendirilmesi ve kontrol tedbirlerinin kararlaştırılması amacıyla yapılması gerekli çalışmalar Risk Değerlendirme olarak tanımlanmaktadır [1]. Kanun sektör ya da büyüklük farklı gözetmeksizin tüm işverenlerin risk değerlendirmesi yapma veya yaptırması yükümlülüğü getirmektedir.

Kanuna istinaden çıkarılan Risk Değerlendirmesi Yönetmeliğine göre risk değerlendirmesi, tüm işyerlerinde kuruluş aşamasından başlamak üzere tehlikeleri tanımlama, riskleri analiz etme, riskleri kontrol altına alma, dokümantasyon ve yapılan çalışmaları yenileme aşamaları izlenerek gerçekleştirilir [2].

Türkiye değişik istatistikler dikkate alınarak yapılan değerlendirmelerde gerek AB ülkelerine gerekse OECD ülkelerine kıyasla iş kazaları açısından en kötü durumdaki ülkeler arasında yer almaktadır. Bunun ülke ekonomisine ciddi ölçülerde zarar verdiği bilinmektedir. Etkin risk değerlendirmesi çalışmalarının iş kazalarını önlemede büyük fayda sağlayacağı beklenmektedir.

#### Risk Değerlendirmesi

AB ülkelerinde iş sağlığı ve güvenliği anlayışı proaktif (önleyici) yaklaşımı esas alır ve risk değerlendirmesi proaktif yaklaşımın temel aracı kabul edilmektedir. Türkiye'de risk değerlendirmesi yakın zamana kadar sadece büyük ölçekli sanayi kuruluşları tarafından yaygın olarak kullanılıyor iken şimdi lokantalar, kuaförler gibi küçük esnafın ve hastaneler, bankalar gibi her tür işletmenin risk değerlendirmesi yapmasını zorunlu kılan büyük çaplı bir kanuni düzenlemeye gidilmiştir. Ancak kanunun uygulama ayağında birçok zorluklar beklenmelidir, çünkü büyük ölçekli firmaların profesyonel ekiplerle uygulamakta olduğu, ciddi istatistiksel teknikler içeren yöntemleri küçük firmaların uygulaması pratikte çok zordur.

Tüm işyerlerinin iş güvenliği uzmanı ve işyeri hekimi çalıştırma yükümlülüğü kısmen bu problemi çözecektir. Danışman olarak görev yapacak bu kişilerin temel görevleri arasında risk değerlendirme çalışmalarının yürütülmesi yer almaktadır.

Diğer taraftan karmaşık yöntemlerle bu işin yapılmasına çalışılması küçük firmalar için zorluk üretmeye devam edecektir. Bunun için AB ülkelerinde kontrol listelerine dayalı ve basit matrislerle risk analizi gerçekleştirilen yöntemler önerilmektedir. Bu yöntemler son dönemde sektör bazlı hazırlanmakta, her sektörün ihtiyacına yönelik açıklamalar içermektedir. Türkiye'de İş Sağlığı ve Güvenliği Genel Müdürlüğü'nün önerdiği bu tür yöntemler kamuoyu ile paylaşılmaktadır. Ayrıntılı bir risk değerlendirmesi rehberi için [3]'e bakılabilir.

Ayrıca karmaşık risk değerlendirme matrisleri ve Fine-Kinney yöntemi [4] gibi yöntemlerin de Türkiye'de yaygın şekilde kullanılmaya başlandığını görmekteyiz.

Türkiye, risk değerlendirmesinin ülke çapında yaygınlaştırılması ve oturtulması sürecinin henüz başındadır. Bu çalışmanın amacı risk değerlendirmesinde mevzuat gerekleri, Türkiye'de yaygınlaşan risk değerlendirmesi yöntemleri ile risk analizinde risk hesaplaması için başvuru yöntemleri hakkında pratik bilgiler vermek ve bunların uygulama zorlukları ve uzun vadede ne ölçüde başarılı olacakları hakkında bir değerlendirme yapmaktır. Diğer bir deyişle yakın geleceğe yönelik bir projeksiyon ortaya koymaktır.

#### KAYNAKLAR

- [1] 6331 sayılı İş Sağlığı ve Güvenliği Kanunu, 20/6/2012 tarihli Resmi Gazete.
- [2] Risk Değerlendirmesi Yönetmeliği, 29/12/2012 tarihli Resmi Gazete.
- [3] Kobi'ler için İş Sağlığı Ve Güvenliği Yönetim Rehberi:Risk Değerlendirmesi,İSG Performans İzleme ve Sağlık Tehlikeleri,Metal Sektörü, Çalışma ve Sosyal Güvenlik Bakanlığı, İş Sağlığı Ve Güvenliği Genel Müdürlüğü, [http://www.isgum.gov.tr/rsm/file/isgdoc/isgip/isgip\\_metal\\_2\\_RA.pdf](http://www.isgum.gov.tr/rsm/file/isgdoc/isgip/isgip_metal_2_RA.pdf)
- [4] Özkılıç, Ö., İş Sağlığı, Güvenliği ve Çevresel Etki Risk Değerlendirmesi, MESS Yayınları, Yayın No:540, 2007.

#### ABSTRACT

##### AN OVERVIEW OF RISK ASSESSMENT METHODS AND PRACTICES IN TURKEY

Turkey is at the beginning of the dissemination of risk assessment process across the country. The aim of this study is to give practical information about the risk assessment methods which are growing up in Turkey, about risk analysis methods applied for the calculation of risk and their application challenges, and to make an assessment of the extent to which it will be successful. Thus it will try to make a projection for the near future.

**Key Words:** Occupational health and safety, Risk assesment methods.

#### BORSA ENDEKSLERİ ARASINDAKİ İLİŞKİNİN KANONİK KORELASYON ANALİZİ İLE İNCELENMESİ

Mehmet Cem ÇATALBAŞ  
Hacettepe Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği, 06800, Ankara, TÜRKİYE  
catalbas@ee.hacettepe.edu.tr

#### Giriş

Borsa endekleri, hisse senetleri arasındaki ilişkiler uzun zamandır istatistik bilimi tarafından incelenmektedir. Bu çalışmada Borsa İstanbul (BIST) için tanımlanmış bileşik endeksler olan BIST100, BIST50 ve BIST30 endekleri ve 22 sektörel endeks arasındaki ilişki kanonik korelasyon analizi ile incelenecektir. Bu ilişkiler ışığında sektörlerin bir birilerine olan etkileri ve bu etkilerin boyutlarının bulunması sağlanacaktır.

#### Endeksler ve İncelenmesi

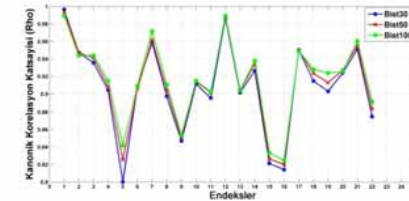
**Endeks genel olarak belirli bir zaman aralığında sektörel faaliyetlerdeki değişimin tanımlanması amacıyla oluşturulmuş verilerdir. Finansal olarak ise belirli özelliklerine göre ayrıştırılmış şirketlere ait hisse senetlerindeki değişimleri temsil etmektedir. Bu çalışmada BIST'a göre ayrıştırılmış hisse senetlerinden oluşan endekslerin çeşitli parametreler ışığında kanonik korelasyon analizi çıktıları irdelenecektir. Bu veriler 168 günlük bir zaman aralığında alınmıştır.**

#### Çizelge 1. Örnek bir endekse ait inceleme aşamasında kullanılan veriler.

Endeks adı	Dün	1.Seans	2.Seans	% Değişim	En Yüksek	En Düşük
XBANK	179.489	175.207	175.937	-1.98	180.438	173.946

#### a. Kanonik Korelasyon Analizi

Kanonik korelasyon analizi (KKA) 1936 yılında Harold Hotelling tarafından bulunmuştur. Genellikle iki bağımsız değişkene sahip veri kümesi arasındaki ilişkiyi incelemek amacıyla kullanılmaktadır. Bulunduğu günden bugüne oldukça fazla uygulama alanı bulan KKA yöntemi ayrıca gelişmiş istatistik analiz yöntemlerinden biri olarak kabul edilmektedir. KKA'de temel amaç veri kümesini oluşturan bilgilerin en yüksek korelasyona ait doğrusal kombinasyonlarının elde edilmesidir [1]. Kanonik korelasyon katsayısı en fazla  $\rho_{maks} = 1$ , en az ise  $\rho_{min} = 0$  değerini almaktadır.



Şekil 2. Sektörel Endeksler ve Bileşik Endeksler Arasındaki İlişki

Şekil 1 de bileşik endeksler olan BIST100, BIST50 ve BIST30 endeksleri ve 22 sektörel (XBANK, XBLSM, XELKT, XFINK, XGIDA, XGMYO, XHOLD v.b) endeks arasındaki en büyük kanonik korelasyon katsayıları(KKK) gösterilmektedir. Çizelge 1'de gösterildiği üzere endeksler arasındaki ilişki incelenirken 6 parametre üzerinden analiz yapılmaktadır.

#### Çizelge 2. Sektörel Endeksler ve Bileşik Endeksler Arasındaki KKK Değerleri

ENDEKS	En Büyük KKK ( $\rho$ )	ENDEKS	En Küçük KKK( $\rho$ )
XBANK	0.989	XSPOR	0.824
XKURY	0.988	XSGRT	0.831
XHOLD	0.971	XGIDA	0.841
XUHIZ	0.960	XILTM	0.853

Çizelge 2 incelendiğinde bileşik borsa endeksi (BIST100) ile en yüksek KKK sahip sektörel endeksin XBANK ve en düşük KKK sahip sektörün ise XSPOR olduğu bulunmaktadır. Sektörel endekslere ait genel veri kümesi ile BIST100 endeksinde ait verilerin KKA sonucunda elde edilen en yüksek kanonik korelasyon katsayısına sahip kanonik çiftlerin katsayıları denklem (1,2) de gösterildiği gibidir[2]. Bu denklemlerde hesaplanan kanonik yüklerden yola çıkılarak hangi sektörlerin bileşik endeks olan(BIST100) endeksinde artı ya da eksi yönde etki yaptığı görülebilmektedir[3].

$$U_1 = 0.616x_1 + 0.116x_2 - 3.069x_3 + 1.438x_4 + 2.201x_5 + 0.1943x_6 \quad (1)$$

$$V_1 = 0.247y_1, \dots, -5.112y_8, \dots, -4.949y_{67} + 0.754y_{68} + 4.192y_{69} + \dots, 3.278y_{79} + \dots, +0.053y_{132} \quad (2)$$

Yukarıdaki bilgiler ışığında BIST100 bileşik endeksinin sektörel endekse en fazla etki ettiği parametreler sırasıyla 2.seans değeri, en yüksek değeri, yüzde değişim, dünkü değeri, en düşük değeri, 1.seans değeri şeklindedir. Sektörel endeksin bileşik endekse en fazla etki ettiği parametreler ise sırasıyla XBLSM endeksinin 1.seans değeri, XKURY dünkü değeri, XKURY 2.seans değeri, XMESY dünkü değeri şeklindedir.

