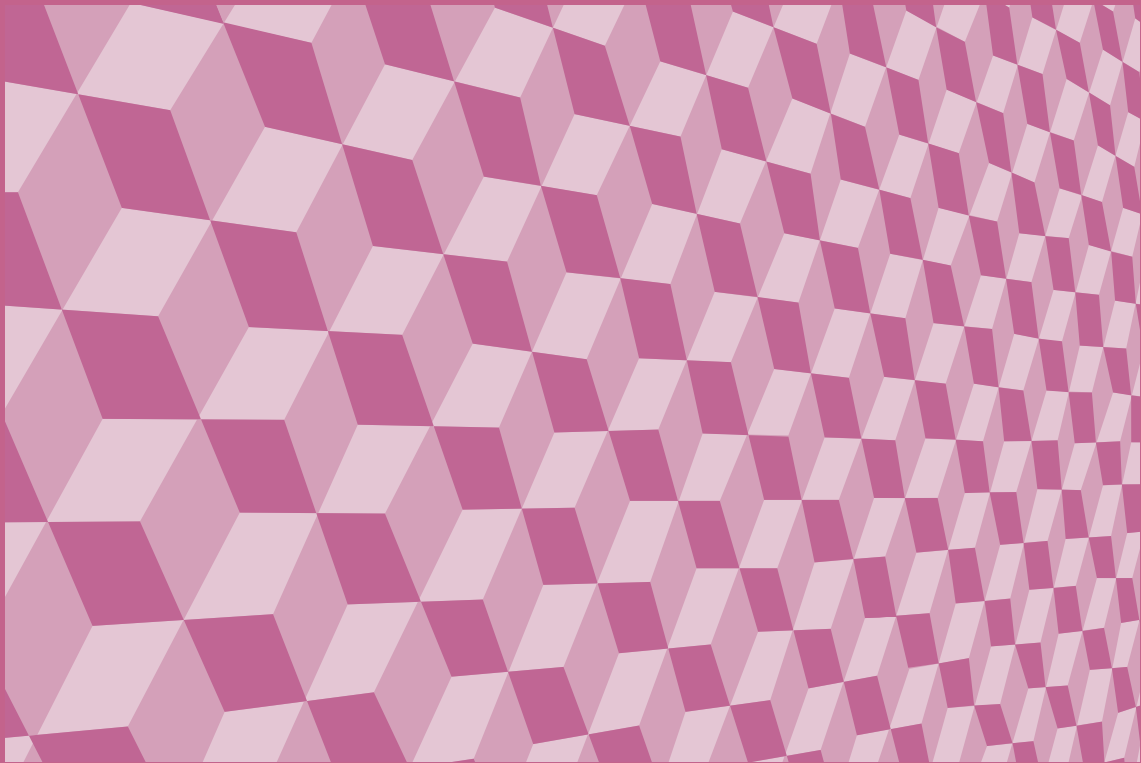




# İSTATİSTİK ARAŞTIRMA DERGİSİ Journal of Statistical Research

**Cilt-Volume: 09 Sayı-Number: 02**  
**Özel Sayı-Special Issue 2012**

ISSN 1303-6319



**Türk İstatistik Derneği, 7. Uluslararası İstatistik Kongresi Bildirileri,**  
**28 Nisan - 01 Mayıs 2011, Antalya**  
The Turkish Statistical Association, Proceedings of 7th Statistics Congress,  
28 April - 01 May 2011, Antalya

**TÜRKİYE İSTATİSTİK KURUMU**  
**Turkish Statistical Institute**



# İSTATİSTİK ARAŞTIRMA DERGİSİ Journal of Statistical Research

**Cilt-Volume: 09 Sayı-Number: 02**  
**Özel Sayı-Special Issue 2012**

**Yayın istekleri için** For publication order

**Döner Sermaye İşletmesi** Revolving Fund Management

**Tel:** + (312) 425 34 23 - 410 05 96 - 410 02 85

**Faks-Fax:** + (312) 417 58 86

**Yayın içeriğine yönelik sorularınız için** For questions about contents of the publication

**Dergi Editörlüğü** Journal Editorship

**Tel:** + (312) 410 03 67 - 284 45 00/171

**Faks-Fax:** + (312) 425 34 05

**İnternet** Internet  
**http://www.tuik.gov.tr** http://www.turkstat.gov.tr

**E-posta** E-mail  
**dergi@tuik.gov.tr** journal@tuik.gov.tr

**Yayın No** Publication Number  
**3944**

**ISSN**  
**1303-6319**

**Türkiye İstatistik Kurumu** Turkish Statistical Institute

**Yücetepe Mah. Necatibey Cad. No: 114 06100 Çankaya-ANKARA / TÜRKİYE**

**Bu yayının 5846 Sayılı Fikir ve Sanat Eserleri Kanununa göre her hakkı Türkiye İstatistik Kurumu Başkanlığına aittir. Gerçek veya tüzel kişiler tarafından izinsiz çoğaltılamaz ve dağıtılamaz.**

Turkish Statistical Institute reserves all the rights of this publication. Unauthorised duplication and distribution of this publication is prohibited under Law No: 5846.

**Türkiye İstatistik Kurumu Matbaası, Ankara** Turkish Statistical Institute, Printing Division, Ankara

**Tel: 0312 410 01 64 \* Fax: 0312 418 50 82**

**Ağustos 2013** August 2013

**MTB: 2013-575 - 500 Adet-Copies**

**Editör Notu**

Değerli Okuyucular,

Türkiye İstatistik Kurumu tarafından 2001 yılından bu yana hakemli olarak yürütülmekte olan "İstatistik Araştırma Dergisi" ile istatistiki araştırmaların niteliğinin yükseltilmesi, kuramsal ve uygulama alanındaki araştırmacılar arasında iletişimin ortak çalışma ve yayınlarla güçlenmesi sağlanmaya çalışılmaktadır.

Evrensel bilimin paylaşılmasını sağlayan bilimsel dergilerin temel işlevi; bilimsel makale yazarının çalışmasını en etkin biçimde ifade etmesine yardımcı olmak ve bilimi anlaşılabilir bir biçimde yayınlamaktır.

Akademisyen, araştırmacı ve okuyucuların artan ilgisine paralel olarak bizlerin çabası, azmi ve kararlılığı da artacak olup, dergimiz daha üst seviyelere taşınacaktır. Dergimizin ulusal ve uluslararası endekslerde taranması çalışmaları da devam etmektedir. Bu kapsamda TÜBİTAK ULAKBİM'e on-line başvuru yapılmış olup, sonuç beklenmektedir. Bu konuya ilişkin olarak alınacak sonuçlar sizlerle paylaşılacaktır.

28 Nisan-01 Mayıs 2011 tarihleri arasında Türk İstatistik Derneği ve Hacettepe Üniversitesi tarafından ortaklaşa düzenlenmiş olan 7. Uluslararası İstatistik Kongresi'nde sunulan ve hakemlik sürecinden geçerek kabul edilmiş makalelerin yer aldığı bu sayımız, İstatistik Araştırma Dergisinin "Özel Sayısı" olarak yayımlanmıştır.

Bu özel sayımızda, kavramsal, kuramsal ve uygulamalı çalışmalar olmak üzere toplam oniki adet çalışmayı siz değerli okuyucularımızla paylaşmanın gururunu taşıyoruz. Bu değerli çalışmaları, bizlerle ve siz değerli okuyucularımız ile paylaşan sayın yazarlara teşekkür ederiz. Ayrıca çalışmaların daha nitelikli hale gelmesinde çok değerli öneri, eleştiri ve katkılarını esirgemeyen sayın hakemlere de şükranlarımızı sunuyoruz.

Dergi'nin basım aşamasına gelmesinde emeğini ve desteklerini esirgemeyen TÜİK Başkanı Sayın Birol AYDEMİR'e, derginin her aşamasında emeği geçen Editör Yardımcısı Sayın Doç. Dr. Özlem İLK'e, dergi çalışmalarını içtenlikle ve azimle yürüten Dergi Sekreteryası'na ve son olarak da emeği geçen diğer tüm TÜİK çalışanlarına teşekkürlerimi iletmek isterim.

Bu sayımızın da akademisyenler ile araştırmacılara faydalı olması temennisi ve gelecek sayılarda hedeflenenler ölçüsünde tekrar buluşmak dileği ile saygılar sunarım.

**Prof. Dr. Fetih YILDIRIM**  
**Dergi Editörü**

**TÜRKİYE İSTATİSTİK KURUMU** **TURKISH STATISTICAL INSTITUTE**  
**İSTATİSTİK ARAŞTIRMA DERGİSİ** **JOURNAL OF STATISTICAL RESEARCH**

**Sahibi** **Owner**  
Türkiye İstatistik Kurumu Adına On Behalf of Turkish Statistical Institute  
Birol AYDEMİR Birol AYDEMİR  
Türkiye İstatistik Kurumu Başkanı President, Turkish Statistical Institute

**Editör** **Editor**  
Prof. Dr. Fetih YILDIRIM Prof. Dr. Fetih YILDIRIM

**Editör Yardımcısı** **Assistant Editor**  
Doç. Dr. Özlem İLK Assoc. Prof. Özlem İLK

**Sekreteryaya** **Secretariat**  
Buket AKGÜN  
Z.Nur EMRE  
Nurdan ELVER

İÇİNDEKİLER	CONTENTS
Sayfa Page	
ÖNSÖZ	III FOREWORD
İÇİNDEKİLER	VII CONTENTS
AMAÇ VE KAPSAM	X AIM AND SCOPE
HAKEM LİSTESİ	XII REFEREE LIST
<b>İstanbul Temiz Su Tüketiminin Bulanık Zaman Serisi Yaklaşımı ile Öngörüsü</b>	<b>1 Forecasting of İstanbul's Clean Water Consumption with Fuzzy Time Series Approaches</b>
<i>Faruk ALPASLAN Özge CAĞCAĞ Damla İLTER Ufuk YOLCU</i>	<i>Faruk ALPASLAN Özge CAĞCAĞ Damla İLTER Ufuk YOLCU</i>
<b>En Düşük ve En Yüksek Altın Fiyatlarının Yapay Sinir Ağları ile Öngörüsü</b>	<b>12 Highest and Lowest Golden Price Forecasting with Artificial Neural Networks</b>
<i>Faruk ALPASLAN Erol EĞRİOĞLU Çağdaş Hakan ALADAĞ Ebrucan TİRİNG</i>	<i>Faruk ALPASLAN Erol EĞRİOĞLU Çağdaş Hakan ALADAĞ Ebrucan TİRİNG</i>
<b>İlişkili Hatalara Sahip Modellerde Splayn ve Çekirdek Regresyon Kestiricilerinin Performansları</b>	<b>20 The Performances of Spline and Kernel Regression Estimators in Models with Correlated Errors</b>
<i>Serdar DEMİR Dursun AYDIN</i>	<i>Serdar DEMİR Dursun AYDIN</i>
<b>Yüksek Öğretimde Öğrenci İşleri Hizmet Kalitesinin Ölçümü ve Bir Örnek Uygulama</b>	<b>29 Measuring the Service Quality of Student Affairs at Higher Education and an Exemplary Application</b>
<i>Mesut KUMRU Pınar YILDIZ KUMRU</i>	<i>Mesut KUMRU Pınar YILDIZ KUMRU</i>

- Neyman, Pólya-Aeppli ve Thomas Dağılımlarının Momentleri Üzerine Bir İnceleme** 41 **An Investigation on the Moments of Neyman, Pólya-Aeppli and Thomas Distributions**  
*Gamze ÖZEL* *Gamze ÖZEL*
- Bireylerin Kredi Kartlarını İptal Ettirme Niyetlerinin İncelenmesinde Lojistik Regresyon Analizi** 48 **Logistic Regression Analysis in Examining the Users' Intention for the Cancellation of Credit Cards**  
*Gültekin ATALIK* *Gültekin ATALIK*  
*Y. Murat BULUT* *Y. Murat BULUT*  
*Zeki YILDIZ* *Zeki YILDIZ*
- Lojistik Regresyon ve Doğrusal Diskriminant Analizlerinde Kullanılan Bazı İndekslerin Karşılaştırılması** 54 **A Comparison of Some Indexes Used in Logistic Regression and Linear Discriminant Analysis**  
*Atilla GÖKTAŞ* *Atilla GÖKTAŞ*  
*Barış KESKİN* *Barış KESKİN*  
*Selen ÇAKMAKYAPAN* *Selen ÇAKMAKYAPAN*
- Tamam Olmayan Piyasada Tek Periyot Karesel Riskten Koruma Probleminde Parametrelerin Belirsizliği Altında Dayanıklı Optimizasyon Yaklaşımı** 65 **Robust Optimization Approach for One Period Quadratic Option Hedging Problem Under the Uncertainty of Parameters in Incomplete Market**  
*Gültaç EROĞLU İNAN* *Gültaç EROĞLU İNAN*  
*Ayşen APAYDIN* *Ayşen APAYDIN*  
*Mustafa Ç. PINAR* *Mustafa Ç. PINAR*
- Bulanık Zaman Serileri ile Ankara Hava Kalitesi Verisinin Çözümlemesi** 77 **Analysis of Ankara Air Quality Data with Fuzzy Time Series**  
*Sibel ALADAĞ* *Sibel ALADAĞ*  
*Çağdaş Hakan ALADAĞ* *Çağdaş Hakan ALADAĞ*  
*Erol EĞRİOĞLU* *Erol EĞRİOĞLU*
- Bulanık Yerel Polinomiyal Parametrik Olmayan Regresyon Modelleri** 84 **Fuzzy Local Polynomial Nonparametric Regression Models**  
*Memmedağa MEMMEDLİ* *Memmedağa MEMMEDLİ*  
*Münevvere YILDIZ* *Münevvere YILDIZ*  
*Özer ÖZDEMİR* *Özer ÖZDEMİR*

- Sayısal Görüntülerdeki Nesnelerin Baskın Noktalarının Algılanmasında İstatistiksel Yaklaşımlar** 93 **Statistical Approaches for the Detection of Dominant Points of Objects in Digital Images**
- Orhan KESEMEN* *Orhan KESEMEN*  
*Çiğdem GÜNGÖR* *Çiğdem GÜNGÖR*
- Üniversite Öğrencilerinin Kan Bağışı Hakkındaki Tutumlarının Lojistik Regresyon Analizi ile İncelenmesi** 101 **Evaluation of the Attitude of University Students on Blood Donation by Logistic Regression**
- Barış ERGÜL* *Barış ERGÜL*  
*Zeynep FİLİZ* *Zeynep FİLİZ*



**AMAÇ VE KAPSAM**

"İstatistik Araştırma Dergisi (İAD)", istatistik araştırmaların niteliğinin yükseltilmesi, istatistik yöntem ve uygulamalarının geliştirilmesi, literatürde yer alan çalışmaların tartışılması, istatistik uygulamalarıyla ilgili anket çalışmalarının ele alınması, kuramsal ve uygulama alanındaki araştırmacılar arasında iletişimin ortak çalışma ve yayınlarla güçlendirilmesi amacıyla, yayımlanan hakemli bir dergidir.

"İstatistik Araştırma Dergisi"nin kapsamında yer alan tematik konular aşağıda özet olarak verilmiştir:

- Bankacılık, Finans, Sigortacılık, Aktüerya ve Risk Yönetimi; Bayesci İstatistik; Benzetim Teknikleri; Bilgi Sistemleri; Biyoistatistik; Bulanık Teori; Demografi; Dency Tasarımı ve Varyans Analizi; Ekonometri; Genel Sayımlar ve Değerlendirmeleri; İstatistik Eğitimi; İstatistik Etiği; İstatistik Kuramı; İstatistiksel Kalite Kontrolü; Kamuoyu ve Piyasa Araştırmaları; Klinik Denemeler; Mühendislikte İstatistik Uygulamaları; Olasılık ve Stokastik Süreçler; Optimizasyon; Örnekleme ve Araştırma Tasarımları; Parametrik Olmayan İstatistiksel Yöntemler; Resmi İstatistikler; Toplum Bilimlerinde İstatistik; Veri Analizi ve Modelleme; Veri Madenciliği; Veri Yönetimi ve Karar Destek Sistemleri; Verimlilikte İstatistiksel Yaklaşımlar; Yönetim Süreçlerinde Performans Analizi; Yöneylem Araştırması; Zaman Serileri; Diğer İstatistiksel Yöntemler gibi istatistiğin her dalında yeni bilgi üretimine yönelik tüm araştırmalar.

**Makale Dili ve Genel Kurallar**

- Bu yayının 5846 Sayılı Fikir ve Sanat Eserleri Kanunu'na göre her hakkı Başbakanlık Türkiye İstatistik Kurumu Başkanlığı'na aittir. Gerçek veya tüzel kişiler tarafından izinsiz çoğaltılamaz ve dağıtılamaz.
- Makale taslakları WORD yazım dilinde, Times New Roman yazı tipinde, 12 punto büyüklükte, satırlar arasında bir satır boşluk bırakılarak yazılmalı, şekil ve grafikler JPG dosyaları olarak hazırlanmalıdır.
- A4 sayfa boyutunda; soldan 3,5 cm, sağdan, yukarıdan ve aşağıdan 2,5 cm boşluk bırakılmalıdır.
- Ana bölüm başlıklarının tümü büyük harf, 12 punto büyüklükte, koyu, ortalı ve Arap rakamları ile numaralandırılarak; alt bölüm başlıklarında ise sadece kelimelerin baş harfleri büyük diğerleri küçük harfle, 12 punto büyüklükte, koyu, sola dayalı ve ana bölüm başlığına endeksli olarak Arap rakamları ile numaralandırılarak yazılmalıdır.
- Makale taslağı yazımında, okuyucunun, çalışmanın her aşamasını anlama ve değerlendirmesine olanak verecek bir anlatım ve plana uyulmalıdır.
- Anlatım olabildiğince sade, anlaşılabilir, öz ve kısa olmalıdır. Gereksiz tekrarlardan, desteklenmemiş ifadelerden ve konu ile doğrudan ilişkisi olmayan açıklamalardan kaçınılmalıdır.
- Yazımda çok genel ifadeler kullanılmamalıdır. Yargı veya kesinlik içeren ifadeler mutlaka verilere/ referanslara dayandırılmalıdır.
- Araştırmacı/araştırmacılar tarafından probleme, hangi kuramsal/kavramsal açıdan yaklaşıldığı, gerekçeleri ile birlikte belirtilmelidir.
- Kullanılan araştırma yönteminin seçilme gerekçesi açıklanmalıdır. Bütün veri toplama araçlarının geçerliliği ve güvenilirliği belirtilmelidir.
- Araştırma sonucunda elde edilen veriler bir bütünlük içinde sunulmalıdır.
- Sadece elde edilen verilere dayanan sonuçlar sunulmalıdır.
- Sonuçların yorumları, varsa, literatürdeki diğer kaynaklarla desteklenerek, değerlendirilmelidir.
- Yararlanılan kaynaklar, çalışmanın kapsamını yansıtacak zenginlik ve yeterlikte olmalıdır.
- Türkçe ve İngilizce özetler; çalışmanın amacı, yöntemi, kapsamı ve temel bulgularını içermelidir.

Ayrıntılı bilgi için, <http://www.tuik.gov.tr> adresinden "İstatistik Araştırma Dergisi Kılavuzu"na bakınız.

## AIM AND SCOPE

“*Journal of Statistical Research (JSR)*” is a refereed journal published with the aim to raise the quality of statistical researches, improve the statistical methodology and applications, discuss the studies included in literature, consider survey studies regarding the statistical application, and strengthen the communication between researchers in the field of theory and application by joint studies and publications.

The contents of the “*Journal of Statistical Research*” are summarized below:

- Researches aimed at producing new knowledge in every field of statistics such as Banking, Finance, Insurance Trade, Actuarial and Risk Management; Bayesian Statistics; Biostatistics; Clinic Tests; Data Analysis and Modeling; Data Management and Decision Support Systems; Data Mining; Demography; Econometrics; Experimental Design and Variance Analysis; Fuzzy Theory; General Census and Evaluation; Information Systems; Non-Parametric Statistical Methods; Official Statistics; Operational Research; Optimization; Sampling and Research Designs; Performance Analysis in Managerial Process; Probability and Stochastic Processes; Public Opinion and Market Researches; Statistical Applications in Engineering; Statistical Approaches in Efficiency; Statistical Ethics; Statistical Quality Control; Statistical Training; Statistics in Social Science; Statistics Theory; Simulation Techniques; Time Series; Other Statistical Methods.

## Article Language and General Rules

- Prime Ministry, Turkish Statistical Institute reserves all the rights of this publication. Unauthorized duplication and distribution of this publication is prohibited under Law No: 5846.
- Article drafts should be prepared in WORD, using Times New Roman font, in 12 point size, with a blank line in between lines. Figures and tables should be prepared as JPG files.
- On A4 paper size; margins should be set as: left 3,5 cm; right, top and bottom 2,5 cm.
- Titles of the main sections should be all capitalized, in 12 point size, bold, centered and numbered with Arabic numerals; only the first letter of the words in the titles of the subsections should be capitalized, with 12 point size, bold, left justified and numbered with Arabic numerals indexed to the titles of the main sections.
- In article draft writing, writer should follow such a plan that reader should be able to understand and evaluate all the steps of the study.
- Narration should be as plain as possible, as well as comprehensible, compact and short. Unnecessary repetitions, unsupported declarations and explanations that are not in direct relation to the topic should be avoided.
- General statements should be avoided in writing. Statements that include judgment or facts must be supported by data/references.
- It should be stated, with justifications, from which theoretical/conceptual aspect the researcher/researchers have approached the problem.
- The reason of choosing the research methodology that is used should be explained. The validity and reliability of all the data collection tools should be presented.
- Data obtained as the result of the research should be presented in unity.
- Results that only rely on the obtained data should be presented.
- The interpretation of the results should be supported and evaluated by the other resources, if any, in the literature.
- Used resources should be in good wealth and proficiency that reflect the scope of the study.
- Turkish and English abstracts should include the goal, methodology, scope and main findings of the study.

For detailed information, please see “A Guide for Journal of Statistical Research” at <http://www.tuik.gov.tr>.

**DERGİNİN BU SAYISINA BİLİMSEL KATKI SAĞLAYAN HAKEMLER**  
**REFEREES WHO PROVIDED SCIENTIFIC CONTRIBUTIONS FOR THIS**  
**VOLUME OF THE JOURNAL**

Prof. Dr	Aşır GENÇ	Selçuk Üniversitesi
Prof. Dr.	Ayşen APAYDIN	Ankara Üniversitesi
Prof. Dr.	Efendi NASİBOV	Dokuz Eylül Üniversitesi
Prof. Dr	Hüseyin TATLIDİL	Hacettepe Üniversitesi
Prof. Dr.	Münevver TURANLI	İstanbul Ticaret Üniversitesi
Prof. Dr.	Ömer L. GEBİZLİOĞLU	Kadir Has Üniversitesi
Prof. Dr	Semra ERBAŞ	Gazi Üniversitesi
Doç. Dr.	Aylin ANIL	Dokuz Eylül Üniversitesi
Doç. Dr	Coşkun KUŞ	Selçuk Üniversitesi
Doç. Dr	Dicle CENGİZ	İstanbul Ticaret Üniversitesi
Doç. Dr	Fatih TANK	Ankara Üniversitesi
Doç. Dr	Kamile ŞANLI KULA	Ahievran Üniversitesi
Doç. Dr	M. Fedai KAYA	Selçuk Üniversitesi
Doç. Dr.	Özlem EGE ORUÇ	Dokuz Eylül Üniversitesi
Doç. Dr.	Özlem Müge AYDIN	Başkent Üniversitesi
Doç. Dr.	Serpil CULA	Başkent Üniversitesi
Doç. Dr.	Sevgi YURT ÖNCEL	Kırıkkale Üniversitesi
Doç. Dr.	Sevil BACANLI	Hacettepe Üniversitesi
Doç. Dr.	Sevtap KESTEL	Orta Doğu Teknik Üniversitesi
Doç. Dr.	Türkan ERBAY DALKILIÇ	Karadeniz Teknik Üniversitesi
Doç. Dr	Zerrin AŞAN	Anadolu Üniversitesi
Doç. Dr	Zeynep FİLİZ	Eskişehir Osman Gazi Üniversitesi
Yrd. Doç. Dr.	Emel BAŞAR	Gazi Üniversitesi
Yrd. Doç. Dr.	Emel ÇANKAYA	Sinop Üniversitesi
Yrd. Doç. Dr.	Filiz YUVA KARDİYEN	Gazi Üniversitesi
Yrd. Doç. Dr.	İhsan KARABULUT	Ankara Üniversitesi
Yrd. Doç. Dr.	Kadir Özgür PEKER	Anadolu Üniversitesi
Yrd. Doç. Dr.	Nimet YAPICI PEHLİVAN	Selçuk Üniversitesi
Yrd. Doç. Dr.	Rukiye DAĞALP	Ankara Üniversitesi
Yrd. Doç. Dr.	Sevcan DEMİR ATALAY	Ege Üniversitesi
Yrd. Doç. Dr.	Yaprak ÖZDEMİR	Gazi Üniversitesi
Dr.	Sibel AÇIK KEMALOĞLU	Ankara Üniversitesi

## İSTANBUL TEMİZ SU TÜKETİMİNİN BULANIK ZAMAN SERİSİ YAKLAŞIMLARI İLE ÖNGÖRÜSÜ

Faruk ALPASLAN\* Özge CAĞCAĞ\*\* Damla İLTER\*\*\* Ufuk YOLCU\*\*\*\*

### ÖZET

*Su tüketiminin doğru öngörülmesi, iklim değişikliklerinin son derece yoğun hissedildiği günümüzde, kısıtlı su kaynaklarının planlanması ve yönetimi açısından büyük önem arz etmektedir. Literatürde, su tüketim öngörüsünde, klasik yöntemler kullanılırken, bulanık mantık ve yapay sinir ağları da son zamanlarda etkin bir şekilde kullanılan yöntemler arasındadır. Tahminde kullanılan klasik yöntemlerin birçok varsayım içermesi nedeniyle, son yıllarda etkin ve daha iyi sonuçlar elde etmek için bulanık mantık ve yapay sinir ağları gibi yöntemlerden yararlanılmaktadır. Bu çalışmada İstanbul su tüketiminin miktarı çeşitli bulanık zaman serisi yöntemleri ile öngörülmesi amaçlanarak elde edilen sonuçlar değerlendirilmiştir.*

**Anahtar Kelimeler:** Bulanık zaman serileri, Öngörü, Su tüketimi.

### 1. GİRİŞ

Dünya üzerinde yaşamın kaynağı olarak tanımlayabileceğimiz su, insanlık tarihi boyunca önemini korumuş hatta yerleşim merkezlerinin kurulmasında belirleyici bir etken olmuştur. Tamamen ikame edilemeyen bir kaynak olan su, yaşayan bütün canlılar için en önemli doğal kaynaklardan biridir. Diğer bir ifadeyle su; hayatın ve canlıların kaynağıdır. Özellikle son 20 yıl içerisinde artan insan nüfusu ve bunun sonucu olarak artan su talebi, küresel bir su krizini gündeme getirmiştir. Bu nedenle ekonomik, politik ve çevresel konulardaki mücadeleler ve çekişmeler çok daha yaygın ve ciddi boyutlara ulaşmıştır. Su kaynakları; miktar, kalite ve tüm diğer sektörel kullanımlar açısından birçok ciddi sorunla karşı karşıyadır. Tüm bunların önlenmesi için su tüketiminin doğru bir şekilde öngörülmesi en temel başlangıç olacaktır. Su tüketiminin nüfus, nem, sıcaklık, ekonomik ve ticari koşullar ve yağış gibi birçok parametre tarafından etkilendiği düşünülürse, su tüketiminin doğru öngörülmesi, iklim değişikliklerinin son derece yoğun hissedildiği bu dönemde, kısıtlı su kaynaklarının planlanması ve yönetimi açısından da büyük önem arz etmektedir.