Çizelge 3. Sektörel Endeksler Arasındaki KKK

En Yüksek KKK			En Düşük KKK		
1.Endeks	2.Endeks	( $\rho$ )	1.Endeks	2.Endeks	( $\rho$ )
XBLSM	XTAST	0,9705	XGIDA	XGMYO	0,696
XKURY	XBANK	0,969	XGIDA	XKAGT	0,704
XHOLD	XKURY	0,966	XKMYA	XSGRT	0,723

Çizelge 3 de ise sektörel endekslerin birbirileri ile olan ilişkisi incelenmiştir. Bu incelemeler ışığında XBLSM ve XTAST endeksleri arasındaki ilişkinin oldukça yüksek olduğu aynı şekilde XKURY ve XBANK endeksleri arasındaki ilişkinin oldukça yüksek olduğu gözlemlenmiştir. XGIDA ve XGMYO endeksleri arasında ise en zayıf ilişkinin olduğu gözlemlenmiştir.

#### KAYNAKLAR

[1] Hotelling, H. (1936). "Relations Between Two Sets of Variates". Biometrika 28 (3-4)

[2] Multivariate Data Analysis, 5th edition by Joseph F. Hair, Jr., Rolph E.

Anderson, Ronald L. Tatham and William C. Black. Copyright © Prentice Hall, Inc. 1998.

[3]Koby Todros and Alfred O. Hero III, "Measure Transformed Canonical Correlation Analysis with Application to Financial Data" Signal Processing Workshop (SAM), 2012

**Anahtar Kelimeler:** Kanonik Korelasyon Analizi, Borsa, Veri Madenciliği, Veri Analizi

## PERIODOGRAM-BASED COINTEGRATION TEST: AN APPLICATION FOR REGIONAL INFLATION IN TURKEY

Serpil TÜRKYILMAZ<sup>1</sup>, Mesut BALİBEY<sup>2</sup>

<sup>12</sup>Bilecik Şeyh Edebalı University, Faculty of Science&Art, Department of Mathematics, Bilecik, Turkey  
serpil.turkyilmaz@bilecik.edu.tr, mesut.balibey@bilecik.edu.tr

### INTRODUCTION

In economics, inflation is described a rise in the general level of prices of goods and services in an economy over a period of time. Inflation-targeting regimes have been adopted to guide monetary policy decisions by many developing countries. Inflation targets are determined by the Central Bank of the Republic of Turkey and government in Turkey together. In practice, central banks use inflation forecasts as intermediate targets.

Inflation in Turkey is observed by the deflator of CPI(consumer price index) which is monthly prepared by the Turkish Statistical Institute, PPI(producer price index) and GDP(the gross domestic product) which is quarterly prepared. Furthermore, Turkey is administratively divided into 81 provinces and five regions identified in demographical studies. This classification includes West, South, Central, Northern and Eastern regions. In addition to the conventional five geographic regions, on 22 nd of september in 2002, a new classification in three different levels happened on behalf of the membership initiative of the european union by the state planning organisation and Turkish statistical agency in three different levels. In this study, the co-integration relationships between 26 statistical regions will be examined by three different method such as Periodogram, Engle-Granger Method and Johansen Method.

### METHODOLOGY

#### 2.1 Periodogram Based Unit Root

The periodogram based unit root test was introduced by Akdi and Dickey(1998). Periodograms can be used in many types of data. For an univariate time series data  $\{Y_t : t \in T\}$ , the periodogram ordinates can be calculated without any model specification as;

$$I_n(w_k) = \frac{n}{2} (a_k^2 + b_k^2), \quad (1)$$

where,

$$a_k = \frac{2}{n} \sum_{t=1}^n (Y_t - \bar{Y}) \cos(w_k t) \quad (2)$$

$$b_k = \frac{2}{n} \sum_{t=1}^n (Y_t - \bar{Y}) \sin(w_k t)$$

$w_k = 2\pi k / n$ .  $\mu$  is the mean of the series. One of the main use of the periodogram is to search whether there is a periodicity in the data or not. This can also be used to test for a unit root (Akdi, Dickey(1998)). The periodogram has many advantages comparing to the time domain approach. The periodogram can be calculated without any model specification and the critical values of the test statistics are free of sample size constraints.

#### 2.2. Periodogram Based Cointegration Tests

When a bivariate series  $Y_t$  is given, each component of the series can be considered as a sum of a stationary and a nonstationary series:

$$\begin{aligned} Y_{1,t} &= a_{11}U_t + a_{12}S_t \\ Y_{2,t} &= a_{21}U_t + a_{22}S_t. \end{aligned} \quad (3)$$

Where  $U_t$  and  $S_t$  represent nonstationary series (unit root series) and stationary series, respectively. From this representation, it can be seen that is:

$$Y_{2,t} - \frac{a_{21}}{a_{11}} Y_{1,t} = (a_{22} - \frac{a_{21}a_{12}}{a_{11}}) S_t = c S_t \quad (4)$$

The Linear combination of two series yields a stationary series with the cointegrating vector  $\underline{a} = (-a_{21}/a_{11}, 1)'$ . The periodogram-based cointegration test developed by Akdi (1995) depends on the regression of the real part of the cross-periodogram ordinates of two series.

$$y_{1,k} = \alpha + \beta y_{2,k} + \eta_k, \quad k=1,2,\dots,[n/2]. \quad (5)$$

Where,  $\hat{\beta}$ , estimate of  $\beta$ , is a consistent estimator of  $(a_{21}/a_{11})$ . In order to determine whether  $Y_{1,t}$  and  $Y_{2,t}$  are cointegrated or not.  $Z_t$  is derived by using the estimated coefficient from the regression of cross-periodogram ordinates (5) as follows:

$$Z_t = Y_t - \hat{\beta} X_t \quad (6)$$

In addition,  $\Delta Z_t$  on  $Z_{t-1}$  is regressed and, usual t-statistics are applied.

## CONCLUSION

In this study, we are going to consider bivariate series and to search whether there are cointegrated or not by using a periodogram based method. In other words, the study examines whether there is any long-term relationship among CPI according to 26 statistical regions by using three different cointegration methods such as Periodogram, Engle-Granger Method and Johansen Method. The results of empirical analysis indicate that there exists no bivariate cointegration relationship among some of CPI series when we use Engle-Granger Regression Method and Periodogram based method. However, the Johansen's Trace and Maximum Eigenvalue Methods does not display any co-integration relationships among these 26 statistical regions. Periodogram-based test has many advantages over conventional tests; this test is model-free, seasonally robust and mean invariant. Furthermore, periodogram method is able to use to support one of the conventional cointegration tests that show results of different co-integration relationships.

## REFERENCES

- [1] Akdi, Y. (1995), Periodogram analysis for unit roots, Ph.D. diss., NCSU.
- [2] Akdi, Y. and D.A. Dickey (1998), Periodograms of unit root time series: Distributions and tests, Communications in Statistics : Theory and Methods, 27(1), 69-87.
- [3] Akdi, Y., H. Berument and S.Y. Cilasun (2006), The relationship between different price indices: Evidence from Turkey, Physica A, Statistical Mechanics and its Applications, 360, 483-492.
- [4] Berument, H., Y. Akdi and C. Atakan (2005), An empirical analysis of Istanbul Stock

Exchange sub-indexes, Studies in Nonlinear Dynamics and Econometrics, 9(3), 1-12.

## ABSTRACT

### AN APPLICATION OF PERIODOGRAM-BASED COINTEGRATION TEST FOR INFLATION ACCORDING TO REGIONS

This paper examines whether there is any long-term relationship among CPI (Consumer Price Index) according to 26 statistical regions by using three different cointegration methods such as Periodogram, Engle-Granger Regression Method and Johansen's Method. Data set covers the period from March 2003 to October 2012 for Turkey. In particular, there exists no bivariate cointegration relationship among some of CPI series when we use the Engle-Granger Regression Method and the periodogram based method. Furthermore, we were unable to obtain any cointegration relationships among these index of 26 statistical region when the Johansen's Trace and Maximum Eigenvalue methods were used. The periodogram-based method has a further advantages over conventional tests. In addition, it is able to use to support one of the conventional cointegration test that show results of different co-integration relationships.

**Key Words:**, Time Series Analysis, Cointegration, Periodogram, Regional Inflation

## KLASİK VE SAĞLAM DEĞİŞKEN SEÇİMİ ÜZERİNE BİR ÇALIŞMA

Onur TOKA<sup>1(\*)</sup>, Meral ÇETİN<sup>2</sup>

<sup>1,2</sup>Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800, Ankara, TÜRKİYE

onur.toka@hacettepe.edu.tr, meral@hacettepe.edu.tr

## GİRİŞ

Çoklu doğrusal regresyon çözümlemesinde verileri en iyi biçimde tanımlayacak gerekli değişkenlerin modele seçilmesi, modele katkısı önemsiz değişkenlerin çıkartılması 'değişken seçimi' ya da 'en iyi altküme seçimi' olarak bilinir. Çoklu doğrusal regresyonda sürekli gündeme gelen sorun, bağımsız değişkenlerin bir altkümесinin seçimidir. Değişken seçimi ile ilgili ölçütlerin çoğu kestirim artkareler ortalamasının basit bir fonksiyonudur.

Klasik model seçim yöntemleri klasik kestiricilerle ve testlere dayalıdır. Örneğin en çok kullanılan Mallows'un Cp ölçütü ve Akeike bilgi ölçütü (AIC) en küçük kareler kestirimine dayalı olduğundan bu ölçütler aykırı değerlere ve hata dağılımının normallik varsayımından sapmalara karşı oldukça duyarlıdır. Bu durumda, model seçim yöntemlerinin sağlam versiyonlarına gerek duyulur. Son yıllarda yapılan araştırmalarda regresyon modeli için test yöntemleri ve sağlam kestiriciler önerilirken, sağlam model seçim yöntemleri ihmal edilmiştir. Bu noktada önerilmiş olan çeşitli yöntemler, benzetim çalışmasıyla klasik model seçim yöntemleriyle karşılaştırılarak yorumlanmıştır.

## SAĞLAM DEĞİŞKEN SEÇİM YÖNTEMLERİ

Regresyon analizinde farklı modeller arasında seçim yapabilmek için kullanılan en güçlü yöntemler AIC ve Mallows'un Cp seçim ölçütleridir. Regresyon analizinde AIC yönteminin hata dağılımlarında normallik varsayımı gerektirmesi ve Mallows'un Cp seçim yönteminde en küçük kareler kestiricisi üzerinden çözümleme yapması iki yöntemde aykırı değerlere karşı duyarlı olmasına sebep olmaktadır.

Sağlam değişken seçimi ile ilgili çalışmalar oldukça azdır. Yapılan çalışmalar daha çok klasik seçim ölçütlerini sağlamlaştırılmasına dayalıdır: Ronchetti (1985), AIC'in sağlam versiyonunu önermiştir. Ronchetti ve Staudte (1994), Mallow'un Cp ölçütünün sağlam bir versiyonunu önerirken, Sommer ve Huggins (1996) Wald testine dayalı kolay genelleştirilebilen bir yöntem önermişlerdir.

Bu çalışmada, Sağlam Cp değişken seçim ölçütü, Sağlam AIC ölçütü ve Sağlam Tp ölçütleri M-kestiricileri kullanılarak hesaplanmış ve klasik değişken seçim ölçütleriyle karşılaştırılmıştır.

## BENZETİM ÇALIŞMASI

Bu çalışmada yukarıda bahsedilen sağlam ve klasik değişken seçim yöntemleri karşılaştırılmıştır. Klasik ve sağlam değişken seçim ölçütlerini karşılaştırabilmek için benzetim çalışması yapılmıştır. S-plus programı kullanılarak bir kodlar hazırlanmıştır. Aykırı değer ve varyans kestiriciler ve değişken seçim ölçütleri üzerindeki etkisini görebilmek amacıyla veri kümesinde farklı aykırı değerler oluşturulmuş ve farklı varyans değerleri ele alınmıştır. Bu etkenler üzerinden elde edilen değişken seçim ölçütlerinin gerçek modeli ve alternatif modelleri seçme yüzdeleri elde edilmiş ve sonuçlar yorumlanmıştır.

## KAYNAKLAR

[1] Çetin, M.(2000),Sağlam Regresyonda Değişken Seçim Kriterleri, Doktora Tezi, Hacettepe Üniversitesi.

[2] Hampel,F.R.,Ronchetti, E. M.,Rousseeuw, P. J., and Stahel, W. A.,1(1986), Robust Statistics:The Approach Based on Influence Functions, New York, John Wiley,.



[3] Ronchetti, E.(1985), Robust model selection in regression, Statistics and Probab. Letters., 3, 21-23.

[4] Ronchetti, E. and Staudte, R.(1994), A Robust version of Mallows's Cp, JASA, Vol. 89, 550-559.

[5] Sommer, S. and Huggins, R. M.,(1996), Variable selection using the Wald test and a Robust Cp, Appl. Statist., 45, 15-29.

## ABSTRACT

### COMPARISON OF CLASSICAL AND ROBUST VARIABLE SELECTION

In linear regression analysis, outliers often have large influence in the model/variable selection process. The aim of this study is to select the subsets of independent variables, which explain dependent variables in the presence of outliers and possible departures from the normality assumption of the error distribution in robust regression analysis. There is a simulation study to compare classic and robust model selection criteria for different variances of error distribution and different number of outlier.

**Key Words:** Model Selection, Robust Model Selection Criteria, Outlier.

## PARÇACIK SÜRÜ OPTİMİZASYONU (PSO) İLE AÇIK OCAK MADEN REZERV ALANININ BELİRLENMESİ

Orhan KESEMEN\*, Ülkü ÜNSAL

Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, Trabzon, TÜRKİYE  
oke semen@gmail.com, ulkunsal@gmail.com

### Giriş

Yeraltında ve/veya bulunan cevherin toprağın veya kayanın üst kısmı kaldırılarak çıkarılması için yapılan çalışmaların tümü açık işletmecilik olarak tanımlanabilir [1]. Maden aramalarında ilgilenilen bir alanın istenilen noktalarında sondaj çalışmaları yapılarak bu noktalardaki rezerv miktarları belirlenmektedir. Bu miktarlardan yararlanarak çıkarılması planlanan açık maden sahasının ekonomikliği incelenir. Açık maden işletmeciliğinde, en az maliyet ile en fazla rezerve ulaşabilmek için iyi planlama ve iyi hesaplama gerekmektedir.

Bu çalışmanın ikinci bölümünde açık maden işletmeciliğinde kullanılan yöntemler belirtilmiştir. Üçüncü bölümde kullanılan optimizasyon yöntemi hakkında bilgi verilmektedir. Dördüncü bölümde ise optimizasyon yönteminin açık maden işletmesinde uygulanması gösterilmiştir.

### Açık İşletmelerde Kullanılan Rezerv Hesaplama Yöntemleri

Açık ocak maden işletmeciliğinde mevcut rezervin hesaplanması için sondaj noktalarından yararlanılarak bazı yöntemler geliştirilmiştir. Bu yöntemlerden en çok kullanılanlar aşağıda açıklanmıştır [2]:

- **Düzenli Bloklar Yöntemi:** Yüzey, her sondaj noktası için bir bloğun köşegenlerinin kesişim yerine gelecek şekilde kare ya da dikdörtgen bloklara ayrılır. Bloğun hacmi, yüzey alanı ile sondajda kesilen kalınlığın çarpımı ile bulunur.
- **Kenar Orta Dikme Yöntemi:** Sondaj noktaları birleştirilerek dar açılı üçgenler oluşturulur. Üçgenlerin kenarortayları çizilerek poligonlar oluşturulur.
- **Açı Ortay Yöntemi:** Sondaj noktaları birleştirilerek oluşturulan dar açılı üçgenlerin açıortayları birleştirilerek poligonlar elde edilir.
- **Üçgen Yöntemi:** Sondaj noktaları köşeleri oluşturacak şekilde eşkenar üçgenler oluşturulur. Tüm üçgenlerin alanı bulunarak damar kalınlığı, yoğunluğu ve jeolojik faktörü ile çarpılarak rezerv miktarı hesaplanır.
- **İzopak Yöntemi:** Cevher alanı aynı düzeydeki kısımların işaretlenmesiyle bulunan eşdeğer eğrilerin arasında kalan hacimlerden hesaplanır.

### Parçacık Sürü Optimizasyonu (PSO)

Parçacık Sürü Optimizasyonu (PSO), 1995 yılında, Kennedy ve R.C. Eberhart tarafından, kuş sürülerinin davranışlarından esinlenerek geliştirilmiş bir optimizasyon yöntemidir [3]. PSO, probleme farklı çözümler getiren parçacık adı verilen elemanlardan oluşur. Parçacıkların oluşturduğu topluluğa sürü adı verilir [4]. Parçacık sürü optimizasyon yönteminin uygulanma adımları Algoritma 1'de verilmiştir.

### Algoritma1: Parçacık Sürü Optimizasyonu

**For** her parçacık için

Parçacığı başlangıç konumuna getir

**End**

**Do**

**For** her parçacık için

Uygunluk değerini hesapla

If uygunluk değeri pbest ten daha iyi ise,

Şimdiki değeri yeni pbest olarak ayarla

**End**

Tüm parçacıkların bulunduğu pbest değerlerinin en iyisini, tüm parçacıkların gbest'i olarak ayarla

**For** her parçacık için

Parçacık hızını güncelle

Parçacık konumunu güncelle

**End**

**While** maksimum iterasyon sayısına veya minimum hata koşulu sağlanana kadar devam et

### PSO'NUN AÇIK MADEN OCAKLARINA UYGULANMASI

Maden alanı yüzeyde bir çokgen çerçeve ile tanımlanarak belirlenebilmektedir. Bu çokgen, en yüksek rezervi ve en az alan verecek bir alanın sınırlarını vermektedir. Bu alan belirlenirken N-boyutlu PSO tekniği kullanılmıştır. Çokgende amaç değeri hesaplanırken en küçük alan içerisinde kar marjını koruyacak şekilde en yüksek maden rezervi verecek şekilde tasarlanmıştır. Burada maden rezerv değişimi bir olasılık yoğunluk fonksiyonuna benzetilirken, rezerv alanın sınırları ise iki boyutta güven aralığının sınırlarını vermektedir.

### KAYNAKLAR

[1] Yrd. Doç. Dr. Bülent HANER, Zonguldak Karaelmas Üniversitesi, Meslek Yüksekokulu, Maden İşletme II Ders Notları

[2] web.deu.edu.tr/maden/docs/maden.../rezerv%20dersnotu-belge.pdf. Erişim Tarihi: 28 Mayıs 2013

[3] KENNEDY, J. EBERHART, R.C., Particle Swarm Optimization, Proc. of IEEE International Conference on Neural Network, Piscataway, NJ. Pp..1942-1948 , 1995

[4] ORTAĞI, Y., GÖLOĞLU, C. Parçacık Sürü Optimizasyonu ile Küme Sayısının Belirlenmesi, Karabük Üniversitesi, Bilgisayar Mühendisliği, XIV. Akademik Bilişim Konferansı, 2012

[5] http://www.yazilimutfagi.com/10058/yapay-zeka/heuristic/parcacik-suru-optimizasyonu-particle-swarm-optimization-.aspx. Erişim Tarihi: 7 Haziran 2013

### ABSTRACT

#### DETERMINATION MINE RESERVE AREA IN OPEN PIT WITH PARTICLE SWARM OPTIMIZATION (PSO)

Nowadays optimization methods used and successful results have been achieved in many areas. In this study, PSO method, open-cast mining operations brings a new approach to the methods used in the removal of the existing reserves. Used method, the maximum reserve amount targeted to be reached with the least cost.

**Key Words:** PSO, Reserve Calculation, Open Pit Mining.

### KÜME SAYISININ OPTİMAL OLARAK BELİRLENMESİNDE YENİ BİR YAKLAŞIM

Orhan KESEMEN\*, Özge TEZEL

Karadeniz Teknik Üniversitesi, Fen Fakültesi,  
İstatistik ve Bilgisayar Bilimleri Bölümü,61080, Trabzon, TÜRKİYE  
okesemem@gmail.com, ozge\_tzl@hotmail.com

#### Giriş

Kümeleme analizi çok değişkenli veri analiz yöntemlerinden biridir. Kümeleme analizi, bir araştırmada incelenen birimleri aralarındaki benzerliklerine ve farklılıklarına göre belirli gruplar içinde toplayarak sınıflandırma yapmayı, birimlerin ortak özelliklerini ortaya koymayı ve bu sınıflar ile ilgili genel tanımlamalar yapmayı sağlayan bir yöntemdir [1]. Burada amaç; gruplanmamış verileri benzerliklerine göre sınıflandırmak ve araştırmacıya uygun, işe yarar özetleyici bilgiler elde etmede yardımcı olmaktır. Bazı kümeleme problemlerinde ise, kümeler birbirinden belirgin bir şekilde ayrılmıyorsa ya da bazı birimler küme üyeliğinde kararsızsa, klasik kümeleme yöntemleri yerine bulanık kümeleme yöntemleri tercih edilmektedir.

Bulanık kümeleme analizinde en iyi küme sayısının belirlenmesi önemli bir problemdir. Bununla birlikte birçok kümeleme algoritması küme sayısının önceden bilinmesini gerektirir. Birçok çalışmada, araştırmacının küme sayısı hakkında ön bilgisinin olmaması, bulunan küme sayısının gerçek küme sayısından az ya da çok olup olmadığının bilinmemesine yol açmaktadır. Eğer bulunan küme sayısı gerçek küme sayısından az çıkarsa, kümelerden bir veya birkaçı birleşmek durumunda olacaktır, çok çıkarsa kümelerden bir veya birkaçı bölünmelere uğrayacaktır. Optimal küme sayısının belirlenme işlemlerine genel olarak Küme Geçerliliği adı verilmektedir [2].

Veriler bir ve iki boyutlu uzayda olduğu zaman verileri görsel olarak yorumlayarak küme sayısına karar verilebilmektedir. Ama uzaydaki boyut sayısı arttıkça görsellik zorlaşmakta ve geçerlilik indekslerine ihtiyaç duyulmaktadır. Bunun için en iyi küme sayısının belirlenmesi için çeşitli küme geçerlilik indeksleri önerilmiştir.