Literatürde, su talebi tahmininde, regresyon, Box-Jenkins ve ARIMA gibi birçok yöntem kullanılmaktadır (Caiado, 2007), (Caiado, 2009), (Lian vd., 2008), (Roberto vd., 2004), (Thomas vd., 2004), (Thomas vd., 2006), (Thomas vd., 2007). Ancak, bu geleneksel yöntemlerin, doğrusal model, gözlem sayısı, normal dağılım gibi birçok varsayım içermesi nedeniyle, son yıllarda bulanık mantık (Altunkaynak vd., 2005); (Froukh, 2001) yapay sinir ağları (Bougadis, 2005), (Calvo vd., 2009), (Calvo vd., 2003), (Cutore vd., 2008), (Fırat vd., 2009) ve diğer bazı farklı yöntemlerden de

\*Prof. Dr., Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Samsun, e-posta: [falpas@omu.edu.tr](mailto:falpas@omu.edu.tr)

\*\*Dr., Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Samsun, e-posta: [ozgecagcag@yahoo.com](mailto:ozgecagcag@yahoo.com)

\*\*\*Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Samsun, e-posta: [damlailter@yahoo.com](mailto:damlailter@yahoo.com)

\*\*\*\*Yrd. Doç. Dr., Giresun Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Giresun, e-posta: [uyolcu@omu.edu.tr](mailto:uyolcu@omu.edu.tr)

(Mohamed vd., 2010), (Nieswiadomy vd., 1989), (Shang vd., 2006) sıklıkla yararlanılmaktadır. Ayrıca zaman serisi öngörüsünde son yıllarda sıklıkla kullanılan bulanık zaman serisi yöntemlerinin üstün öngörü performansı birçok çalışmada ortaya konmuştur (Alpaslan vd., 2011), (Egrioglu vd., 2009), (Uslu vd., 2010). Bu çalışmada da, İstanbul su tüketimi, bulanık zaman serisi yöntemleri ile öngörülmektedir. Böylece klasik zaman serisi yaklaşımlarındaki kısıtlamalara gerek duyulmamakta ve bulanık zaman serisi yaklaşımlarının üstün öngörü performansı ile daha doğru tahminler elde edilmesi amaçlanmaktadır.

Bulanık zaman serisi öngörü yöntemleri, üzerinde birçok çalışma yapılan, bulanıklaştırma, bulanık ilişkilerin belirlenmesi ve durulaştırma aşamalarından oluşur. Bulanıklaştırma aşamasında genellikle sübjektif olarak belirlenen evrensel küme parçalanmaları kullanılmakla birlikte, zaman serisinin mutlak ilk farklar serisine bağlı ortalamaya ve dağılıma dayalı yaklaşımların yanında, klasik optimizasyon, genetik algoritmalar, parçacık sürü optimizasyonu ve evrensel küme parçalanmasını kullanmayan bulanık kümeleme gibi sübjektif kararlara ihtiyaç duymayan yöntemlerin kullanıldığı çalışmalar da mevcuttur. Bulanık ilişkilerin belirlenmesinde ise, ilk çalışmalarda karmaşık matris işlemleri kullanılmış, bu çalışmaları bulanık mantık grup ilişki tablolarının ve son yıllarda da yapay sinir ağlarının kullanıldığı çalışmalar izlemiştir. Durulaştırma aşamasında hemen tüm çalışmalarda merkezileştirme yöntemi kullanılmaktadır.

Çalışmanın ikinci bölümünde, temel bulanık zaman serisi tanımları verilmiş, üçüncü bölümde ise uygulamada çözümlemesi yapılacak bulanık zaman serisi öngörü yöntemleri tanımlanmıştır. Dördüncü bölümde “İstanbul temiz su tüketim miktarı” zaman serisi literatürde kullanılan beş farklı bulanık zaman serisi yöntemiyle çözümlenerek uygulama sonuçları verilmiştir. Son bölümde ise elde edilen sonuçlar değerlendirilmiştir.

## 2. BULANIK ZAMAN SERİLERİ

Bulanık zaman serisi tanımı ilk olarak, Song ve Chissom tarafından yapılmıştır (Song vd., 1993). Bazı temel bulanık zaman serisi tanımları aşağıdaki gibi verilebilir.

$U = \{u_1, u_2, \dots, u_b\}$ , evrensel küme ve evrensel kümenin belirlenen bir sabit aralık uzunluğuna göre parçalanması ile elde edilen  $u_i$ 'ler alt aralıklar ve  $b$  alt aralık sayısı olmak üzere bulanık kümeler,

$$A_i = f_{A_i}(u_1)/u_1 + f_{A_i}(u_2)/u_2 + \dots + f_{A_i}(u_b)/u_b \quad (1)$$

şeklinde tanımlanır. Burada,  $f_{A_i}$ ,  $A_i$  bulanık kümesinin üyelik fonksiyonunu gösterir ve  $f_{A_i} : U \rightarrow [0,1]$  olacaktır.  $f_{A_i}(u_a)$  ise  $u_a$  alt aralığının  $A_i$  bulanık kümesine ait olmasının üyelik derecesidir ve  $1 \leq a \leq b$  olmak üzere  $f_{A_i}(u_a) \in [0,1]$  olacaktır.

**Tanım 1.**  $Y(t)(t = \dots, 0, 1, 2, \dots)$  reel değerli zaman serisine uygun evrensel küme tanımı ve alt aralıkların tespit edilmesinden sonra bunlara bağlı olarak elde edilen  $A_i$  bulanık

kümelerinden oluşan yeni zaman serisi,  $F(t)$  bulanık zaman serisi olarak adlandırılır [22].

**Tanım 2.** “\*” herhangi bir operatörü ve  $R$  bulanık ilişki matrisini göstermek üzere, eğer  $F(t)$  bulanık zaman serisi yalnızca bir gecikmeli  $F(t-1)$  bulanık zaman serisinden etkilenmekte ise,  $F(t)$  ile  $F(t-1)$  bulanık zaman serisi arasındaki bulanık ilişki,

$$F(t) = F(t-1) * R(t, t-1) \quad (2)$$

şeklinde ifade edilir ve birinci dereceden bulanık zaman serisi öngörü modeli olarak adlandırılır. Bu ilişki Song ve Chissom tarafından aşağıdaki gibi tanımlanmıştır (Song vd., 1993).

$$F(t-1) \rightarrow F(t) \quad (3)$$

Sonuç olarak  $F(t-1) = A_i$  ve  $F(t) = A_j$  olduğunda  $F(t)$  ile  $F(t-1)$  bulanık zaman serisi arasındaki bulanık ilişki, (4) ifadesi ile verilebilir.

$$A_i \rightarrow A_j \quad (4)$$

**Tanım 3.**  $F(t)$  bulanık zaman serisi olsun. Eğer  $F(t)$ ,  $F(t-1)$ ,  $F(t-2)$ , ...,  $F(t-n)$  gecikmeli bulanık zaman serilerinden etkileniyor ise bulanık mantık ilişki takipteki gibi ifade edilebilir ve  $n$ 'nci dereceden bulanık zaman serisi öngörü modeli olarak adlandırılır [10].

$$F(t-n), \dots, F(t-2), F(t-1) \rightarrow F(t) \quad (5)$$

### 3. UYGULAMADA KULLANILAN ZAMAN SERİSİ YÖNTEMLERİ

İlk olarak Song ve Chissom (Song vd., 1993) tarafından ortaya konan bulanık zaman serileri, son yıllarda üzerinde oldukça yoğun çalışılan bir alan haline gelmiştir. Bulanıklaştırma, bulanık ilişkilerin belirlenmesi ve durulaştırma gibi üç temel aşamadan oluşan bulanık zaman serilerinin bu üç aşaması üzerine de birçok yaklaşım önerilmiştir. Song ve Chissom ilk bulanık zaman serisi tanımını yapmasının yanında bulanık ilişkileri bulanık ilişki matrisini elde ettiği matris işlemleri ile gerçekleştirmektedir (Song vd., 1993). Chen ise, bulanık ilişkileri belirlemede hem birinci dereceden hem de yüksek dereceden yöntemlerde kullanılan bulanık mantık ilişki ve grup ilişki tablolarından yararlanmıştır (Chen, 1996), (Chen, 2002). Aladağ vd. ise bulanık ilişkilerin belirlenmesinde yapay sinir ağlarını (YSA) kullanmışlardır (Aladağ vd., 2009). Bulanık ilişkilerin belirlenmesi üzerine katkılar sağlayan bu çalışmaların yanında, bulanıklaştırma aşamasında, Huarng everensel küme parçalanmasında belirlenen aralık uzunluklarının yöntemin öngörü performansı üzerinde etkin bir rol oynadığını göstermiş ve aralık uzunluklarının ortalamaya ve dağılıma dayalı olarak belirlendiği iki yaklaşım ortaya koymuştur (Huarng, 2001). Bunun yanında Eğrioğlu vd. ise, aralıkların belirlenmesinde optimizasyonu temel alan bir yaklaşım önermişleridir (Eğrioğlu vd., 2010).

“İstanbul temiz su tüketim miktarı” zaman serisinin öngörülmesinde kullanılan yöntemler çözümleme aşamalarının bazılarında benzerlik gösterirken, bazılarında da farklı yaklaşımlar içermektedir. Bu bakımdan tüm yöntemlere ait tek bir algoritma farklı aşamaları belirtilerek aşağıdaki gibi verilebilir.

**ADIM 1:** Evrensel küme ve alt aralıklar tanımlanır. Veri setindeki en büyük ve en küçük değer sırasıyla  $D_{\min}$  ve  $D_{\max}$ , ayrıca  $D_1$  ve  $D_2$  keyfi pozitif sayı olmak üzere evrensel küme aşağıdaki gibi belirlenir.

$$U = [D_{\min} - D_1, D_{\max} + D_2]. \quad (6)$$

Bu adımda aralıkları Eğrioğlu vd. öngörü hatasını en küçük yapacak şekilde (Egrioglu vd., 2010), Huarng zaman serisinin mutlak ilk farklarına bağlı olarak ortalamaya ve dağılıma dayalı yaklaşımlarla (Huarng, 2001) belirlerken çözümlemesi yapılan diğer yöntemler subjektif olarak belirlemektedir (Aladag vd., 2009), (Chen, 1996), (Chen, 2002), (Huarng, 2001).

**ADIM 2:** Evrensel küme ve parçalanmalara bağlı olarak bulanık kümeler tanımlanır.

$$A_j = a_{j1} / u_1 + a_{j2} / u_2 + \dots + a_{jb} / u_b, \quad (7)$$

$$a_{jk} = \begin{cases} 1 & , k = j \\ 0.5 & , k = j-1, j+1 \\ 0 & , \text{aksi halde} \end{cases}, \quad j=1, \dots, b \quad (8)$$

**ADIM 3:** Gözlemler bulanıklaştırılır.

Her bir veri bulunduğu aralığın en büyük üyelik değerine sahip olduğu bulanık küme ile eşleştirilerek zaman serisi bulanıklaştırılır.

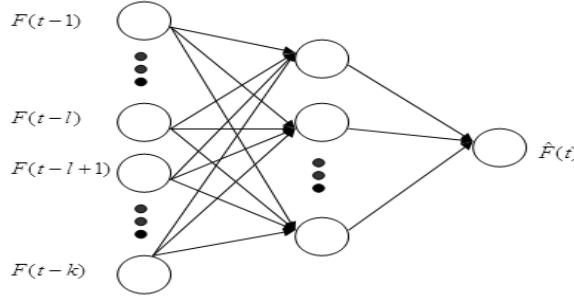
**ADIM 4:** Bulanık ilişkiler belirlenir.

- Song-Chissom matris işlemlerinden yararlanmıştır (Song vd., 1993). Bulanık ilişki  $A_k \rightarrow A_l$  olduğunda bu ilişkiye karşılık gelen ilişki matrisi  $R_i = A_k^T A_l$  ile elde edilir. Burada çarpım işleminde “min” operatörü kullanılır. Bulanık zaman serisi için birinci dereceden bulanık ilişkiyi ifade eden ilişki matrisi ise  $R = \cup_i R_i$  şeklinde elde edilir.
- Birinci dereceden bulanık zaman serisi öngörü modelinde, bulanık ilişkiler  $A_1 \rightarrow A_2, A_1 \rightarrow A_1, A_1 \rightarrow A_3$  şeklinde iken, bulanık mantık grup ilişki  $A_1 \rightarrow A_2, A_3, A_1$  şeklinde olmaktadır (Chen, 1996), (Huarng, 2001).
- Yüksek dereceli bulanık zaman serisi öngörü modelinde bulanık mantık ilişki ve grup ilişkiler sırasıyla (Chen, 1996) ve (Chen, 2002)’de verilmektedir.

$$\begin{matrix} A_{in}, A_{i(n-1)}, \dots, A_{i1} \rightarrow A_{j1} \\ \vdots & \vdots \\ A_{in}, A_{i(n-1)}, \dots, A_{i1} \rightarrow A_{jp} \end{matrix} \quad (9)$$

$$A_{in}, A_{i(n-1)}, \dots, A_{i1} \rightarrow A_{j1}, A_{j2}, \dots, A_{jp} \quad (10)$$

- Bulanık ilişkilerin YSA ile belirlendiği yüksek dereceli bulanık zaman serisi öngörü modelinde, YSA'nın girdileri bulanık gecikmeli değişkenlerden, çıktısı öngörüdür, hedef değeri ise gerçek verinin bulanık değerinden oluşur (bakınız Şekil 1) (Aladag vd., 2009).



Şekil 1. İleri beslemeli yapay sinir ağı mimarisi

**ADIM 5:** Bulanık öngörüler elde edilir.

- Song-Chissom, bulanık öngörüü  $\hat{F}(t) = \max\{\min\{F(t-1) \circ R\}\}$  ifadesi ile elde eder (Song vd., 1993).
- Chen (Chen, 1996) ve Huarng'ın (Huarng, 2001) birinci derece modelinde üç farklı durum söz konusudur;

*Durum 1.* Bulanık grup ilişki tablosundan sadece  $A_i \rightarrow A_j$  ilişkisi var ise bulanık öngörü  $A_j$ 'dir.

*Durum 2.* Bulanık grup ilişki tablosunda  $A_i \rightarrow A_i, A_j, \dots, A_k$  ise bulanık öngörü  $A_i, A_j, \dots, A_k$ 'dir.

*Durum 3.* Bulanık grup ilişki tablosunda  $A_i \rightarrow Boş$  ise bulanık öngörü  $A_i$ 'dir.

- Chen (Chen, 2002) ve Eğrioğlu vd.'nin [Eğrioğlu vd., 2010) yüksek dereceden modellerinde de üç durum vardır.

*Durum 1.* n'nci dereceden bulanık mantık grup ilişki tablosunda,  $A_{in}, A_{i(n-1)}, \dots, A_{i1} \rightarrow A_j$  ilişkisi var ise bulanık öngörü  $A_j$ 'dir.