#### Yöntemler

##### • Bölünme Katsayısı (PC) ve Bölünme Entropisi (PE)

Bezdek tarafından önerilen bu küme geçerlilik indekslerinin dezavantajı sadece üyelik derecelerini kullanması ve kümelerin veri yapılarını göz önünde bulundurmamasıdır. [3]

##### • Xie-Beni İndeksi(XB)

Xie ve Beni tarafından (1991) geliştirilen bu indeks, yoğunluk ve ayrılma geçerlilik fonksiyonu olarak da bilinir. Bu küme geçerliliği indeksi hesaplanırken üyelik değerlerinden, verilerin küme merkezlerine olan uzaklıklarından faydalanılmıştır.

##### • Fukuyama and Sugeno İndeksi(FS)

Fukuyama ve Sugeno tarafından geliştirilen bu indekste de yoğunluk ve ayrılma kavramlarından faydalanılmıştır.

##### • K İndeksi

Kwon tarafından önerilen bu indekste Xie ve Beni indeksindeki küme sayısının veri sayısına yakın olması eğilimi azaltılmaya çalışılmıştır.

##### • CWB İndeksi

Rezaee tarafından önerilen bu indekste küme merkezleri arasındaki minimum ve maksimum uzaklıktan, kümelerin varyanslarından faydalanılmıştır.

##### • B<sub>crit</sub> İndeksi

Boudraa varyans odaklı bir geçerlilik indeksi önermiştir.

##### • SV İndeksi

Kim iki kümenin birbirinden ne kadar ayrıldığını ölçmek için, hem iki küme arasındaki alt ayrılma indeksini hem de üst ayrılma indeksini kullanan bir küme geçerlilik indeksi önermiştir.

#### Önerilen Yöntem

Şimdiye kadar önerilen yöntemler üyelik değerlerini kullanarak küme içi benzerlikleri temel alan yöntemlerdir. Bu çalışmada ise sadece küme içi benzerlik değil küme dışı farklılıklarda dikkate alınmıştır. Bir elemanın bir kümeye ait olup olmaması üyelik değeri ile belirlenmektedir. Bir elemanın küme merkezine yaklaşması üyelik değerinin bire uzaklaşması ise üyelik değerinin sıfıra yaklaşmasına neden olmaktadır. Bir elemanın tüm kümelere üyelik değerinin toplamı bire eşit olduğundan, elemanın ait olduğu kümeye üyelik değeri (a) olarak alınırsa, üye olmama değeri ise (1 - a) olacaktır. Bu iki değer oranlanmasıyla elde edilen formülasyon optimal küme sayısının belirlenmesinde etkili bir yöntem olmaktadır.

#### KAYNAKLAR

- [1] Güler, N., Yüksek lisans Tezi, Bulanık Kümeleme Analizi ve Bulanık Modellemeye Uygulamaları, 2006
- [2] Atatürk Ü. İİBF Dergisi, 10. Ekonometri ve İstatistik Sempozyumu Özel Sayısı, 2011 (ALPASLAN, ERİLLİ, YOLCU, EĞRİOĞLU, & ALADAĞ, 2011)
- [3] Kim, Dae-Won, Lee, Kwang H., & Lee, Doheon On cluster validity index for estimation of the optimal number of fuzzy clusters, 2004

#### ABSTRACT

##### A NEW APPROACH TO THE DETERMINATION OF THE OPTIMAL NUMBER OF CLUSTERS

Deciding the optimal number of clusters is an important problem for fuzzy clustering analysis. In fuzzy clustering, each point has a degree of belonging to clusters and it is called membership value. In this study, we use membership value and nonmembership value. The main purpose of this study, determine the optimal number of cluster for separate data points which use similarities or dissimilarities between data points with using new method and find the best validity index for all data sets.

#### DEĞİŞTİRİLMİŞ YAPAY ARI KOLONİSİ ALGORİTMASI KULLANILARAK YOL GÜZERGAHI ÜZERİNDE DAĞITIM MERKEZLERİNİN YERLERİNİN BELİRLENMESİ

Orhan KESEMEN\*, Yeşim YEGİNOĞLU, Eda ÖZKUL

Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, Trabzon, TÜRKİYE, okesemen@gmail.com, yesimyeginoglu@gmail.com, eda.ozkul.gs@gmail.com

#### Giriş

Tedarik zinciri yönetimi stratejisi sürekli artan rekabet ortamında şirketlerin faaliyetlerini sürdürebilmelerinde çok önemli bir yer tutmaktadır. Taşıma maliyetleri şirketlerin giderlerinin büyük bir kısmını oluşturmaktadır. İşletmelerin ürünlerini geniş pazarlara kısa zamanda hızlı teslimat ve düşük maliyetler ile ulaştırabilmeleri için öncelikle yapmaları gereken dağıtım merkezleri için uygun yerler belirlemektir. Dağıtım merkezlerinin yerlerinin belirlenmesi, işletmelerin karşılaştığı en büyük sorunlardan biridir. Dağıtım merkezlerinin yerlerinin belirlenmesinde pazar araştırmalarının ve talep kestirimlerinin yapılması son derece önemlidir.

İşletmeler, ulaştırma sisteminin yetersiz olması ve taşıma maliyetlerinin fazla olması gibi nedenlerden dolayı dağıtım merkezlerini pazarlara yakın yerlere kurmak zorunda kalabilir.

Bu çalışmada belli bir bölgede dağıtım merkezleri için yerler belirlenecektir. Dağıtım merkezlerinin şehir merkezlerinde mi yoksa şehirlerarası yollar üzerinde mi kurulacağı da büyük bir problemdir. Şehir merkezinde kuruluş maliyeti fazla olmasına rağmen şehir dışında kurulduğunda da taşıma maliyeti fazla olacaktır. Fakat şehirlerarası yolların üzerinde dağıtım merkezlerinin kurulması tek bir dağıtım merkezinden birden çok pazara dağıtım yapmasına olanak sağlamaktadır.

#### Yöntem

Amaç fonksiyonu belirlenirken bazı şehirlerin daha kalabalık olduğu dolayısıyla pazar payının daha yüksek yani talebin daha fazla olacağı göz önünde bulundurularak dağıtım merkezlerinin bu şehirlere daha yakın olması gerektiği göz önünde bulundurulmalıdır.

Şehirlerarasındaki yollar doğrularla belirlenmektedir. Ürünler ilk aşamada tek bir üretim merkezinden dağıtım merkezlerine gönderileceği için dağıtım merkezleri arasında bağlar bulunmaktadır. Dağıtım merkezleri arasında bağların bulunması da bir karayolu üzerinde birden fazla dağıtım merkezi kurulmaması kısıtını meydana getirmektedir. Yollar iki boyutlu uzayda ayrı ayrı doğruların birleşimi biçiminde tanımlanmaktadır. Dolayısıyla mevcut optimizasyon yöntemleri alan içinde arama yaparken bu çalışmada doğrular üzerinde bir kısıtlamaya gidildiğinden farklı bir yaklaşım gerekmiştir. Bunun için yapay arı koloni algoritması mevcut probleme göre uyarlanarak geliştirilmiştir.

Yapay arı kolonisi (ABC) algoritması, Karaboğa tarafından arıların doğadaki davranışları temel alınarak geliştirilen popülasyon tabanlı bir optimizasyon tekniğidir [1].

Verilen karayolunda her doğru bir arama alanı olarak düşünülmektedir. Bir bölgede K tane dağıtım noktası seçilecekse bunların K tane ayrı doğru üzerinde olması gerekmektedir. Aynı doğru üzerinde olan dağıtım noktaları maliyet açısından anlamsız olmaktadır. K tane dağıtım noktası farklı K tane doğru üzerine yerleştirilmesiyle bir tane başlangıç çözümü elde edilmektedir. Bu çözümlerden N tane oluşturularak başlangıç popülasyonu elde edilmektedir. Her çevrimde, her çözümün bir dağıtım noktası doğru üzerinde ileri veya geri yönde rastgele miktarda hareket ederek çözüm iyileştirilmeye çalışılır. Eğer iyileştireme işlemi limiti aşarsa yeni bir başlangıç noktası seçilir. Tüm çevrimlerin sonucunda elde edilen en iyi genel çözüm problemin çözümü olarak kabul edilmektedir.

Oluşturulan her çözümün şehirlerarasındaki ulaşım ve taşıma maliyetlerinin belirlenmesinde Dijkstra algoritması kullanılmaktadır.

## KAYNAKLAR

[1] Karaboğa, D. ve Akay, B., Artificial Bee Colony Algorithm on Training Artificial Neural Networks, Signal Processing and Communications Applications, 2007. SIU 2007, IEEE 15th. 11-13 June 2007, 1 – 4.

## ABSTRACT

### DETERMINATION OF THE LOCATIONS OF DISTRIBUTION CENTERS ON THE ROAD USING MODIFIED ARTIFICIAL BEE COLONY ALGORITHM

Nowadays, a lot of organizations use distribution centers in order to distribute their products easily. And it makes determine the location of distribution centers is extremely important. Although general optimizations methods work on areas, this study's aim is determine the highways as line segments and specify optimum number of distribution center using modified ABC algorithm.

**Key Words:** Distribution center, Artificial bee colony algorithm.

## THE PREDICTION OF TURKISH LOCAL DERBY MATCHES

\*Ömer Ozan EVKAYA<sup>1</sup>, Sezgin ÇİFTÇİ<sup>2</sup>

1 Atılım University, Faculty of Arts & Sciences, Department of Mathematics, 06836, Ankara, Turkey, ozanevkaya@gmail.com

2. Baskent University, Faculty of Commercial Sciences, Department of Insurance and Risk Management, 06810, Ankara, Turkey, sezginciftci@yahoo.com

### Introduction

The usage of statistical methods to the soccer match data has gained more attention after the betting market became important for the economy in recent years. In this study, the Turkish derby match results predicted based on new generated explanatory variables using accessible 10 year data on the net. Two different regression models are considered for the statistical analysis. Firstly, the multiple linear regression (MLR) is used to interpret all derby matches in one single linear equation. The impacts of booking cards are also considered and MC simulation is used to estimate the number of cards for the future match. Alternatively, the logit/probit regression models are utilized to analyze each derby match separately. The set of distinct predictors is considered and the best model selection is carried out according to the area under the Receiver Operating Characteristic (ROC) curve.

### Methodology and Findings

In order to predict the derby match result, the available past data is collected firstly. The combination of overall performance of each team and head-to-head statistics are used to generate meaningful explanatory variables. Afterwards, the set of new predictors is considered to estimate the most likely winner of any derby match. One of the classical approaches is the linear regression to describe the result of a match based on estimators. For this reason, Multiple Linear Regression (MLR) method is used for simplicity. Moreover, the unpredictable number of booking cards is simulated by Monte Carlo (MC) technique by assigning prior distribution. After the selection of best MLR model for the home/away scoring ratio which represents the winner of the match, the MC simulation is conducted to derive winning probability for each team in different derby matches. The Table 1 summarizes the simulation result in a percentage.

Apart from MLR, the any derby Match Result (MR) is simplified as a binary outcome as follows;

$$MR = \begin{cases} 0, & \text{if home team did not lose any points (win)} \\ 1, & \text{if home team lost any points (draw or loss)} \end{cases} \quad (1)$$

Based on the above dichotomous MR value, the logit and probit regression models are used using different predictor variables. Under this modelling, each derby match is considered separately and the most significant logit/probit model is decided by using the area under the ROC curve. The results of probit model are significant for each derby and the predictor names are changed from match to match. Moreover, some derby matches are not modeled by logit/probit regression approach. The Table 2 shows us the model fit results.

**Table1. Simulated Probability Summary**

Derby Match		Probabilities (%)		
Home Team	Away Team	Win	Draw	Loss
Galatasaray(GS)	Fenerbahçe(FB)	87.7	12.2	0.1
Galatasaray(GS)	Besiktas(BJK)	94.8	5	0.2
Galatasaray(GS)	Trabzonspor(TS)	88	11.9	0.1

Fenerbahçe(FB)	Galatasaray(GS)	99.8	0.2	0
Fenerbahçe(FB)	Besiktaş(BJK)	91.2	8.8	0
Fenerbahçe(FB)	Trabzonspor(TS)	99.9	0.1	0
Besiktaş(BJK)	Galatasaray(GS)	54.8	44.3	0.9
Besiktaş(BJK)	Fenerbahçe(FB)	39	59.4	1.6
Besiktaş(BJK)	Trabzonspor(TS)	37.6	59.4	3
Trabzonspor(TS)	Galatasaray(GS)	16.4	74.6	9
Trabzonspor(TS)	Fenerbahçe(FB)	44	54.8	1.2
Trabzonspor(TS)	Besiktaş(BJK)	43.9	53.9	2.2

Table2. Logit/Probit Model Fit Results

Derby Match	Predictor Name	Coeff. Significancy	Area under ROC curve
GS - FB	$HT_{SSR}$	0.145	0.9667
GS - BJK	$HDA_{SPC}, HDA_{BCP}$	0.181 , 0.180	0.9583
GS - TS	$AT_{SSR}, A_{BCP}$	0.329 , 0.617	0.9643
FB - GS	$AT_{SSR}, A_{BCP}$	0.234 , 0.196	0.9444
FB - BJK	$HT_{SSR}, A_{BCP}$	0.214 , 0.218	0.8214
BJK - FB	$HT_{SSR}, HDA_{BCP}$	0.374 , 0.395	0.9167
TS - GS	$HDA_{SSR}$	0.186	0.8889
TS - FB	$AT_{SPR}, A_{BCP}$	0.348 , 0.108	0.9583
TS - BJK	$H/A_{SSR}, H_{BCP}$	0.596 , 0.517	0.9643

### Conclusions and Discussions

Because of the limited data set, in two proposed models, the results are not satisfactory. In MLR approach, the coefficients of predictors is not significant as it is expected. Moreover, even if the area under ROC curve seems to be so meaningful, the significancy of the coefficients of predictor variables still doubtful for some derby matches. Despite these shortcomings, the home team always defeats its opponent with high probabilities as given in Table 1. Especially, Fenerbahçe seems to be most powerful team among the others. Moreover, in logit/probit model,  $A_{BCP}$  has impact on the result of a match in most of the time. In general, it sounds logical, the away team affected negatively when the aggressiveness of it exceeds some certain level. Besides,  $HT_{SSR}$  and  $AT_{SSR}$  affects the winning capability of each team.

### REFERENCES

- [1] Lahvicka J. (2012), Using Monte Carlo Simulation to Calculate Match Importance: The case of English Premier League.
- [2] Aldrich J. H. and Nelson F. D. (1984), Linear Probability, Logit and Probit Models, Quantitative Applications of Social Science.
- [3] Cleves M. A. (2002), Comparing Areas under Receiver Operating Characteristic Curves from two or more Probit or Logit Models, The Stata Journal.
- [4] Website : <http://www.football-data.co.uk/turkeym.php>, November 2012.
- [5] Website : <http://www.transfermarkt.com.tr>, November 2012.

**Key Words:** derby match, MLR, MC simulation, logit/probit regression, ROC curve

A REVIEW ON ESTIMATION DISTRIBUTION ALGORITHMS IN DISCRETE AND  
CONTINUOUS OPTIMIZATION PROBLEMS

Serpil GÜMÜŞTEKİN<sup>1</sup>, Talat ŞENEL<sup>1</sup>

<sup>1</sup>Fen Edebiyat Fakültesi, Ondokuz Mayıs Üniversitesi, Samsun, Türkiye, e-mail: serpil.gumustekin@omu.edu.tr  
<sup>1</sup>Fen Edebiyat Fakültesi, Ondokuz Mayıs Üniversitesi, Samsun, Türkiye, e-mail: tlsenel@omu.edu.tr

Estimation Distribution Algorithms (EDAs) is a quite recent topic in optimisation techniques. They are a set of algorithms that belong to the field of Evolutionary Computation and they have been applied to a wide set of academic and real-world optimization problems. In EDAs there is neither crossover nor mutation operator. New population is generated by sampling the probability distribution, which is estimated from a database containing selected individuals of previous generation. Different approaches have been proposed for the estimation of probability distribution. In this paper we provide a review of different EDA approaches and present solutions based on discrete and continuous optimization in binary and non-binary search spaces. Problems are solved and the experimental results comparing every algorithms.

**Anahtar Kelimeler:** Estimation Distribution Algorithms, Optimization, Without Dependencies, Bivariate Dependencies, Multivariate Dependencies

- [1] Armañanzas, R., Inza, I., Santana, R., Saegys, Y., Flores, J. L., Lozano, J. A., Van de Peer, Y., Blanco, R., Robles, V., Bielza, C., and Larrañaga, P. (2008). A review of estimation of distribution algorithms in bioinformatics. *BioData Mining*, 1(6):1–12.
- [2] Baluja, S. (2006). Incorporating a priori knowledge in probabilistic-model based optimization. In Pelikan, M., Sastry, K., and Cantú-Paz, E., editors, *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications*, Studies in Computational Intelligence, pages 205–222. Springer-Verlag.
- [3] Bengoetxea, E., Miquélez, T., Larrañaga, P., and Lozano, J. A. (2002). Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation, chapter Experimental results in function optimization with EDAs in continuous domain, pages 177–190. Kluwer Academic Publishers, Boston/Dordrecht/London.
- [4] Brownlee, A., McCall, J., Zhang, Q., and Brown, D. (2008). Approaches to selection and their effect on fitness modelling in an estimation of distribution algorithm. In *Proceedings of the 2008 Congress on Evolutionary Computation (CEC-2008)*, pages 2621–2628, Hong Kong. IEEE Press.
- [5] Chow, C. K. and Liu, C. N. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14(3):462–467.

ÇOK BOYUTLU ÇAPRAZ TABLOLARIN ANALİZİNDE 1. TİP HATA YAPMA OLASILIĞI  
BAKIMINDAN PEARSON VE LOG-LIKELIHOOD RATIO Kİ-KARE İSTATİSTİKLERİNİN  
KARŞILAŞTIRILMASI

Şengül CANGÜR<sup>\*</sup>, Handan ANKARALI  
Düzce Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim A.D., 81620, Düzce, Türkiye  
E-mail: sengulcangur@duzce.edu.tr

Pearson ki-kare ve Log-Likelihood ratio (LR) test istatistikleri, çapraz tabloların satır ve sütunlarına yerleştirilen kategorik yapıdaki değişkenler arasındaki ilişkilerin analizinde sıklıkla kullanılmaktadır. Ancak bu iki test istatistiğinin farklı koşullardaki sonuçları açısından literatürde çok net bir bilgi mevcut olmadığı için tercih söz konusu olduğunda araştırmacılar hangi testi kullanacaklarına çoğu zaman kesin karar verememektedir. Bu çalışmadaki amaç, testlerin seçiminde önemli rol oynayan 1. tip hata yapma olasılığı bakımından söz konusu iki testi çeşitli koşullarda karşılaştırmaktır. Bu amacı gerçekleştirmek için bir simülasyon çalışması planlanmış ve bu çalışmada sıfır hipotezinin geçerli olduğu farklı boyutta 4 çapraz tablo, her tabloda 3 farklı örneklem genişliği ve yine her tabloda 2 farklı marjinal satır ve marjinal sütun olasılıkları olmak üzere toplam 24 farklı koşul incelenmiştir. Her bir koşul 5000 kez tekrarlanarak Pearson ki-kare ve LR istatistiklerinin gözlenen 1. tip hataları hesaplanmıştır. Simülasyon sonuçları incelendiğinde, toplam örneklem genişliği 100 ve 1000 iken satır ve sütun marjinal olasılıklarının dengeli veya dengesiz olmasından etkilenmeksizin, her iki testin gerçekleşen 1. tip hatalarının benzer ve başlangıçta belirlenen %5 seviyesinde olduğu gözlenmiştir. Ancak örnek genişliğinin 50 olması durumunda satır veya sütun olasılıkları dengeli iken LR testinin 1. tip hata olasılığı daha yüksek çıkmış, ancak Pearson ki-kare testinin hata olasılığı %5 seviyesinde korunmuştur. Küçük örneklem genişliğinde ve satır ve sütun olasılıklarının dengesiz dağılımında ise her iki testin hataları hesaplanamamıştır. Tablo boyutu büyüdükçe ve örneklem genişliği 1000 iken sonuçların benzer olduğu ancak örneklem genişliği 100 veya altı durumlarda Pearson ki-kare testinin beklenenden daha az, LR testinin ise beklenenden daha çok hipotezi ret ettiği belirlenmiştir. Bu sonuç örnek genişliğinin azalması ile birlikte satır ve sütun olasılıklarının dengeli dağılımı dengeszleştirilince daha net gözlenmektedir. Kare tablolarda ise satır ve sütun olasılıkları dengeli dağılmış ve örneklem genişliği 100 ve üzeri olduğunda, test sonuçlarında gerçekleşen 1. tip hataların benzer ve belirlenen seviyeyi koruduğu belirlenmiştir. Örneklem genişliği küçüldükçe veya satır-sütun olasılıkları dengesiz dağıldığında LR testinin gözlenen hatası beklenen seviyeyi aşmış, ancak Pearson ki-kare istatistiğinin hatası %5 civarında veya biraz daha küçük çıkmıştır. Kare tablonun boyutu arttıkça bu sonuç daha net gözlenmiştir. Sonuç olarak, örneklem genişliği, tablonun kare veya dikdörtgen yapıda olması ve satır ve/veya sütun marjinal olasılıklarının dengeli veya dengesiz olmasının LR testinin gerçekleşen 1. tip hatasını daha olumsuz etkilediği görülmüştür. Küçük örneklemelerde ise Pearson ki-kare istatistiğinin daha az hipotezi ret ettiği söylenebilir. Bu durum Pearson ki-kare istatistiğinde yanlış negatiflik oranının, LR ki-kare istatistiğinde ise yanlış pozitiflik oranının daha yüksek olduğunu açıklar. Bu sonuçlara göre küçük örneklem söz konusu olduğunda her iki testin de 1. tip hata açısından olumsuz etkilendiği söylenebilir.