*Durum 2.* n'nci dereceden bulanık mantık grup ilişki,  $A_{in}, A_{i(n-1)}, \dots, A_{i1} \rightarrow A_{j1}, A_{j2}, \dots, A_{jp}$  ise bulanık öngörü  $A_{j1}, A_{j2}, \dots, A_{jp}$  olur.

*Durum 3.* n'nci dereceden bulanık mantık grup ilişki tablosunda,  $A_{in}, A_{i(n-1)}, \dots, A_{i1} \rightarrow Boş$  ilişkisi mevcut ise reel öngörü,  $A_{in}, A_{i(n-1)}, \dots, A_{i1}$  bulanık



kümelerine bağlı olarak,  $u_{in}, u_{i(n-1)}, \dots, u_{i1}$  aralıklarının orta noktaları,  $m_{in}, m_{i(n-1)}, \dots, m_{i1}$  olmak üzere,

$$\frac{1 \times m_{in} + 2 \times m_{i(n-1)} + \dots + n \times m_{i1}}{1 + 2 + \dots + n} \quad (11)$$

ifadesi ile elde edilir.

- Aladağ vd. (Aladağ vd., 2009) yönteminde ise YSA'nın çıktı değeri  $A_j$  iken bulanık öngörü de  $A_j$  olacaktır.

**ADIM 6:** Durulaştırma.

Song-Chissom yönteminde üç farklı durumda gerçekleşir;  $\hat{F}(t) = \max\{\min\{F(t-1) \circ R\}\}$  ifadesi,

*Durum 1.* Tek bir “en büyük” üyeliğe sahipse duru öngörü bu üyeliğe karşılık gelen aralığın orta noktası olacaktır.

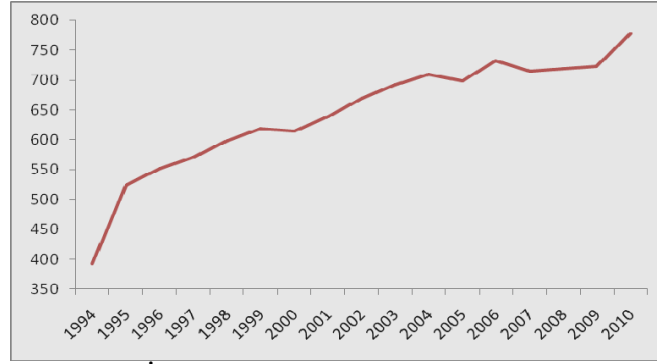
*Durum 2.* Birden fazla ve ardışık “en büyük” üyeliklere sahipse duru öngörü bu üyeliklere karşılık gelen aralıkların orta noktalarının ortalaması olacaktır.

*Durum 3.* Birden fazla ve ardışık olmayan “en büyük” üyeliklere sahipse duru öngörü tüm üyeliklerin standartlaştırılması ile elde edilir (Song vd., 1993).

Diğer yöntemlerde, merkezileştirme yöntemi kullanılır. Bulanık öngörü  $A_j$  ise, durulaştırılmış öngörü,  $A_j$  bulanık kümesinde en yüksek üyelik değerine sahip  $u_j$  aralığının orta noktası olacaktır. Bulanık öngörü  $A_{j1}, A_{j2}, \dots, A_{jp}$  ise duru öngörü bu bulanık kümelerde en yüksek üyelik değerine sahip aralıkların orta noktalarının ortalaması olacaktır.

#### 4. UYGULAMA

Bu çalışmada, grafiği Şekil 2’de verilen 1994-2010 yıllarına ait “İstanbul temiz su tüketim” verisinin öngörülmesi amaçlanmıştır. Bu amaçla beş farklı bulanık zaman serisi yöntemi ile çözümlene yapılmıştır. Çözümlemede zaman serisinin son 3 yıla ait verisi test kümesi olarak kullanılarak bu yıllara ait öngörüler elde edilmiştir.



Şekil 2. İstanbul temiz su tüketim miktarı zaman serisi

Uygulamada;

- Veri setindeki en büyük ve en küçük değer sırasıyla  $D_{\min} = 392$  ve  $D_{\max} = 778$  olarak tespit edildi.  $D_1$  ve  $D_2$  keyfi pozitif sayıları sırasıyla 42 ve 22 seçilerek evrensel küme  $U = [350, 800]$  olarak belirlendi.
- Eğrioğlu vd. yönteminde aralık uzunluğu (30,50) aralığında optimize edilirken, diğer yöntemlerde 30,35,40,45 ve 50 olarak alındı (Eğrioğlu vd., 2010).
- Huarng yönteminde ortalamaya ve dağılıma dayalı yaklaşımda aralık uzunluğu sırasıyla 10 ve 20 olarak belirlendi (Huarng, 2001).
- Aladağ vd., Chen ve Eğrioğlu vd. yüksek dereceli bulanık zaman serisi yöntemlerinde model derecesi 2- 5 arasında alınmıştır (Aladağ vd., 2009); (Chen, 2002) ; (Eğrioğlu vd., 2010).
- Aladağ vd. yönteminde gizli tabaka birim sayısı, ağın genelleştirme yeteneğini kaybetmemesi için, 1-5 aralığında sınırlandırılmıştır (Aladağ vd., 2009).
- Böylece Song-Chissom ( Song vd., 1993) yönteminde 5, Chen (Chen, 1996) yönteminde 5, Huarng (Huarng, 2001) yönteminde 2, Chen (Chen, 2002) yüksek derece yönteminde 20, Eğrioğlu vd. (Eğrioğlu vd., 2010) yönteminde 4 ve Aladağ vd. (Aladağ vd., 2009) yönteminde 100 olmak üzere toplam 136 farklı çözümlenme yapılmıştır.

Çözömlenmeler sonucunda yöntemlere ait en iyi durumlar Tablo 1'de özetlenmiştir. Elde edilen sonuçlar hata kareler ortalaması karekökü (HKOK) ve ortalama mutlak yüzdellik hata (OMYH) kriterlerine göre değerlendirilmiştir.

$$HKOK = \sqrt{\frac{\sum_{t=1}^n (X_t - \hat{X}_t)^2}{n}} \quad (12)$$

$$OMYH = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - \hat{X}_t}{X_t} \right| \quad (13)$$

Burada  $X_t$ ,  $t$  anına ait gerçek değeri,  $\hat{X}_t$ ,  $t$  anına ait tahmin değeri,  $n$  ise gözlem sayısını ifade etmektedir.

Tablo 1. Yöntemlerin en iyi sonuçları

	Model	Aralık	GTBS*	HKOK	OMYH
	Derecesi	Uzunluğu			
Song – Chissom [22]	1	45,00	-	28,1114	0,0306
Chen [9]	1	45,00	-	28,1114	0,0306
Huarng [6]	1	10,00	-	30,9946	0,0277
Chen [10]	4	20,00	-	33,5509	0,0271
Chen [10]	4	45,00	-	26,4339	0,0223
Eğrioğlu vd. [12]	2	42,44	-	15,1483	0,0162
Aladağ vd. [1]	4	45,00	3	10,0125	0,0113

\*GTBS: Gizli Tabaka Birim Sayısı

Tablo 1 incelendiğinde en iyi öngörü performansının Aladağ vd. (Aladağ vd., 2009) tarafından önerilen yöntemde 10,0125 HKOK ve 0,0113 OMYH değerleri ile elde edilmiştir. Bunun yanında tüm yöntem sonuçlarının oldukça düşük hatalar içerdiği ve dolayısıyla tatmin edici bir öngörü performansına sahip oldukları söylenebilir.

## 5. SONUÇLAR VE TARTIŞMA

İnsanların hayatlarını sürdürebilmeleri için gerekli en temel şey olan su tarih boyunca önemini korumuş ve eksikliği insanlık için var olabilecek bir sorun haline dönüşmüştür. Kişi başına düşen günlük su miktarı ülkeden ülkeye değişmektedir. Örneğin bu miktar Almanya için 620 litre iken Türkiye’de 90-136 litre arasında değişmektedir. Afrika ve Orta doğu ülkelerinin birçoğunda bu miktar, ancak 2-3 litredir. Dünyada 80 ülkede yaşayan yaklaşık 2 milyar insan, kurak mevsimde bu kadarını da bulamamaktadır. 1998’de 28 ülke su kıtlığı çekmişken bu sayının 2025’te 56’ya yükselmesi beklenmektedir. Tüm bu rakamlar dikkate alındığında su tüketiminin doğru öngörülmesi, kısıtlı su kaynaklarının planlanması ve yönetimi açısından da büyük önem arz etmekle beraber gerekenlerin yapılması açısından da oldukça önemlidir.

Bu çalışmada, “İstanbul temiz su tüketim miktarı” bazı bulanık zaman serisi yöntemleri ile öngörülmüş ve oldukça etkileyici sonuçlar elde edilmiştir. Sonuçlar incelendiğinde öngörü hatasının %1 ile %3 arasında olduğu görülür. Böylece bulanık zaman serisi öngörü yöntemlerinin, son yıllarda doğru öngörülmesi oldukça önem kazanmış “İstanbul temiz su tüketim miktarı” zaman serisi için oldukça etkileyici bir öngörü performansı sergilediği söylenebilir.

## 6. KAYNAKLAR

Aladag, C. H., Basaran, M. A., Egrioglu, E., Yolcu, U., Uslu, V. R., 2009. Forecasting in High Order Fuzzy Times Series by Using Neural Networks to Define Fuzzy Relations, Expert Systems with Applications, 36, 4228-4231.

Alpaslan, F., Cagcag, O., Aladag, C. H., Yolcu, U., Egrioglu, E., 2011. A Novel Seasonal Fuzzy Time Series Method, FUZZYSS'11: The Second International Fuzzy

Systems Symposium, Proceeding Book, Editors: C. Gokceoglu, H. C. Aladag, A. Akgun, Page: 50-55.

Altunkaynak, A., Ozger, M., Cakmakçı, M., 2005. Water Consumption Prediction of Istanbul City by Using Fuzzy Logic Approach, *Water Resources Management*, 19, 641-654.

Bougadis, J., Adamowski, K., Diduch, R., 2005. Short-term Municipal Water Demand Forecasting, *Hydrological Processes*, 19, 137-148.

Caiado, J., 2007. Forecasting Water Consumption in Spain Using Univariate Time Series Models, Munich Personal RePEc Archive (Online at, <http://mpra.ub.uni-muenchen.de/6610/>).

Caiado, J., 2009. Performance of Combined Double Seasonal Univariate Time Series Models for Forecasting Water Consumption, Munich Personal RePEc Archive, (Online at <http://mpra.ub.uni-muenchen.de/15242/>).

Calvo, I. P., Gutierrez-Estrada, J. C., 2009. Improved Irrigation Water Demand Forecasting Using a Soft-Computing Hybrid Model, *Biosystems Engineering*, 102, 202-218.

Calvo, I. P., Roldan, J., Lopez-Luque, R., Gutierrez-Estrada, J. C., 2003. Demand Forecasting for Irrigation Water Distribution Systems, *Journal of Irrigation and Drainage Engineering*, 129(6), 422-431.

Chen, S. M., 1996. Forecasting Enrollments Based on Fuzzy Time-Series, *Fuzzy Sets and Systems*, 81, 311-319.

Chen, S. M., 2002. Forecasting Enrollments Based on High Order Fuzzy Time Series, *Cybernetics and Systems*, 33:1-16.

Cutore, P., Campisano, A., Kapelan, A., Modica, C., Savic, D., 2008. Probabilistic prediction of urban water consumption using the SCEM-UA algorithm, *Urban Water Journal*, 5(2), 125-132.

Egrioglu, E., Aladag, C. H., Yolcu, U., Uslu, V. R., Basaran, M. A., 2010. Finding an Optimal Interval Length in High Order Fuzzy Time Series, *Expert Systems with Applications*, 37, 5052-5055.

Egrioglu, E., Aladag, C. H., Yolcu, U., Basaran, M. A and Uslu, V. R., 2009. A New Hybrid Approach Based on SARIMA and Partial High Order Bivariate Fuzzy Time Series Forecasting Model. *Expert Systems with Applications*, 36, 7424-7434.

Firat, M., Yurdusev, M. A., Turan, M. E., 2009. Evaluation of Artificial Neural Network Techniques for Municipal Water Consumption Modeling, *Water Resour Manage*, 23, 617-632.

- Froukh, M. L., 2001. Decision-Support System for Domestic Water Demand Forecasting and Management, *Water Resources Management*, 15, 363-382.
- Huang, K., 2001. Effective Length of Intervals to Improve Forecasting in Fuzzy Time Series, *Fuzzy Sets and Systems*, 123, 387-394.
- Lian, T. H., Liu, Q. J., Wang, J. C., 2008. Water Demand Forecast Based on ARIMA Time-series Identification, *Control Engineering of China*, 81.
- Mohamed, M. M., Al-Mualla, A. A., 2010. Water Demand Forecasting in Umm Al-Quwain (UAE) Using the IWR-MAIN Specify Forecasting Model, *Water Resource Manage*, 44, 4093-4120.
- Nieswiadomy, M. L., Molina, D. J., 1989. Comparing Residential Water Demand Estimates Under Decreasing and Increasing Block Rates Using Household Data, *Land Economics*, 65(3), 280-289.
- Roberto, M. E., Céline, N., 2004. Is All Domestic Water Consumption Sensitive to Price Control, *Applied*, 36, 1697-1703.
- Shang, F., Uber, J. G., Waanders, B. G. B., Boccelli, D., Janke, R., 2006. Real Time Water Demand Estimation in Water Distribution System, 8th Annual Water Distribution Systems Analysis Symposium, Cincinnati, Ohio, USA, August 27-30.
- Song, Q., Chissom, B. S., 1993. Fuzzy Time Series and Its Models, *Fuzzy Sets and Systems*, 54, 269-277.
- Thomas, M., Fullerton, J., Arturo, E., 2004. Short-term Water Consumption Dynamics in El Paso, Texas, *Water Resources Research*, 40, doi: 10.1029/2004WR00326.
- Thomas, M. F. J., Roberto, T., Jorge, E. M. C., 2007. An Empirical Analysis of Tijuana Water Consumption, *Atl. Econ J*, 35, 357-369.
- Thomas, M. F. J., Roberto, T., Martha, P. B., 2006. Short-Term Water Consumption Patterns in Ciudad Juárez, Mexico, *Atlantic Economic Journal*, 34, 467-479.
- Uslu, V. R., Aladag, C. H., Yolcu, U., Egrioglu, E., 2010. A new hybrid approach for forecasting a seasonal fuzzy time series. 1<sup>st</sup> International Symposium On Computing In Science & Engineering, Izmir, Turkey.