**KAYNAKLAR**

- [1] Agresti A. and Yang M. C. (1987), An empirical investigation of some effects of sparseness in contingency tables, *Computational Statistics & Data Analysis*, 5(1): 9-21.
- [2] Bradley D. R., Bradley T. D., McGrath S. G. and Cutcomb S. D. (1979), Type I error rate of the chi-square test of independence in R×C tables that have small expected frequencies, *Psychological Bulletin*, 86(6): 1290-7.
- [3] Lydersen S., Pradhan V., Laake P. and Senchaudhuri P. (2005), Power comparison of exact tests for association in unordered rxc contingency tables using standard, mid p, and randomized test versions, *Journal of Statistical Computation and simulation*, 75(6): 447-58.
- [4] Lydersen S., Pradhan V., Senchaudhuri P. and Laake P. (2007), Choice of test for association in small sample unordered rxc tables, *Statistics in Medicine*, 26(23): 4328-43.

UYARLAMALI AĞ İLE TIP-II BULANIK MANTIĞA DAYALI

YOL KAYBI TAHMİNLERİ

ABSTRACT

COMPARISON OF PEARSON CHI-SQUARE AND LOG-LIKELIHOOD RATIO STATISTICS IN MULTIDIMENSIONAL CROSS-TABLES WITH REGARD TO TYPE I ERROR

Pearson Chi-square and Log-Likelihood Ratio (LR) are frequently used to investigate the relation between categorical variables in cross-tables. However, exact details are not available in literature about the results of two test statistics in different conditions. In some cases, researchers cannot decide on which test statistic should use. This study aims to compare these tests in various conditions, as regards Type-I errors. According to simulation results, LR test is more adversely affected by the structure of table (square/rectangular), sample size, and the balanced/unbalanced marginal row and/or column probabilities than Pearson-test. Also both tests are negatively affected by small sample sizes even if Pearson-statistics rejects less null hypothesis. Predictive value of false-positive of LR test is high when predictive value of false-negative of Pearson Chi-square test is high.

**Key Words:** Pearson Chi-square, Log-Likelihood Ratio, Cross-Tables

Türkan ERBAY DALKILIÇ	Kamile ŞANLI KULA	Berna Yeşim HANCI
Karadeniz Teknik Üniversitesi,	Ahi Evren Üniversitesi,	50 kersey crescent, RG141SZ, Speen, Newbury, UK.
Fen Fakültesi,	Fen Edebiyat Fakültesi,	yesim.hanci1@vodafone.com
İstatistik ve Bilgisayar Bilimleri Bölümü	Matematik Bölümü	
tedalkilic@gmail.com	sanli2004@hotmail.com	

Giriş

Kablosuz iletişimde mobil terminallerinin pozisyonlarının belirlenmesi önemli bir problemdir. Mobil terminallerinin konumlarının belirlenmesinde faydalanılan metotlardan biri radyo dalgaları yol kaybı ölçümlerinin kullanılmasıdır. Bu çalışmada, tip-II bulanık mantığa dayalı uyarlamalı ağın kullanıldığı bir algoritma önerilmiş ve bu algoritmadan elde edilen sonuçlar yol kaybı modelleri için en yaygın kullanılan yöntemlerden biri olan Bertoni-Walfisch modelinden elde edilen sonuçlar ile karşılaştırılmıştır. Çalışmada kullanılan yayılım ölçümleri İstanbul'da 900 MHz bandında toplanmıştır.

Tip-II Bulanık Mantık

Tip-II bulanık sistemler, tip-II bulanık kümeleri içeren bulanık eğer-ise kurallarından oluşurlar. Temel olarak bir tip-II bulanık küme üyelik fonksiyonu hakkında belirsizliğe sahip bir kümedir. Bulanıklığın sadece dilsel değişkenlerle sınırlanmadığı üyelik fonksiyonunun tanımında da devam ettiği tip-II bulanık kümenin, geleneksel (tip-I) bulanık mantığın bir genelleşmesi olduğunu söylenebilir. Bir tip-II bulanık küme bulanık üyelik fonksiyonu ile karakterize edilirler, bu kümenin her bir elemanı için üyelik derecesi  $[0,1]$  aralığında bulanık bir kümedir.

Uyarlamalı Ağ İle Tip-II Bulanık Mantığa Dayalı Yol Kaybı Tahmini

Regresyon modelinin parametrelerinin belirlenmesi süreci bağımsız değişkenlerin sınıf ya da düzey sayılarının ve önsel parametrelerin belirlenmesi ile başlar. Önsel parametreler dağılımı karakterize eden parametrelerdir ve bu çalışmada bağımsız değişkenlerin normal dağılıma sahip olması durumu ele alındığından normal üyelik fonksiyonuna ilişkin merkez ( $m$ ) ve yayılım ( $\sigma$ ) parametreleri ile ilgilenilecektir. Bağımsız değişkenlerin normal dağılımdan gelmesi durumunda sonsal parametre olarak da tanımlanabilen regresyon modelinin bilinmeyen parametrelerinin belirlenmesi için önerilen yöntemle ilişkin algoritma ana hatları ile aşağıdaki gibi tanımlanabilir.

**ADIM 1:** Önsel parametreler belirlenir.

**ADIM 2:** Uyarlamalı ağın birinci tabakasındaki sinir fonksiyonları bağımsız değişkenlerin geldiği dağılıma ait üyelik fonksiyonunun Normal Dağılım fonksiyonu olduğu düşünüldüğünde, üyelik fonksiyonları aşağıdaki biçimde tanımlanır. Burada merkez parametresi olan  $m$  bulanık bir parametredir ve  $m \in [m_1, m_2]$  aralığında değerler alır.

$$\mu_{F_h}(x_i) = \exp \left\{ - \left[ \frac{x_i - m_h}{\sigma_h} \right]^2 \right\}$$

**ADIM 3:** Bağımsız değişkenlerin üyelik fonksiyonlarını karakterize eden önsel parametrelere merkez parametresi  $m$ 'nin bulanık sayı olması durumunda, regresyon modelinin bilinmeyen katsayıları olan sonsal parametre seti  $c_i^L = (a_i^L, b_i^L)$  ( $i=1, \dots, p$ ) biçiminde elde edilir.

**ADIM 4:** Adım 3'de elde edilen sonsal parametre seti  $c_i^L = (a_i^L, b_i^L)$  kullanılarak,

$Y^L = c_0^L + c_1^L x_1 + c_2^L x_2 + \dots + c_p^L x_p$  biçiminde ifade edilen regresyon modelleri oluşturulur.

**ADIM 5:** Modele ilişkin hata hesaplanır. Eğer  $\varepsilon < 0$  ise ulaşılan sonsal parametre, kurulacak olan regresyon modellerinin parametreleri olarak elde edilmiştir, sürece son verilir. Eğer  $\varepsilon \geq 0$  ise adım 6'ya geçilir.  $0$  karar verici tarafından belirlenen küçük sabit bir değer,

**ADIM 6:** Adım 1'de belirlenen merkezi önsel parametreler güncellenir. Belirlenen en küçük hatayı veren önsel parametreler ve bu parametrelere ilişkin modellerden elde edilen tahmin çıktı olarak alınır.

**Anahtar Kelimeler:** Uyarlamalı ağ, tip-II bulanık mantık, yol kaybı.

## KAYNAKLAR

- [1] Chi-Bin, C. and Lee, E. S. (2001), Switching Regression Analysis by Fuzzy Adaptive Network, European Journal of Operational Research, 128, 647-663
- [2] Castillo, O., Melin, P.(2008) "Type-2 fuzzy logic: theory and applications", Springer.
- [3] Jyh-Shing Roger J. (1993), ANFIS: Adaptive-Network-Based Fuzzy Inference System, IEEE Transaction on Systems, Man and Cybernetics, 23, No:3, 665-685
- [4] J. Walfisch and H.L. Bertoni,(1998) "A theoretical model of UHF propagation in urban environments", IEEE Trans. Antennas and Propagation, Vol. 36, No. 12, pp. 1788-1796.
- [5] Dubois, D. and Prade, H., 1980, Fuzzy Sets and Systems: Theory and Application, Academic, New York.

## BULANIK REGRESYONDA GUSTAFSON-KESSEL KÜMELEME ALGORİTMASINA DAYALI PARAMETRE TAHMİNİ

Türkan ERBAY DALKILIÇ

Ersegül GÖKTÜRK ŞAHİN

Karadeniz Teknik Üniversitesi,

Karadeniz Teknik Üniversitesi,

Fen Fakültesi,

Fen Fakültesi,

İstatistik ve Bilgisayar Bilimleri Bölümü

İstatistik ve Bilgisayar Bilimleri Bölümü

tedalkilic@gmail.com

ersegul@hotmail.com

### Giriş

Regresyon çözümlemesinde verilerin farklı dağılımlara sahip olması durumu tahmin sürecinde klasik çözümlemelerin dışına çıkılmasını gerektirir. Böyle durumlarda Bulanık kümeleme yöntemleri alternatif yöntemler olarak kendini göstermektedir. Bulanık regresyon çözümlemesinin önemli adımlarından biri veri setini meydana getiren kümelerin belirlenmesi ve bu kümelere yer alan verilerin tahmine katkılarının derecelerini belirleyecek üyelik derecelerinin elde edilmesidir. Bu çalışmada veri setlerinin üyelik derecelerinin belirlenmesi aşamasında Gustafson-Kessel kümeleme algoritmasından faydalanılmış, elde edilen üyelik derecelerine dayalı bulanık regresyon çözümlemesi için bir algoritma önerilerek algoritmadan elde edilen tahminler mevcut yöntemler ile elde edilen tahminler ile karşılaştırılmıştır.

### Gustafson-Kessel Algoritması

Literatürde yer alan birçok bulanık kümeleme algoritmalarından biri de Gustafson-Kessel algoritmasıdır. Diğer c-ortalama dayalı bulanık kümeleme algoritmalarından farkı veri kümelerinin elips biçiminde dağılım göstermeleri durumunda, verilerin var olan kümelere ait olma durumlarını belirlemede daha hassas olmasıdır. Algoritmanın işleyişinde öncelikle başlangıç değerleri belirlenir bunlar; küme sayısı, iterasyon sayısı, kabul edilebilir hata miktarı, ve başlangıç üyelik derecelerdir. Daha sonra her küme için küme merkezleri Eşitlik 1 ile belirlenir;

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (1)$$

Merkezleri belirlenen her bulanık küme için bulanık kovaryans matrisi Eşitlik (2) ile hesaplanır.

$$F_i = \frac{\sum_{j=1}^n u_{ij}^m (x_j - v_i) (x_j - v_i)^T}{\sum_{j=1}^n u_{ij}^m} \quad (2)$$

Bulanık kümelere yer alan her veri için Mahalanobis uzaklığı;

$$D_{ij, A_i}^2 = (x_j - v_i)^T A_i (x_j - v_i) \quad (3)$$

biçimde elde edilir. Burada,

$$A_i = \sqrt{\det(F_i)} F_i^{-1} \quad (4)$$

Gustafson-Kessel algoritması Eşitlik (5) ile verilen amaç fonksiyonunun minimize edilmesine dayanır bu fonksiyonu minimum yapan üyelik dereceleri verilerin var olan bulanık kümelere optimal üyeliklerini göstermektedir.



$$J(X, u, v) = \sum \sum u_{ij}^m D_{ij, A_i}^2 \quad (5)$$

#### G-K Kümeleme Algoritmasına Dayalı Parametre Tahmini

Regresyon modelinin parametrelerinin belirlenmesi ve tahmin değerlerin elde edilmesi sürecinde gözlemlerin aynı sınıfa ait olup olmamaları önemlidir. Farklı sınıflardan gelen veriler birer küme olarak düşünülürken bulanık regresyon modellerinin elde edilerek hatası küçük tahminlere ulaşmak için verilerin kümelere ait olma derecelerinin elde edilmesi gerekir. Bu çalışmada mevcut bulanık kümeleme yöntemlerinden G-K bulanık kümeleme algoritması kullanılarak üyelik dereceleri elde edilmiş ve elips biçimli dağılım gösteren kümelere ait olma dereceleri belirlenmiştir. Daha sonra elde edilen bu optimal üyelik dereceleri her küme için kurulan modelde ağırlık olarak yer almış ve bu ağırlıklara bağlı tahminler elde edilmiştir. Tahminler mevcut c-ortalamaya dayalı kümeleme algoritmalarından elde edilen üyelik dereceleri kullanılarak elde edilen tahmin sonuçları ile karşılaştırılmıştır.

**Anahtar Kelimeler:** G-K Algoritması, bulanık regresyon çözümlemesi.

#### KAYNAKLAR

- [6] Davari, A., Marhaban, M.H., Noor, S.B.M., Karimadini, M., and Karimodini, A., (2011), "Parameter Estimation of K-distributed sea clutter based on fuzzy inference and Gustafson-Kessel Clustering", Fuzzy Sets and Systems, Vol. 163, pp. 45-53.
- [7] Krishnapuram, R., and Kim, J., (1999) "A note on the Gustafson-Kessel and Adaptive Fuzzy Clustering Algorithms" IEEE Transaction on Fuzzy Systems, Vol. 7, No. 4, pp. 453-461.
- [8] Kim, Y., Kim, D.W., Lee, D., and Lee, H.K., (2004), "A cluster validation index for GK cluster analysis based on relative degree of sharing" Information Sciences Vol. 168, pp. 225-242.

#### PARAMETER ESTIMATION IN FUZZY REGRESSION BASED GUSTAFSON-KESSEL ALGORITHM

##### ABSTRACT

In this paper, fuzzy membership degrees of each data obtained using the Gustafson-Kessel fuzzy clustering algorithm and determined degrees of belonging distributed in an ellipse-shaped clusters. Then, these membership degrees are used by weight in estimation process in fuzzy regression analysis. The predicted values are compared with estimates obtained from other classical methods.

**Key Words:** G-K algorithm, fuzzy regression analysis.

#### YAPAY SINIR AĞLARI VE SOSYAL AĞ MERKEZİLİK ÖLÇÜLERİ İLE MÜŞTERİ KAYBI ANALİZİ

Vildan GÜLPINAR  
İstanbul Aydın Üniversitesi , Anadolu Bil MYO, Florya Kampüsü, Küçükçekmece İstanbul Türkiye  
E-mail: vildangulpinar@aydin.edu.tr

Dilek ALTAŞ  
Marmara Üniversitesi , İktisat Fakültesi, Göztepe Kampüsü, 34722, İstanbul Türkiye  
E-mail: dilekaltas@marmara.edu.tr

##### Giriş

İletişim toplulukları, her bireyin ağıdaki diğer bireylerle zayıf ya da güçlü ilişki kurabildiği Sosyal Ağlar (SA) olarak ele alındığında, müşterilerin ağıdaki etkisinin ortaya çıkarılması için Sosyal Ağ Analizi (SAA) güçlü bir yöntem kabul edilmektedir.

Müşteri kaybı olayı, zincirleme bir etkiye sahiptir. Bu zincir süreci, SAA'nın güçlü bir düğümü tarafından hatta daha az gelir sağlayan fakat müşteri bağlılığı sürecinde güçlü etkiye sahip bir müşteri tarafından da başlatılabilir. Bu nedenle firmalar, etkileri yüksek müşterileri ortaya çıkarabilmelidirler. SAA'da güçlü ve merkezi düğüm olarak tanımlanan bu müşteriler, "müşteri kaybı" gibi zincirleme etki yaratan olaylarda ne kadar etkili olabildikleri dikkate alınarak seçilebilir. Böylece uygulanacak pazarlama yöntemi, etkileri yüksek bireylere göre belirlenebilir.

SA merkezilik ölçüleri ile müşteriler arasındaki bağlantıların ağırlıkları, zincir sürecinin ağıdaki etkisini ortaya koyabilmektedir. Müşteri bağlantılarının anlaşılması ve SA yapısında her zincirin merkezi çekirdeklerinin belirlenmesi kitlesel müşteri kaçışlarını önlemek ve böylece gelir kaybindan kurtulmanın en iyi yoludur. SAA, topluluk içindeki müşteri kaybı olaylarında olası korelasyonları açığa çıkararak, olayın SA'nın çekirdek bir düğümü tarafından tetiklendiğinde daha güçlü etkileme gücüne sahip olduğu, oysa çevresel bir düğüm tarafından tetiklendiği zaman etkisinin daha az olduğunu kanıtlar.

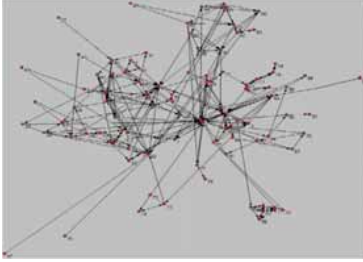
Yukarıda anlatılanlar ışığında makalenin amacı, müşteri kayıplarının çok hızlı yaşandığı telekomünikasyon piyasasında müşteri kaybını Yapay Sinir Ağları (YSA), yöntemiyle tahmin etmek ve gelen ve giden arama miktarları dikkate alınarak SAA sonucunda etkili müşteriler değerlendirilerek etkin bir pazarlama yöntemi sunmaktır.

##### Uygulama

Analizlerde kullanılacak veriler, Türkiye'nin farklı illerinde ikamet etmekte olan 100 farklı GSM operatörü kullanıcılarına anket uygulanarak elde edilmiştir. 14 sorudan oluşan anket 2 bölümden oluşmaktadır.

SAA'ya uygun olabilmesi için veriler "kartopu örnekleme" tekniğiyle elde edilmiştir. YSA, MATLAB 11 programında analiz edilmiştir. SAA için, özellikle görsel gösterim avantajı ve kullanım kolaylığından dolayı Pajek programı kullanılmıştır.

Müşterilerin SA'daki görüntüleri Şekil 1'de gösterilmektedir.



Şekil 1. Sosyal Ağ Görüntüsü

Bu çalışma ile telekomünikasyon piyasasında müşteri kaybı tahmini yapılmıştır. Daha sonra müşterilerin her biri ağın düğümlerini ve müşterilerin birbirlerini arama miktarları ağın bağlantı değerlerini göstermek üzere iletişim ağı incelenmiştir. Elde edilen bulgular ışığında, telekomünikasyon piyasasında kaybedilme riski taşıyan abonelerin tahminini veren YSA sonuçları dikkate alınarak, iletişim ağındaki bazı müşterilerin SA'daki etkinlikleri şu şekilde incelenebilir:

- Ağın hem önemli (hublar/ oteritelere göre) hem de güçlü (merkezilik ölçülerine göre) düğümleri arasında yer alan 1 nolu müşterinin kaybedilme riski -1,6118 olarak tahmin edilmiştir ve bu değer kaybedilmeye en yakın müşteri grubuna işaret etmektedir. Arasındalık merkeziliği en yüksek yani ağın ikili düğümlerinin yolları arasında en fazla bulunan müşteri olduğu düşünüldüğünde, kaybedilmesi durumunda, kaybı diğer müşterilere en kolay ve en hızlı (yakınlık merkeziliğinden dolayı) yayacak ve zincirleme kaybı başlatacak müşteri olacaktır.

#### KAYNAKLAR

- [1] Pinheiro, Carlos Andre Reis - Marus Helfert,(2009), "Mixing Scores from the Artificial Neural Network Analysis to Improve the Customer Loyalty", **IEEE Computer Society**, s.954-959.
- [2] Geppert, Carl, (2003), Customer Churn Management: Retaining High-Margin Customer with Relationship Management Techniques, s.2-3. [https://www.amr.kpmg.com/microsite/kpmgme/downloads/CHURN\\_02\\_26final.pdf](https://www.amr.kpmg.com/microsite/kpmgme/downloads/CHURN_02_26final.pdf).

#### ABSTRACT

#### CUSTOMER CHURN ANALYSIS WITH THE HELP OF THE ARTIFICIAL NEURAL NETWORKS AND SOCIAL NETWORK ANALYSES CENTRALITY MEASURES

This paper is intended to estimate the loss of customers in the Turkish telecommunication market with the help of the Artificial Neural Networks (ANN) and examine the positions and effects on the network of the customers for whom there is a risk of loss, by analyzing the customer communication network with the help of the Social Network Analysis (SNA).

**Key Words:** Artificial Neural Networks, Social Network Analysis, Customer Churn Management, Telecommunication Market

#### VERİ ZARFLAMA ANALİZİ İLE TÜRKİYE'DEKİ İLLERİN KÜLTÜREL AÇIDAN DEĞERLENDİRİLMESİ

\*Hakan ALTUNAY, Kezban BULUT

\*Bursa Teknik Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Endüstri Mühendisliği Bölümü, Osmangazi, Bursa, hakanaltunay@rocketmail.com

Kırıkkale Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Yahşihan, Kırıkkale, kezbanbulut@yahoo.com

Genişletilmiş Özet

Veri Zarflama Analizi (VZA); doğrusal programlamanın özel bir uygulama şekli olup, aynı amaç ve hedeflere sahip sistemlerin göreceli olarak verimliliğini ölçmede kullanılır [4]. Veri Zarflama Analizi ilk olarak Charnes, Cooper ve Rhodes tarafından 1978 yılında kâr amaçlı kurulmayan ve kamu hizmeti sağlayan kuruluşların örgütsel performansını izleyebilmek amacıyla geliştirilmiştir [1]. Yöntem, günümüzde sağlık sektörü, üniversiteler, bankacılık, ulaşım gibi pek çok alanda uygulanma alanı bulmaktadır.

VZA, doğrusal programlama prensiplerine dayanan, spesifik olarak Karar Verme Birimlerinin (KVB) kullandıkları girdileri hangi etkinlik derecesinde çıktıya dönüştürdüğünü tespit etmemize imkan sağlayan ve duyarlılık analizi yöntemiyle kaynakların daha etkin kullanılması amacıyla girdi ve çıktıların ayarlanmasını mümkün kılan bir teknik olarak ifade edilmektedir [3].