## FORECASTING OF ISTANBUL'S CLEAN WATER CONSUMPTION WITH FUZZY TIME SERIES APPROACHES

### ABSTRACT

*Accurate forecasting of water consumption is very important for planning and managing water sources during now that the global warming and climate change have distinctly appeared. While the classic methods have been frequently used for the forecasting of water consumption in the literature, fuzzy logic and artificial neural networks have also been among the efficiently used methods over the last years. Since conventional estimation methods require several assumptions, methods such as fuzzy logic and artificial networks are utilized to obtain more efficient and better results in the recent years. In this study, forecasting the quantity of İstanbul water consumption with various fuzzy time series approaches is aimed, and the results are evaluated.*

**Keywords:** Fuzzy time series, Forecasting, Water consumption.

## EN DÜŞÜK VE EN YÜKSEK ALTIN FİYATLARININ YAPAY SİNİR AĞLARI İLE ÖNGÖRÜSÜ

Faruk ALPASLAN\* Erol EĞRİOĞLU\*\* Çağdaş Hakan ALADAĞ\*\*\*  
Ebrucan TİRİNG\*\*\*\*

### ÖZET

Gözlemleri gün içinde değişen borsa indeksi, altın fiyatları veya döviz kuru gibi zaman serilerinin yapay sinir ağları ile çözümlenmesi bu çalışmanın temel aldığı problemdir. Altın fiyatları gibi zaman serileri gün içinde değişen aralık değerlere sahip olduğundan klasik zaman serileri analizi ile bu tür verilerin çözümlenmesinde en düşük veya en yüksek değerlerden oluşan zaman serileri ayrı olarak modellenmektedir. Belirli bir gün için altın fiyatı en düşük ve en yüksek değer arasında değişmektedir. Bu çalışmada, altın fiyatı gibi zaman serilerinin, yapay sinir ağı ile en düşük ve en yüksek değerlerinin tahmin edilmesi için üç farklı yaklaşım karşılaştırılmıştır. İlk yaklaşım en düşük ve en yüksek değerlerin ayrı yapay sinir ağları ile çözümlenmesi, ikinci yaklaşım ise en düşük ve en yüksek değerlerin aynı yapay sinir ağı üzerinde çözümlenmesi yaklaşımıdır. Üçüncü yaklaşım ise merkez ve açıklıklara dayalı yaklaşımıdır. Yaklaşımlar T.C. Merkez bankası internet sitesinden (EVDS) alınan 30/03/2011 ve 30/03/2012 tarihleri arasındaki günlük altın fiyatları zaman serisine uygulanmıştır. Uygulama sonucunda elde edilen sonuçlar tartışılmıştır.

**Anahtar Kelimeler:** Aralık değerli zaman serisi, Öngörü, Yapay sinir ağları.

### 1. GİRİŞ

Son yıllarda zaman serisi öngörüsü için yapay sinir ağlarına olan ilgi gittikçe artmaktadır. Yapay sinir ağlarının normallik, doğrusallık gibi zor kısıtlara sahip olmaması uygulamada daha çok tercih edilmesine sebep olmaktadır. Yapay sinir ağları öngörü probleminin çözümünde birçok avantajına rağmen göz ardı edilemeyecek problemlere de sahiptir. Yapay sinir ağları ile zaman serisi öngörüsü üzerine yapılan çalışmalar (Zhang vd., 1998) tarafından özetlenmiştir. Özellikle yapay sinir ağının mimari yapısının belirlenmesi ciddi bir tartışma alanıdır. Öngörü probleminde yapay sinir ağının mimari seçim problemi için (Eğrioğlu vd., 2008); (Eğrioğlu vd., 2008); (Aladağ vd., 2010) çalışmalarında model seçim ölçütü ve stratejileri önermişlerdir. Mimari seçim probleminin tartışıldığı diğer bir çalışma ise (Alpaslan vd., 2010)'dir. Mimari belirleme kuşkusuz ilgilenilen öngörü problemine uygun olmak zorundadır. Bu çalışmada gözlemleri gün içinde değişen borsa indeksi, altın fiyatları veya döviz kuru gibi zaman serilerinin yapay sinir ağları ile çözümlenmesi tartışılmıştır. Yapay

\*Prof. Dr., Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Samsun, e-posta: [falpas@omu.edu.tr](mailto:falpas@omu.edu.tr)

\*\*Doç. Dr., Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Samsun, e-posta: [erole@omu.edu.tr](mailto:erole@omu.edu.tr)

\*\*\*Doç. Dr., Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara, e-posta: [aladag@hacettepe.edu.tr](mailto:aladag@hacettepe.edu.tr)

\*\*\*\*Arş. Gör., Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Samsun, e-posta: [cbrutiring@hotmail.com](mailto:cbrutiring@hotmail.com)

sinir ağları ile zaman serisi öngörüsü literatürde yaygın olarak uygulanmıştır. Gözlemleri gün içinde değişen borsa indeksi, altın fiyatları veya döviz kuru gibi zaman serilerinin yapay sinir ağları ile çözümlenmesi problemi literatürde yeni bir konudur. Altın fiyatları gibi zaman serileri gün içinde değişen değerlere sahip olduğundan, klasik zaman serileri analizinde açılış, kapanış, en düşük veya en yüksek değerlerden oluşan zaman serileri ayrı olarak modellenmektedir. Belirli bir gün için altın fiyatı en düşük ve en yüksek değer arasında değişmektedir. Bu çalışmada, altın fiyatı gibi zaman serilerinin, yapay sinir ağı ile en düşük ve en yüksek değerlerinin tahmin edilmesi için üç farklı yaklaşım karşılaştırılmıştır. İlk yaklaşım en düşük ve en yüksek değerlerin ayrı yapay sinir ağları ile çözümlenmesi, ikinci yaklaşım ise en düşük ve en yüksek değerlerin aynı yapay sinir ağı üzerinde çözümlenmesi yaklaşımıdır. Üçüncü yaklaşım ise merkez ve açıklıklara dayalı yaklaşımıdır. Çalışmanın ikinci bölümünde ileri beslemeli yapay sinir ağları ile zaman serisi öngörüsünün nasıl yapılacağı özetlenmiştir. Üçüncü bölümde aralık verilerinin yapay sinir ağları ile çözümlenmesi için kullanılan üç yaklaşım tanıtılmıştır. Dördüncü bölümde altın fiyatları zaman serisi üzerinde üçüncü bölümde verilen yöntemler uygulanmıştır. Son bölümde ise elde edilen sonuçlar tartışılmıştır.

## 2. İLERİ BESLEMELİ YAPAY SİNİR AĞLARI İLE ZAMAN SERİSİ ÖNGÖRÜSÜ

Son yıllarda literatürdeki zaman serisi uygulamalarının büyük çoğunluğunda yapay sinir ağları kullanılmaktadır. Zaman serilerinde, yapay sinir ağları yönteminin kullanımının yaygınlaşmasına aşağıda sıralanan üstünlükleri neden olmaktadır.

- Zaman serisinin eğrisel ya da doğrusal yapıya sahip olup olmadığı test edilmeksizin çözümlenebilir.
- Geleneksel zaman serisi yöntemlerinden daha iyi öngörü sonuçları elde edilebilmektedir.
- Geleneksel eğrisel zaman serisi modelleri sadece belli eğrisel yapılar için kullanıldığından genel olarak yeterli esnekliğe sahip değildir. Ancak yapay sinir ağları ile çözümlenmede zaman serisinin eğrisel yapısının nasıl olduğu önemli değildir.
- YSA yöntemi, geleneksel zaman serisi tahmin yöntemleri gibi karmaşık bir teoriye sahip değildir, anlaşılması daha kolaydır.

Yapay sinir ağları yöntemiyle zaman serilerinde öngörü elde etmeyi, 7 adımda özetlemek mümkündür (Günay vd., 2007).

**Adım 1:** Aktivasyon fonksiyonunun seçimi ve verinin ön işlenmesi.

İlk olarak yapay sinir ağının gizli tabakası ve çıktı tabakasında kullanılacak aktivasyon fonksiyonunun türüne karar verilir. Bu çalışmanın uygulama kısmında gizli tabaka birimlerinde lojistik aktivasyon fonksiyonu kullanılmıştır. Lojistik aktivasyon fonksiyonu formül 1’ de verildiği gibidir. Literatürde lojistik aktivasyon fonksiyonunun öngörü problemlerinde başarılı sonuçlar ürettiği bilinmektedir (Eğrioğlu vd., 2008).

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (1)$$



Verilerin kullanılan aktivasyon fonksiyonuna uygun olacak aralığa dönüştürülmesi gerçekleştirilir. Eğer lojistik aktivasyon fonksiyonu kullanılacaksa,  $x_i$ , girdi değerlerini göstermek üzere

$$x'_i = \frac{x_i - \text{Min}(x_i)}{\text{Maks}(x_i) - \text{Min}(x_i)} \quad (2)$$

şeklinde girdi değerleri (0,1) aralığına dönüştürülebilir. Burada,  $\text{Maks}(x_i)$ , girdi değerlerinin en büyüğünü,  $\text{Min}(x_i)$ , girdi değerlerinin en küçüğünü göstermektedir.

**Adım 2:** Veri organizasyonu: Eğitim ve test kümelerinin büyüklüğünün veri kümesinin yüzde kaç olacağına karar verilir.

**Adım 3:** Modelleme: Girdi sayısı, gizli tabaka sayısı, gizli tabakadaki birim sayısı ve çıktı tabakasındaki birim sayısı, bu birimlerde kullanılacak aktivasyon fonksiyonu, öğrenme algoritması ve bu algoritmanın parametreleri ve performans ölçütü belirlenerek, kullanılacak yapay sinir ağı modeli kurulur.

**Adım 4:** Girdi değerlerinin oluşturulması: Yapay sinir ağının girdi değerleri gecikmeli zaman serileridir.  $X_t$ , zaman serisi için girdi değerleri oluşturulurken, girdi tabakasındaki birim sayısı  $m$  ile gösterilsin,  $m$  tane gecikmeli zaman serisi  $X_{t-1}, X_{t-2}, \dots, X_{t-m}$ , şeklinde oluşturulur.

**Adım 5:** En iyi ağırlık değerlerinin hesaplanması: Eğitim kümesi üzerinden, seçilen öğrenme algoritması ile en iyi ağırlık değerleri bulunur. Elde edilen en iyi ağırlık değerleri kullanılarak, kurulan yapay sinir ağı modelinin çıktı değerleri hesaplanır.

**Adım 6:** Performans ölçütünün hesaplanması: Yapay sinir ağının test kümesi tahminleri elde edilir. Adım 5'te elde edilen çıktı değerlerine ve bu adımda elde edilen değerlere, Adım 1'de uygulanan dönüşümün tersi uygulanır. Bu dönüşüm sonucu elde edilen değerler sırasıyla, eğitim kümesinin tahminlerini ve test kümesinin tahminlerini oluşturur. Test kümesinin tahminleri ile test kümesindeki verilerin arasındaki farka dayalı olarak, seçilen performans ölçütü hesaplanır. Yapay sinir ağının verilerdeki ilişkiyi ne kadar iyi öğrendiğini belirlemek için kullanılan birçok performans ölçütü vardır. En çok kullanılan performans ölçütlerinden biri hata kareler ortalaması karekök (RMSE), diğeri yön doğruluğu (DA)'dır.  $z_p$ , yapay sinir ağı çıktı değerini,  $d_p$ , hedef çıktı değerini ve  $p$ , test kümesinin büyüklüğünü göstermek üzere,

$$RMSE = \sqrt{\frac{\sum_{p=1}^P (d_p - z_p)^2}{P}} \quad (3)$$

$$DA = \frac{1}{T} \sum_{i=1}^T a_i, \quad a_i = \begin{cases} 1 & \text{if } (d_{i+1} - d_i)(z_{i+1} - d_i) > 0 \\ 0 & \text{Aksi halde} \end{cases} \quad (4)$$

eşitliği ile hesaplanmaktadır.

**Adım 7:** Öngörü: Son olarak, Adım 5'te bulunan en iyi ağırlık değerleri kullanılarak, test kümesinden sonraki zamanlar için yani, gelecek için öngörü değerleri elde edilir.

### 3. ARALIK VERİLERİNİN ÇÖZÜMLENMESİ İÇİN YAPAY SINIR AĞLARI YAKLAŞIMLARI

Aralık verilerinin çözümlenmesi literatürde yeni araştırılan bir konudur. Aralık verilerinin yapay sinir ağları ile çözümlenmesinde alternatif yaklaşımlar ortaya atılmıştır. (Lima Neto vd., 2008); (Billard vd., 2000); (Maia vd., 2010); (Maia vd., 2008) aralık değerli tahmin çalışmalarından bazılarıdır. Bu bölümde, çalışmada uygulanan üç farklı aralık değerli zaman serisi yaklaşımı özetlenmiştir.

**Yaklaşım 1.** Aralık verilerin çözümlenmesi için (Maia vd., 2008) çalışmasında aralık zaman serisinin alt ve üst sınır değerlerinin ( $XU_t$  - Üst sınır değeri,  $XL_t$  - Alt sınır değeri) yapay sinir ağları ile ayrı ayrı çözümlenmesini gerçekleştirmiştir. Bu yöntemde alt ve üst sınır değerleri için tahminler aşağıdaki modellerden elde edilir.

$$X\hat{U}_t = f_1(XU_{t-1}, \dots, XU_{t-m}), \quad X\hat{L}_t = f_2(XL_{t-1}, \dots, XL_{t-m}) \quad (5)$$

Burada,  $f_1$  ve  $f_2$  sırasıyla alt sınır ve üst sınır değerleri zaman serilerinin gecikmeli değişkenlerinin doğrusal olmayan ve sinir ağından elde edilen fonksiyonlarıdır.