Yapılan bu çalışmada Veri Zarflama Analizi tekniğinden yararlanılarak Türkiye sınırları içerisinde bulunan 81 ilimizin kültürel açıdan değerlendirilmesi sağlanmıştır. Çalışmaya kaynak oluşturan istatistiksel veriler Türkiye İstatistik Kurumu (TÜİK) tarafından yayınlanan Kültür İstatistikleri 2011 isimli yayınından alınmıştır [5]. Çalışmada VZA çözümü için Frontier Analyst paket programından yararlanılmıştır. Bu program, dünyada halen 35 ülkede kullanılmakta olup, akademik çevrelerce de geniş ölçüde kabul görmüştür. Çizelge 1.'de VZA modelinde kullanılan girdi ve çıktılar verilmiştir.

Çizelge 1. VZA modelinin girdileri ve çıktıları

Girdiler	Çıktılar
X1: Matbaa sayısı	Y1: Müze ve ören yeri ziyaretçi sayısı/il nüfusu
X2: Müze ve ören yeri sayısı	Y2: Halk Kütüphanelerinden yararlanan kullanıcı sayısı/il nüfusu
X3: Halk Kütüphanesi sayısı	Y3: Kütüphanelerden ödünç verilen materyal sayısı/il nüfusu
X4: Kütüphanelerdeki toplam kitap sayısı	Y4: Tiyatro seyirci sayısı/il nüfusu
X5: Tiyatrolarda oynanan gösteri sayısı	Y5: Sinema seyirci sayısı/il nüfusu
X6: Tiyatrolardaki toplam koltuk sayısı	
X7: Sinemalardaki toplam koltuk sayısı	
X8: Sinemalarda gösterilen film sayısı	
X9: İlde yaşayan halkın gelir seviyesi	

Kültür, toplumların binlerce yıllık birikimlerinin çeşitli boyutlarda bir bileşimi ve hayata yansıma biçimidir. Kültür istatistikleri de bu birikimlerin sayısal göstergeleridir. Kültür değerlerini yansıtan kültür istatistiklerinin bir bölümü kütüphaneler, müzeler, tiyatro, opera, bale ve sinemalara ilişkin bilgilerdir [2]. Çalışmada amaç Türkiye’de bulunan söz konusu illerin her birisi için yapılan kültürel yatırımlara karşın, bu illerde yaşayan halkın bu yatırımlara ne oranda cevap verdiğinin göreceli olarak değerlendirilmesidir. Bu sayede halkın olası kültürel ihtiyaçlarına paralel olarak iyileştirme önerileri sunulabilecektir. Diğer açıdan da ihtiyaç fazlası bölgeler belirlenerek yapılacak gereksiz yatırımların önüne geçilebilmesi sağlanacaktır.

## KAYNAKLAR

[1] A. Charnes, W.W. Cooper and E. Rhodes (1978), "Measuring the Efficiency of Decision Making Units," European Journal of Operational Research 2, pp.

[2] Çakmak, Z., Uzgören, N., Keçek, G., (2005), Kümeleme Analizi Teknikleri İle İllerin Kültürel Yapılarına Göre Sınıflandırılması Ve Değişimlerinin İncelenmesi, Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, Sayı: 12, Haziran 2005.

[3] Eroğlu, Ergün ve Melek C. Atasoy (2006), "Veri Zarflama Analizi ile Etkinlik Ölçümü ve Etkin Karar Birimlerinin Duyarlılık Analizi", İstanbul Üniversitesi İşletme Fakültesi Dergisi, 35(2), ss. 91–106.

[4] Onaran, S. ,2006. Veri Zarflama Analizi Kullanılarak Üniversite Kütüphanelerinin Performanslarının Değerlendirilmesi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

[5] TÜİK, Kültür İstatistikleri (2011)

## ABSTRACT

### EVALUATION OF CULTURALLY THE CITIES IN TURKEY BY USING DATA ENVELOPMENT ANALYSIS

Data Envelopment Analysis (DEA) is a non-parametric productive efficiency measurement method for operations with multiple inputs and multiple outputs. . In this study we analyzed culturally relative comparison of the cities in Turkey by using Data Envelopment Analysis. This study consist of two phase. In first phase, inputs and outputs used in the DEA model are determined. Second stage Data Envelopment Analysis results are evaluated. According to the results obtained from the study is suggested that some improvements. In this way, investments will be made to meet the cultural needs of the people.

**Key Words:** Data Envelopment Analysis, Efficiency Analysis, Operational Research

### SEDAN VE REASÜRÖR AÇISINDAN OPTİMAL REASÜRANS ANLAŞMALARI Yasemin GENÇTÜRK, Murat KIRKAĞAÇ, Rümeyza KARATAŞ\*

Hacettepe Üniversitesi Beytepe Kampüsü Aktüerya Bilimleri Bölümü ANKARA  
yasemins@hacettepe.edu.tr murat.k@hacettepe.edu.tr rmys@hacettepe.edu.tr

#### Reasürans nedir? Reasürans türleri nelerdir?

Reasürans, sigorta edilmiş riskin, belirli bir kısmının veya tamamının yeniden sigorta edilmesidir. Sigorta şirketleri, teminat verdikleri rizikolarda büyük hasarların aynı zamanda meydana gelme ihtimaline karşı, hasar ödemelerinde zorlanmamak için reasürans (mükerrer sigorta) yaptırır. Reasürans işlemlerinde riski devreden şirkete sedan, devir alan şirkete reasürör denir.

Reasürans türleri temel olarak ihtiyari ve otomatik olmak üzere ikiye ayrılır.

**İhtiyari (Facultative) Reasürans Anlaşmaları :** Sigorta şirketinin riski reasüre etmede, reasürör de kabul etmede serbest olduğu anlaşmalardır. . Büyük ve spesifik risklerin reasüransı genellikle ihtiyari reasürans ile yapılır. Bölümlü ve bölümsüz reasürans olmak üzere ikiye ayrılır. Belirli başlı bölümlü reasürans anlaşmaları, kotpar(quota-share) ve eksedan(surplus) olmak üzere iki çeşittir. Bölümlü olmayan reasürans anlaşmaları ise, hasar fazlası(excess of loss), hasar oranı fazlası(stop loss), toplam hasar fazlası ve toplam hasar oranı fazlası olarak verilebilir.

**Otomatik (Automatic) Reasürans Anlaşmaları :** Reasürans şartlarının isteğe bağlı olmadığı anlaşmalardır. Reasürör ve sigorta şirketleri işleri otomatik olarak devir ve kabule zorunludur. Sorumluluk paylaşımı sigortalı meblağ yani azami teminat ya da hasar üzerine yapılır. İhtiyari reasürans anlaşmalarında olduğu gibi otomatik reasürans anlaşmalarının da, bölümlü ve bölümsüz türleri vardır.

Reasürans anlaşmalarında sigorta şirketinin üstlendiği risk sedan ile reasürör arasında paylaşıldığından, sigorta şirketi topladığı primin bir kısmını reasüröre devreder. Sigortalının ödemesi gereken primin hesaplanmasında kullanılan farklı yöntemler vardır.

#### Prim Hesaplama Yöntemleri

Belirli başlı prim hesaplama yöntemleri Net Prim, Beklenen Değer ve Varyans prensibi olarak sıralanabilir.

$$\text{Net prim prensibi : } \pi_x = E[X]$$

Net prim, sigortacının risk altındaki beklenen toplam hasar miktarına eşittir.

$$\text{Beklenen değer prensibi : } \pi_x = (1 + \theta)E[X]$$

0 prim yükleme faktörüdür ve sıfırdan büyüktür,  $\theta > 0$ . Bu prensibe göre hesaplanan prim aynı ortalamaya sahip farklı riskler için aynıdır.

$$\text{Varyans prensibi : } \pi_x = E[X] + \alpha V[X], \alpha > 0$$

Varyans prensibinde yüklem toplam hasar miktarının varyansı ile orantılıdır. Bu prensip beklenen değer prensibine göre daha duyarlıdır yani riskteki değişkenliği dikkate alır.

Bu eşitliklerde  $X$ , belirli bir zamanda sigortacının toplam hasar veya kayıp miktarının gösteren raslantı değişkenidir.

#### Optimal Reasürans Anlaşmaları

Optimal reasürans anlaşmaları, optimizasyon kriterleri ve prim hesabında kullanılan prensibe bağlı olarak değişmektedir. Optimal reasürans anlaşmasının belirlenmesine ilişkin çalışmalar incelendiğinde, çoğunlukla sedan şirketin (sigortacı) faydasının göz önünde bulundurulduğu görülmektedir. Arrow (1963) ve Kalutza (2008) sigorta şirketinin beklenen fayda fonksiyonu maksimize eden optimal reasürans anlaşmasının belirlenmesi; Vajda (1962) ve Bowers (1997) sigorta şirketinin riskinin varyansını minimize eden optimal reasürans anlaşmasının belirlenmesi, Cai (2008), Balbas (2009) ve Weng (2009) VaR ve CTE gibi belli başlı risk ölçümlerini minimize eden optimal reasürans anlaşmasının belirlenmesine ilişkin çalışmalar yapmışlardır. Ancak sigortacı açısından optimal olan reasürans anlaşmaları reasürör açısından kabul edilemez olabilir. Bu nedenle her iki tarafın da faydası gözetilerek optimal anlaşmanın belirlenmesi önemlidir. Bu çalışmada hem sedan hem de rasürörün birleşik yaşam olasılığı ile birleşik kar olasılığını maksimize eden optimal reasürans anlaşmaları incelenmiştir.

#### KAYNAKLAR

- [1] Balb'as, A., B. Balb'as, and A. Heras, 2009, Optimal Reinsurance With General Risk Measures, Insurance: Mathematics and Economics, 44(3): 374-384.
- [2] Bowers, N. L., H. U. Gerber, J. C. Hickman, D. A. Jones, and C. J. Nesbitt, 1997, Actuarial Mathematics, 2nd edition (Schaumburg, IL: Society of Actuaries).
- [3] Cai, J., K. S. Tan, C. G. Weng, and Y. Zhang, 2008, Optimal Reinsurance Under VaR and CTE Risk Measures, Insurance: Mathematics and Economics, 43: 185-196.
- [4] Cai J, (2012), Optimal Reciprocal Reinsurance Treaties between Insurers and Reinsurers, Department of Statistics and Actuarial Science University of Waterloo, Canada.
- [5] Cai J, Ying Fang, Zhi Li, Gordon E. Willmot. (2013), Optimal Reciprocal Reinsurance Treaties Under The Joint Survival Probability and The Joint Profitable Probability, The Journal of Risk and Insurance.

#### OPTIMAL RECIPROCAL REINSURANCE TREATIES ABSTRACT

When studies related to optimal reinsurance are examined, it is usually seen that only insurer's point of view is taken into account. However, an optimal reinsurance contract for an insurer may not be optimal for a reinsurer and it might be unacceptable for a reinsurer.

Therefore, it is important to design an optimal reinsurance contract in which interests of both an insurer and reinsurer are considered. This study examines an optimal reinsurance contract in which joint survival probability and profitable probability of reinsurer and insurer are maximized.

**Key Words:** Types of Reinsurance Contracts, Premium Principles, Optimal Reinsurance.