**Yaklaşım 2.** (Maia vd., 2008) çalışmasında aralık değerli yapay sinir ağının çözümlenmesinde zaman serisinin alt ve üst sınır değerlerinin yapay sinir ağları ile birlikte çözümlenmesi gerçekleştirilmiştir. Bu yöntemde tek bir yapay sinir ağı oluşturulmakta ve alt sınır üst sınır değerleri aşağıdaki modelden elde edilmektedir.

$$[X\hat{L}_t, X\hat{U}_t] = f_3(XU_{t-1}, \dots, XU_{t-m}, XL_{t-1}, \dots, XL_{t-m}) \quad (6)$$

Burada  $f_3$  alt sınır ve üst sınır değerleri zaman serilerinin gecikmeli değişkenlerinin doğrusal olmayan ve sinir ağından elde edilen fonksiyonlardır. Bu modelde yapay sinir ağının girdileri alt sınır ve üst sınır değerleri olan zaman serilerinin gecikmeli değişkenleri  $XU_{t-1}, \dots, XU_{t-m}, XL_{t-1}, \dots, XL_{t-m}$  ve çıktıları ise  $X\hat{L}_t, X\hat{U}_t$  zaman serileridir.

**Yaklaşım 3.** (Lima vd., 2008) çalışmasında aralık verilerine doğrusal regresyon uygulanması için merkez ve açıklıklara dayalı yaklaşımı önermiştir. Bu yaklaşım zaman serilerinin çözümlenmesi için (Maia vd., 2010); (Maia vd., 2008) çalışmalarında yapay sinir ağları ve SARIMA modelinin, yapay sinir ağları ve Holt üstel düzleştiriminin melezlendiği yöntemlerde uygulanmıştır. Bu çalışmada üçüncü bir yaklaşım olarak merkez ve açıklıklara dayalı yaklaşım yapay sinir ağları ile zaman serisi tahmininde uygulanmıştır. Merkez ve açıklık zaman serileri ( $XC_t$ - Merkez zaman serisi,  $XR_t$ - Açıklık zaman serisi ) aşağıdaki gibi oluşturulur.

$$XC_t = \frac{XU_t + XL_t}{2}, \quad XR_t = \frac{XU_t - XL_t}{2} \quad (7)$$

$XC_t$  ve  $XR_t$  zaman serileri ayrı ayrı ileri beslemeli yapay sinir ağları ile çözülür. Bu durumda  $XC_t$  ve  $XR_t$  zaman serilerinin tahmini aşağıdaki modellerden elde edilir.

$$X\hat{C}_t = f_4(XC_{t-1}, \dots, XC_{t-m}), X\hat{R}_t = f_5(XR_{t-1}, \dots, XR_{t-m}) \quad (8)$$

Burada,  $f_4$  ve  $f_5$  sırasıyla merkez ve açıklık değerleri zaman serilerinin gecikmeli değişkenlerinin doğrusal olmayan ve sinir ağından elde edilen fonksiyonlardır. Merkez ve Açıklık serileri tahmin edildikten sonra,  $X\hat{L}_t, X\hat{U}_t$  zaman serileri aşağıdaki denklemlerden bulunur.

$$X\hat{L}_t = X\hat{C}_t - X\hat{R}_t, X\hat{U}_t = X\hat{C}_t + X\hat{R}_t \quad (9)$$

#### 4. UYGULAMA VE SONUÇLAR

Uygulamada 30.03.2011 ile 30.03.2012 tarihleri arasındaki en düşük ve en yüksek günlük ABD doları /Ons altın fiyatları zaman serisi kullanılmıştır. Toplam 235 gözleme sahip zaman serisinin son 35 gözlemi karşılaştırma için kullanılmıştır. Yani 13.02.2012 ile 30.02.2012 tarihleri arasındaki veriler test kümesi olarak alınmıştır.

İlk olarak en düşük ve en yüksek altın fiyatları zaman serisi yaklaşım 1 ile çözümlenmiştir. Çözümlemeden test kümesi için elde edilen sonuçlar Tablo 1'de verilmiştir. Her iki zaman serisinin çözümlenmesinde tüm tabakalarda lojistik aktivasyon fonksiyonu kullanılmıştır. Yapay sinir ağının girdi sayısı ve gizli tabaka birim sayısı 1 ile 10 arasında değiştirilmiş ve toplam 100 yapay sinir ağı mimarisi içinden en iyi test performansı veren mimariler seçilmiştir. En düşük altın fiyatı ve en yüksek altın fiyatı serisi için aynı mimari 1-4-1'den en iyi test kümesi performansına ulaşılmıştır. 1-4-1 mimarisi girdi tabakasında  $X_{t-1}$  bir adım gecikmeli değişkeninin olduğu, gizli tabakada 4 birimin kullanıldığı mimaridir.

**Tablo 1. Yaklaşım 1 ile en düşük ve en yüksek altın fiyatları zaman serisi için elde edilen sonuçlar**

	MİMARİ	RMSE	DA
En düşük altın fiyatı	1-4-1	17,0621	0,5294
En yüksek altın fiyatı	1-4-1	16,2269	0,5294

İkinci olarak en yüksek ve en düşük altın fiyatları zaman serisinin gecikmeli değişkenlerinin aynı anda yapay sinir ağına girdi olarak verildiği ve en yüksek en düşük altın fiyatının iki çıktısı olarak kullanıldığı yapay sinir ağı mimarisinden en düşük ve en yüksek altın fiyatları birlikte yaklaşım 2 ile tahmin edilmiştir. Çözümlemede kullanılan yapay sinir ağı modelinde tüm tabakalarda lojistik aktivasyon fonksiyonu kullanılmıştır. Girdi sayısı her iki zaman serisi için 1 ile 10 ve gizli tabaka birim eleman sayısı da 1 ile 10 arasında değiştirilmiştir. En iyi test kümesi sonucunu veren mimari 2-5-2 mimarisi yani girdi tabakasında  $X_{t-1}, X_{t-2}, Y_{t-1}, Y_{t-2}$  gecikmeli değişkenleri ve gizli tabakada 5 birimin kullanıldığı yapay sinir ağı mimarisidir. Elde edilen sonuçlar Tablo 2'de verilmiştir.

Tablo 2. Yaklaşım 2 ile en düşük ve en yüksek altın fiyatları zaman serisi için elde edilen sonuçlar

	MİMARİ	RMSE	DA
En düşük altın fiyatı	2-5-2	19,0759	0,6471
En yüksek altın fiyatı		16,4254	0,5000

Veriye üçüncü olarak yaklaşım 3 uygulanmıştır. Uygulamada kullanılan yapay sinir ağı tüm tabakalarda lojistik aktivasyon fonksiyonunun kullanıldığı yapay sinir ağıdır. Elde edilen en iyi mimari için sonuçlar Tablo 3'de verilmiştir. Verinin merkezinin ve açıklığının tahmininde kullanılan yapay sinir ağında girdi ve gizli tabaka birim sayıları 1 ile 10 arasında değiştirilmiştir ve test kümesi için en iyi sonuçları veren mimariler kullanılmıştır. Verinin merkezi için en uygun mimari 1-4-1 yani girdi tabakasında  $X_{t-1}$  bir adım gecikmeli değişkenin olduğu, gizli tabakada 4 birimin kullanıldığı mimaridir. Verinin açıklığı için ise en uygun mimari 5-2-1 mimarisi olarak elde edilmiştir.

Tablo 3. Yaklaşım 3 ile en düşük ve en yüksek altın fiyatları zaman serisi için elde edilen sonuçlar

	MİMARİ	RMSE	DA
En düşük altın fiyatı	1-4-1, 5-2-1	16,8592	0,5000
En yüksek altın fiyatı		15,5734	0,5294

Bu çalışmada kullanılan tüm yapay sinir ağlarında lojistik aktivasyon fonksiyonu kullanılmıştır. Lojistik aktivasyon fonksiyonuna dayalı ileri beslemeli yapay sinir ağlarının başarılı öngörü sonuçları ürettiği (Aladağ vd., 2010); (Alpaslan vd., 2010); (Eğrioğlu vd., 2008) çalışmalarında gösterilmiştir. Bu çalışmada amaç aralık tahminli öngörü yöntemlerinin karşılaştırılması olduğundan aktivasyon fonksiyonunun belirli bir türünde sabitlenmiştir. Ancak diğer aktivasyon fonksiyonunun aralık değerli zaman serisinde etkisi ayrı bir araştırma konusu olarak incelenebilir. Çalışmada kullanılan tüm yapay sinir ağlarının eğitiminde Levenberg-Marquardt yöntemi kullanılmıştır. Özellikle yapay sinir ağının ağırlıklarının sayısı çok fazla olmadığında kısa sürede yakınsama sağlayan ve ikinci dereceden bir yaklaşımdır. Tüm yöntemlerin uygulanmasında "Matlab 2011b Neural Network Toolbox" kullanılmıştır.

## 5. TARTIŞMA

Bu çalışmada aralık değerli zaman serileri tahmini için üç yaklaşım karşılaştırılmıştır. Yaklaşım 1'de zaman serisinin alt ve üst sınır değerlerinin yapay sinir ağları ile ayrı ayrı çözümlenmesi gerçekleştirilmiştir. Yaklaşım 2'de aralık değerli yapay sinir ağının çözümlenmesinde zaman serisinin alt ve üst sınır değerlerinin yapay sinir ağları ile birlikte çözümlenmesi gerçekleştirilmiştir. Bu yöntemde tek bir yapay sinir ağı oluşturulmaktadır. Yaklaşım 3'te ise merkez ve açıklıklara dayalı yaklaşım yapay sinir ağları ile zaman serisi tahmininde uygulanmıştır. Bu üç yaklaşımın değerlendirilmesi RMSE, MAPE ve DA performans ölçütleri ile yapılmıştır. Verilerin çözümlenmesi için kullanılan yapay sinir ağları yaklaşımlarından RMSE ölçütüne göre Yaklaşım 3 en iyi sonuçları vermiştir. DA ölçütü tek başına kullanılmamasına rağmen, öngörünün yön doğruluğunu gösterebilen bir ölçüttür. DA ölçütüne göre Yaklaşım 2 en düşük altın fiyatını yönünü tahmin etmede diğer yöntemlerden başarılı görülmektedir. Sonuç olarak

her üç yönteminde oldukça düşük RMSE değerleri üretmesi aralık değerli bir zaman serisi olan altın fiyatlarının öngörüsü için yapay sinir ağlarının uygulanabileceğini göstermiştir.

## 6. KAYNAKLAR

Aladağ, C. H., Eğrioğlu, E., Gunay, S., Başaran, M. A., 2010. Improving Weighted Information Criterion by Using Optimization, *Journal of Computational and Applied Mathematics*, 233, 2683-2687.

Alpaslan, F., Tiring, E., Eğrioğlu, E., 2010. İleri Beslemeli Yapay Sinir Ağları İle Öngörü İçin Gizli Tabaka Sayısı Üzerine Bir Araştırma, 7. İstatistik Günleri Sempozyumu Bildiriler Kitabı, 53-58, Ankara.

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.

Eğrioğlu, E., Aladağ, Ç. H., Günay, S., 2008. A New Architecture Selection Strategy in Solving Seasonal Autoregressive Time Series by Artificial Neural Networks, *Hacettepe Journal of Mathematics and Statistics*, Volume 37, Issue 2.

Eğrioğlu, E., Aladağ, Ç. H., Gunay, S., 2008. A New Model Selection Strategy in Artificial Neural Network, *Applied Mathematics and Computation*, 195, 591-597.

Günay, S., Eğrioğlu, E., Aladağ, Ç. H., 2007. Tek Değişkenli Zaman Serileri Analizine Giriş, *Hacettepe Üniversitesi Yayınları*.

Billard, L., Diday, E., 2000. Regression Analysis for Interval-valued Data, *Data Analysis, Classification and Related Methods, Proceedings of the Seventh Conference of the International Federation of Classification Societies (IFCS'00)*, Springer, Belgium, 369-374.

Maia, A. L. S., De Carvalho, F. A. T., 2010. Holt's Exponential Smoothing and Neural Network Models for Forecasting Interval-valued Time Series, *International Journal of Forecasting* doi:10.1016/j.ijforecast.2010.02.012.

Maia, A. L. S., De Carvalho, F. A. T., Ludermir, T. B., 2008. Forecasting Models for Interval-valued Time Series, *Neurocomputing*, 71, 3344-3352.

Zhang, G., Patuwo, B. E., Hu, Y. M., 1998. Forecasting with Artificial Neural Networks: The State of The Art, *International Journal of Forecasting*, 14, 35-62.

## HIGHEST AND LOWEST GOLDEN PRICE FORECASTING WITH ARTIFICIAL NEURAL NETWORKS

### ABSTRACT

*This study is focused on analyzing time series such as share price stockholders, golden prices and exchange rates, for which observations can change within the day, by using artificial neural networks. When such time series are analyzed using conventional methods, time series which consist of minimum and maximum values are modeled separately. Golden price is changed between minimum and maximum values for a given day. In this study, three different approaches, which are used for analyzing time series such as golden prices, are compared to forecast the minimum and maximum values with artificial neural networks. While time series composed of minimum and maximum values are analyzed with different neural network models in the first method, these two time series are analyzed with the same neural network model in the second method. The third method is based on center and intervals. Methods are employed on the golden price time series which include daily observations between 30/03/2011 and 30/03/2012, and taken from the website of the Central Bank of Republic of Turkey. Results obtained from this application are discussed.*

**Keywords:** Time series including interval value, Forecasting, Artificial neural networks.

# İLİŞKİLİ HATALARA SAHİP MODELLERDE SPLAYN VE ÇEKİRDEK REGRESYON KESTİRİCİLERİNİN PERFORMANSLARI

Serdar DEMİR\*

Dursun AYDIN\*\*

## ÖZET

*Parametrik olmayan regresyon kestiricilerinin performanslarının, ilişkili hatalara sahip modellerde ciddi biçimde azaldığı iyi bilinmektedir. Bu çalışmada, regresyon fonksiyonunun parametrik olmayan kestirimi için sıkça kullanılan kübik düzleştirme splaynı ile Nadaraya-Watson çekirdek kestiricilerinin hatalı ilişkilere sahip modellerdeki performansları incelenmiştir. Bunun için bir benzetim çalışması gerçekleştirilerek, elde edilen kestirimlerin hata kareler ortalamaları karşılaştırılmıştır. Benzetim sonuçlarına göre, Nadaraya-Watson çekirdek kestiricisi küçük örneklerde iyi performans gösteriyorken, splayn kestiricisi büyük örneklerde iyi performans göstermektedir.*

**Anahtar Kelimeler:** Çekirdek kestiricisi, ilişkili hatalar, Kübik splayn, Nadaraya-Watson.

## 1. GİRİŞ

Parametrik olmayan regresyon kestiricileri, bağımsız değişkene ilişkin gözlem değerleri  $x_i$  ( $x_1 < x_2 < \dots < x_n$ ) ve bağımlı değişkene ilişkin gözlem değerleri  $y_i$  olmak üzere,

$$Y_i = m(x_i) + \varepsilon_i, \quad (i = 1, 2, \dots, n) \quad (1)$$

biçimindeki bir regresyon modelinin bilinmeyen  $m(x)$  fonksiyonunun kestirimi ile ilgilenmektedirler. Burada,  $\varepsilon_i$  hatalarının sıfır ortalamalı ve  $\sigma^2 V$  varyans-kovaryans matrisi ile ilişkili hatalar oldukları varsayılmaktadır.  $V$  matrisi bir ağırlıklandırma matrisidir.

Parametrik olmayan regresyon kestirimleri için yaygın biçimde kullanılan kestiriciler, çekirdek (kernel) regresyon ve splayn (spline) düzleştirme kestiricilerdir (Hardle, 1991) (Eubank, 1999), (Green vd., 1994). Özellikle zaman serileri çözümlemelerinde sıkça karşılaşılan hataların ilişkili olması durumu, çekirdek ve splayn tipi kestiricilerin performanslarını ciddi biçimde etkilemektedir. Diggle ve Hutchinson (1989); Hurvich ve Zeger (1990); Khon vd. (1992); Wang (1998); Krivobokova ve Kauermann (2007); Liu (2009) ve Morton vd. (2009) ilişkili hataların varlığında splayn düzleştirme kestiricileri üzerinde çalışmalar yapmışlardır. Altman (1990); Chu ve Marron (1991); Hart (1991); Herrmann (1992); Rio (1996); Ray (1997); Opsomer vd. (2001); Carroll vd. (2002); Park vd. (2006); Kim vd. (2009); De Brabanter vd. (2010) ve Lee vd. (2010) hataların ilişkili olduğu durumlarda çekirdek tipi kestiricileri ve bant genişliği seçim yöntemlerini incelemişlerdir. Bu çalışmalar, genel olarak eşit aralıklı sabit tasarımı model (örneğin,  $x_i = i/n$ ) varsayımı yapılarak gerçekleştirilmiştir. İlişkili hatalara sahip modellerde, çekirdek ve splayn kestiricilerine ilişkin söz konusu bu

\*Yrd. Doç. Dr., Muğla Sıtkı Kocaman Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [serdardemir@mu.edu.tr](mailto:serdardemir@mu.edu.tr)

\*\*Doç. Dr., Muğla Sıtkı Kocaman Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [duaydin@mmu.edu.tr](mailto:duaydin@mmu.edu.tr)

çalışmalarda kestiricilerin özellikleri kuramsal ve deneysel olarak ayrı ayrı incelenmekte, ancak doğrudan performansları karşılaştırılmamaktadır. Bunun nedeni, kuramsal karmaşıklık ve benzetimlerdeki işlem zorluklarıdır.

Bu çalışmada, rastgele tasarımı model varsayımı altında Nadaraya-Watson çekirdek regresyon kestiricisi ile kübik splayn düzleştirme kestiricisinin performanslarının karşılaştırılması amaçlanmıştır. Kestiricilerin performansları, bir benzetim çalışması ile elde edilen verilere ilişkin kestirimlerin hata kareler ortalamalarına dayalı olarak karşılaştırılmaktadır. İncelenen kestiricilere ilişkin bilgiler bir sonraki bölümde sunulmaktadır. Üçüncü bölümde, kestiricilerin performanslarını karşılaştırmak amacıyla yapılan benzetim çalışmasının ayrıntıları ve sonuçları verilmektedir.

## 2. NADARAYA-WATSON ÇEKİRDEK VE KÜBİK SPLAYN KESTİRİCİLERİ

Çalışmanın bu kısmında, parametrik olmayan regresyon uygulamalarında en sık kullanılan Nadaraya-Watson çekirdek kestiricisi ve kübik splayn kestiricisi tanıtılmaktadır.

### 2.1 Nadaraya-Watson Çekirdek Regresyon Kestiricisi

Parametrik olmayan regresyon fonksiyonu kestirimi uygulamalarında sıkça kullanılan bir kestirici, Formül (2) ile verilen

$$\hat{m}_K(x) = \frac{\sum_{i=1}^n Y_i K[(x-X_i)/h]}{\sum_{i=1}^n K[(x-X_i)/h]} \quad (2)$$

Nadaraya-Watson (NW) çekirdek kestiricisidir (Nadaraya,1964); (Watson,1964) Düzleştirme parametresi olan  $h$  sabiti, kestirimin düzleştirme düzeyini kontrol etmekte ve bant genişliği olarak adlandırılmaktadır. Kestiricinin performansında önemli katkısı olan bant genişliğinin seçimi için çeşitli yöntemler mevcuttur. Başlıca seçim yöntemleri çapraz-geçerlilik, genelleştirilmiş çapraz-geçerlilik, ceza fonksiyonları, plug-in ve bootstrap yöntemleridir (Pagan vd.,1999). Hangi yöntemin daha iyi olduğu konusu tartışmalı bir konudur. Bu çalışmada, splayn kestiriciler için de sıkça kullanılan genelleştirilmiş çapraz-geçerlilik yöntemi ele alınmaktadır. Genelleştirilmiş çapraz-geçerlilik (GÇG) yöntemi ile,

$$GÇG(h) = \frac{\sum_{i=1}^n [Y_i - \hat{m}_K(X_i)]^2}{\left[1 - K(0) / \sum_{i=1}^n K\left(\frac{x_i - x_j}{h}\right)\right]^2} \quad (3)$$

biçimindeki fonksiyonu minimum yapan  $h$  bant genişliği optimal bant genişliği değeri olarak seçilir (Hardle,1991).

Formül 2'de,  $K(\cdot)$  çekirdek fonksiyonu simetrik bir yoğunluk fonksiyonudur. Uygulamada sıkça kullanılan çekirdek fonksiyonları, formül 4 ile verilen Gaussian çekirdek fonksiyonu ve (5) formülü ile verilen Epanechnikov çekirdek fonksiyonudur.



$$K(u) = \exp(-u^2/2)/\sqrt{2\pi}, \quad -\infty < u < \infty \quad (4)$$

$$K(u) = 3(1-u^2)/4, \quad |u| \leq 1 \quad (5)$$

## 2.2 Kübik Splayn Düzleştirme Kestiricisi

Parametrik olmayan regresyon modelinde hataların ilişkili olması durumunda regresyon fonksiyonu  $m(x)$ 'in kestirimi için Diggle ve Hutchinson (1989) tarafından kübik splayn düzleştirmeye dayalı bir yaklaşım önerilmiştir. Bilinen bir  $\rho > 0$  parametresi için,  $\mathbf{V} = [v_{ij}]$  iken,  $v_{ij} = \exp(-\rho |x_i - x_j|)$  otokorelasyon fonksiyonunu kullanan, eşit aralıklı tasarım noktaları (sabit tasarım) gerektirmeyen bu yöntemin splayn düzleştirmeye dayalı çözümü,

$$S[m(x)] = (\mathbf{Y} - \mathbf{M})^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{M}) + \lambda \mathbf{M}^T \mathbf{C} \mathbf{M} \quad (6)$$

biçimindeki cezalı kareler toplamı (penalized sum of squares) ölçütünün  $m(x)$ 'e göre minimum yapılmasını amaçlamaktadır. Formül 6'nın sağ tarafındaki ilk terim verilerin uyumunu (goodness-of-fit), ikinci terim ise düzleştirmeyi (smoothness) tanımlayan ifadelerdir. Sabit bir değer olan  $\lambda$  ise, verilerin uyumu ile düzleştirme arasındaki dengeyi sağlamaktadır. Formül 6 ile bulunan  $m(x)$  kestirimi, splayn düzleştirme kestirimi  $\hat{m}_S(x)$  olarak tanımlanır. Eşitlikte,  $\mathbf{Y} = [y_1 \cdots y_n]^T$  ve  $\mathbf{M} = [m(x_1) \cdots m(x_n)]^T$  biçimindedir.  $\mathbf{C}$  matrisi yarı-pozitif tanımlı bir ceza matrisidir ve  $\mathbf{C} = \mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T$  biçiminde elde edilir. Burada,  $\mathbf{Q}$  matrisi elemanları,  $h_i = x_{i+1} - x_i$  ( $i = 1, 2, \dots, n-1$ ) olmak üzere,  $i = 2, 3, \dots, n$  ve  $j = 2, 3, \dots, n-1$  için

$$q_{ij} = \begin{cases} 1/h_{j-1}, & i = j-1 \text{ ise} \\ -1/h_{j-1} - 1/h_j, & i = j \text{ ise} \\ 0, & |i-j| \geq 2 \text{ ise} \\ 1/h_j, & i = j+1 \text{ ise} \end{cases}$$

biçiminde tanımlanan  $n \times (n-2)$  boyutlu bir matristir.  $\mathbf{R}$  matrisi elemanları,  $i=2,3,\dots,n$  ve  $j=2,3,\dots,n-1$  için

$$r_{ij} = \begin{cases} h_{j-1}/6, & i = j-1 \text{ ise} \\ (-h_{j-1} - h_j)/3, & i = j \text{ ise} \\ 0, & |i-j| \geq 2 \text{ ise} \\ h_j/6, & i = j+1 \text{ ise} \end{cases}$$

biçiminde tanımlanan  $(n-2) \times (n-2)$  boyutlu bir matristir (Green vd., 1994); (Eubank, 1999). Diggle ve Hutchinson (1989), Formül 6'da  $a_i = \exp\{-\rho(x_{i+1} - x_i)\}$  olarak,  $\mathbf{V}^{-1}$  yerine simetrik  $\mathbf{W}$  bant matrisini kullanmıştır.  $\mathbf{W} = [w_{ij}]$  matrisinin sıfırdan farklı olan elemanları aşağıdaki formüller yardımıyla hesaplanmaktadır (Diggle vd., 1989):

$$w_{11} = 1/(1-a_1^2), \quad i = 2, \dots, n-1.$$

$$w_{ii} = 1 + a_{i-1}/(1-a_{i-1}^2) + a_i/(1-a_i^2), \quad i = 2, \dots, n-1.$$

$$w_{nn} = 1/(1-a_n^2), \quad i = 1, \dots, n.$$

$$w_{i+1,1} = w_{i,i+1} = -a/(1-a_1^2), \quad i = 1, \dots, n.$$

Yukarıda verilen tanımlamalar ışığında, 6 formülünü minimum yapan kübik splayn düzleştirme kestiricisi

$$\hat{\mathbf{m}}_S = \begin{bmatrix} \hat{m}_S(X_1) \\ \vdots \\ \hat{m}_S(X_n) \end{bmatrix} = (\mathbf{W}^{-1} + \lambda \mathbf{C})^{-1} \mathbf{W}^{-1} \mathbf{Y} = \mathbf{S}_\lambda \mathbf{Y} \quad (7)$$

biçimindedir. Burada  $\mathbf{S}_\lambda$  matrisi şapka ya da düzleştirme matrisi olarak adlandırılmaktadır.

Formül 6'daki düzleştirme parametresi  $\lambda$  bir denge unsuru olarak, kübik splayn kestiricisinin performansında önemli rol oynamaktadır.  $\lambda$ 'nın küçük seçilmesi fazlasıyla pürüzlü bir tahmine, büyük seçilmesi ise fazlasıyla düz bir tahmine yol açabilmektedir.  $\lambda$ 'nın belirlenmesine ilişkin literatürde pek çok yöntem (çapraz geçerlilik, genelleştirilmiş çapraz geçerlilik, Akaike bilgi ölçütü, Mallows'un  $C_p$  ölçütü, Plug-in vd.) mevcuttur. Kübik splayn kestiricisi için  $\lambda$  parametresinin seçiminde en çok tercih edilen yöntem, genelleştirilmiş çapraz-geçerlilik (GÇG) yöntemidir (Eubank,1999). Bu yöntem, hem etkin olması hem de özellikle kolay uygulanabilir olması nedeniyle ilgi çekmektedir. İlişkili hatalara sahip modeller için, Diggle ve Hutchinson (1989) tarafından uyarlanan,

$$G\check{C}G(\lambda) = [\mathbf{Y} - \hat{\mathbf{m}}_S]^T \mathbf{W} [\mathbf{Y} - \hat{\mathbf{m}}_S] / \{\text{tr}[\mathbf{I} - \mathbf{S}_\lambda]\}^2 \quad (8)$$

biçimindeki genelleştirilmiş çapraz-geçerlilik fonksiyonu kullanılmaktadır. Bu fonksiyonu minimum yapan  $\lambda$  değeri, optimal düzleştirme parametresi değeri olarak seçilmektedir.

### 3. BENZETİM ÇALIŞMASI

İlişkili hatalara sahip 1 formülü ile verilen regresyon modellerinde bilinmeyen  $m(x)$  fonksiyonun kestiriminde, Nadaraya-Watson çekirdek regresyon kestiricisi ( $\hat{m}_K$ ) ile kübik splayn düzleştirme kestiricisinin) performanslarını karşılaştırmak için bir benzetim çalışması yapılmıştır. Bu benzetim çalışması için (Chu,1991); (Rio, 1996); (Goijjier, 2002); (Kim, 2009); (Brabanter, 2010) ve (Lee, 2010) tarafından yapılan benzetim çalışmaları dikkate alınarak,  $m(x) = x^3(1-x^3)$  regresyon fonksiyonu kullanılmış ve

$$y_i = x_i^3(1-x_i^3) + \varepsilon_i \quad (9)$$

modeline uygun olarak veriler üretilmiştir. Burada,  $x_i$ 'ler  $[0,1]$  aralığında tekbiçimli (uniform) dağılımdan rastgele çekilmiştir. Modelin  $\varepsilon_i$  hataları,  $\varepsilon_0 \sim N(0, 0.0001/(1-\rho^2))$  olmak üzere,

$$\varepsilon_i = \rho \varepsilon_{i-1} + u_i, \quad i=1, 2, 3, \dots, n$$

biçimindeki birinci dereceden otoregresif ( $AR(1)$ ) bir süreçten üretilmiştir. Sürecin  $u_i$  hataları, sıfır ortalamalı ve 0.0001 varyanslı bir normal dağılımdan rastgele çekilmiştir.

Otokorelasyon katsayısı  $\rho = 0.30, 0.60, 0.90$  değerlerinin herbiri için  $n = 25, 50, 100$  büyüklüğünde rastgele örneklemeler üretilmiştir. Benzetimler için tekrar sayısı  $r = 500$  olarak alınmıştır.

Nadaraya-Watson kestirimleri için Gaussian çekirdek fonksiyonu kullanılmıştır. Bant genişliği  $h$  Formül 3 ile verilen genelleştirilmiş çapraz-geçerlilik yöntemiyle belirlenmiştir. Splayn kestirimleri için düzleştirme parametresi  $\lambda$  ise Formül 8 ile verilen genelleştirilmiş çapraz-geçerlilik yöntemiyle belirlenmiştir.

Her bir örneklem genişliği (25, 50, 100) ve her bir otokorelasyon katsayısı (0.30, 0.60, 0.90) için üretilen verilere ilişkin Nadaraya-Watson çekirdek kestirimleri ve splayn düzleştirme kestirimleri hesaplanmıştır. Daha sonra her bir kestirimin hata kareler ortalaması bulunmuştur. Bu işlem 500 kez tekrarlanmış ve 500 kestirim üzerinden

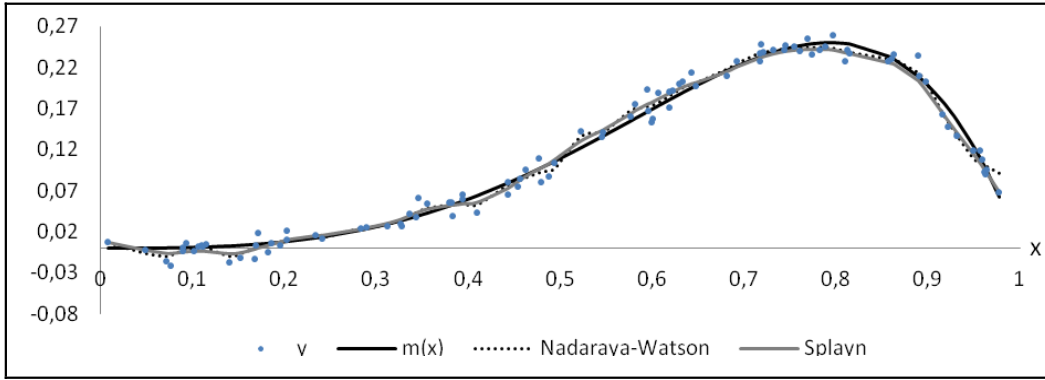
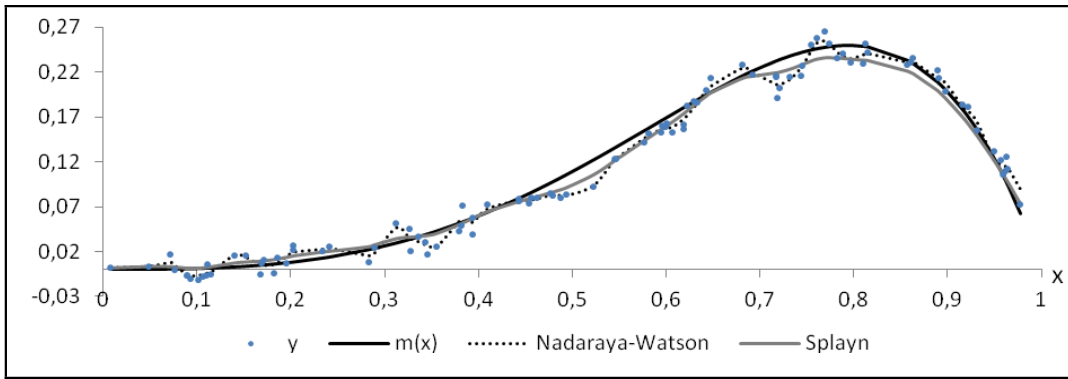
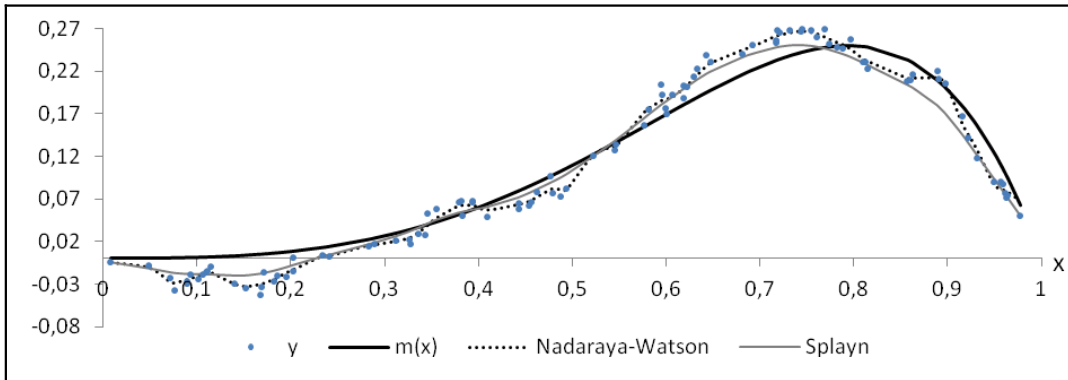
Ortalama Hata Kareler Ortalamaları ( $\text{OHKO} = \frac{\sum_{i=1}^r \text{HKO}_i}{r}$ ) bulunmuştur. Her bir duruma ilişkin kestirimlerin OHKO değerleri Tablo 1’de verilmektedir.

Tablo 1’den,  $\rho$ ’nun her üç değeri için  $n=25$  örneklem büyüklüğünde Nadaraya-Watson kestirimlerinin daha küçük hata kareler ortalamalarına sahip oldukları görülmektedir. Buna karşın,  $n=50$  ve  $n=100$  örneklem genişliklerinde ise splayn kestirimleri daha küçük hata kareler ortalamalarına sahip olmaktadır. Örneklem genişliği arttıkça, splayn kestiricisinin Nadaraya-Watson kestiricisine göre daha iyi performans gösterdiği söylenebilir. Ayrıca beklenildiği gibi, hataların ilişki düzeyleri arttıkça tüm örneklem büyüklüklerinde hem Nadaraya-Watson hem de splayn kestirimlerin hata kareler ortalamalarının büyüdüğü görülmektedir.

Tablo 1. Kestirimlerin ortalama hata kareler ortalamaları

$\rho$	$n$	$\text{OHKO}(\widehat{m}_K)$	$\text{OHKO}(\widehat{m}_S)$	$\text{OHKO}(\widehat{m}_S)/\text{OHKO}(\widehat{m}_K)$
0.30	25	0.0000927	0.0001107	1.194681
	50	0.0000684	0.0000507	0.741656
	100	0.0000499	0.0000366	0.732469
0.60	25	0.0001611	0.0001756	1.089935
	50	0.0001231	0.0001047	0.850226
	100	0.0001032	0.0000834	0.808648
0.90	25	0.0007215	0.0007377	1.022485
	50	0.0006614	0.0006062	0.916553
	100	0.0005553	0.0004495	0.809438

Rastgele seçilen  $n=100$  büyüklüğündeki üç benzetim örneklemini için  $m(x)$  fonksiyonunun Nadaraya-Watson çekirdek ve splayn düzleştirme kestirimlerinin grafikleri Şekil 1, Şekil 2 ve Şekil 3’de verilmektedir.

Şekil 1.  $\rho=0.30$  ve  $n=100$  için splayn ve Nadaraya-Watson kestirimleriŞekil 2.  $\rho=0.60$  ve  $n=100$  için splayn ve Nadaraya-Watson kestirimleriŞekil 3.  $\rho=0.90$  ve  $n=100$  için splayn ve Nadaraya-Watson kestirimleri

Şekil 1, Şekil 2 ve Şekil 3'ten, hatalar arasındaki ilişki arttıkça Nadaraya-Watson çekirdek ve splayn düzleştirme kestirimlerinin kötüleştiği görülmektedir. Bununla beraber, splayn kestirimlerinin Nadaraya-Watson kestirimlerine göre daha pürüzsüz eğriler ortaya çıkardığı söylenebilir.

#### 4. SONUÇ VE TARTIŞMA

Bu çalışmada, hataları ilişkili olan rastgele tasarım modellerinde Nadaraya-Watson çekirdek kestiricisi ile kübik splayn düzleştirme kestiricisinin performansları, hata kareler ortalaması ölçütüne dayalı olarak karşılaştırılmıştır. Bu amaçla gerçekleştirilen benzetim sonuçlarına göre, küçük örneklerde Nadaraya-Watson çekirdek kestiricisinin, büyük örneklerde ise splayn kestiricisinin daha iyi performans gösterdiği söylenebilir. Ayrıca, beklenildiği gibi hatalardaki ilişki düzeyi arttıkça iki kestiricinin de performanslarının kötüleştiği görülmüştür. Bununla beraber, hataların ilişki düzeyinin artmasından, küçük örneklerde Nadaraya-Watson kestiricisi daha fazla etkilenirken, büyük örneklerde ise splayn kestiricisi daha fazla etkilenmektedir. Sonuç olarak, ilişkili hatalara sahip regresyon modellerinde, küçük örneklem durumunda Nadaraya-Watson çekirdek kestiricisinin kullanılması, büyük örneklem durumunda ise kübik splayn kestiricinin kullanılması önerilmektedir. Gelecekteki çalışmalarda, farklı bant genişliği ya da düzleştirme parametresi seçim yöntemleri bakımından karşılaştırmalar yapılması planlanmaktadır.

#### 5. KAYNAKLAR

- Altman, N. S., 1990. Kernel Smoothing of Data with Correlated Errors, *JASA*, 85, 749-759.
- Carroll, R. J., Linton, O., Mammen, E., Xiao, Z., 2002. More Efficient Kernel Estimation in Nonparametric Regression with Autocorrelated Errors, *STICERD-Econometrics Paper Series*, 2002/435.
- Chu, C. K., Marron, J. S. 1991. Comparison of Two Bandwidth Selectors with Dependent Errors, *Ann. Statist.*, 19, 1906-1918.
- De Brabanter, K., De Brabanter, J., Suykens, J., De Moor, B., 2010. Kernel Regression with Correlated Errors. *Proc. of the 11th International Symposium on Computer Applications in Biotechnology (CAB)*, Leuven, Belgium, Jul. 2010, 13-18.
- De Gooijer, J. G., Gannoun, A., Larramendy, I., 2002. Nonparametric Regression with Serially Correlated Errors. *Ann. Ins. Stat. Univ. Paris*, 46, Fasc. 1-2.
- Diggle, P. J., Hutchinson, M. F., 1989. On Spline Smoothing With Autocorrelated Errors. *Australian Journal of Statistics*. 31. 166-182.
- Eubank, R., 1999. *Nonparametric Regression and Spline Smoothing*, New York, Dekker.
- Green, D. J., Silverman, B. W., 1994. *Nonparametric Regression and Generalized Linear Models*, London, Chapman & Hall.
- Hardle, W., 1991. *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.
- Hart, J. D., 1991. Kernel Regression Estimation with Time Series Errors, *J. R. Statist. Soc. B*, 53, 173-187.

- Hermann, E., Gasser, T., Kneip, A., 1992. Choice of Bandwidth for Kernel Regression When Residuals are Correlated, *Biometrika*, 79, 783-795.
- Hurvich, C. M., Zeger, S. L., 1990. A Frequency Domain Selection Criterion for Regression with Autocorrelated Errors, *J. Am. Statist. Ass.*, 85, 705-714.
- Kim, T. Y., Park, B. U., Moon, M. S., Kim, C., 2009. Using Bimodal Kernel Inference in Nonparametric Regression with Correlated Errors, *J. Multivariate Analysis*, 100, 1487-1497.
- Kohn, R., Ansley, C. G., Wong, C. M., 1992. Nonparametric Spline Regression with Autoregressive Moving Average Errors, *Biometrika*, 79, 335-346.
- Krivobokova, T., Kauermann, G., 2007. A Note on Penalized Spline Smoothing with Correlated Errors. *Journal of the American Statistical Association*, 102 (480), 1328-1337.
- Lee, Y. K., Mammen, E., Park, B. U., 2010. Bandwidth Selection for Kernel Regression with Correlated Errors, 44, 4, 327-340.
- Liu Jun, M., 2009. Nonlinear Time Series Modeling Using Spline-Based Nonparametric Models, AMATH'09 Proceedings of the 15th American Conference on Applied Mathematics.
- Morton, R., Kang, E. L., Henderson, B., 2009. Smoothing Splines for Trend Estimation and Prediction in Time Series, *Environmetrics*, 20, 249-259.
- Nadaraya, E. A., 1964. On Estimating Regression, *Theory Pb. Appl.*, Vol.10, 186-190.
- Opsomer, J., Wang, Y., Yang, Y., 2001. Nonparametric Regression with Correlated Errors, *Statist. Sci.* 16, 134-153.
- Pagan, A., Ullah A., 1999. *Nonparametric Econometrics*, Cambridge, Cambridge University Press.
- Park, B. U., Lee, Y. K., Kim, T. Y., Park, C., 2006. A Simple Estimator of Error Correlation in Nonparametric Regression Models, *Scand. J. Statist.* 33, 451-462.
- Ray, B. K., Tsay, R. S., 1997. Bandwidth Selection for Kernel Regression with Long-range Dependent Errors, *Biometrika*, 84(4), 791-802.
- Rio, A. Q., 1996. Comparison of Bandwidth Selectors in Nonparametric Regression Under Dependence, *Comput. Statist. Data Anal.*, 21, 563-580.
- Wang, Y., 1998. Smoothing Spline Models with Correlated Random Errors, *JASA*, 93, 34-348.
- Watson, G. S., 1964. Smooth Regression Analysis, *Sankhya, Series A*, 26, 359-372.

## THE PERFORMANCES OF SPLINE AND KERNEL REGRESSION ESTIMATORS IN MODELS WITH CORRELATED ERRORS

### ABSTRACT

*It is well known that the performances of nonparametric regression estimators severely decrease in the models with correlated errors. In this paper, the performances of the cubic smoothing spline and the Nadaraya-Watson kernel estimators, which are often used for nonparametric estimation of regression function, are investigated in the models with correlated errors. For this purpose, a simulation study is performed and the mean square errors of the estimators are compared. The simulation results show that the Nadaraya-Watson kernel estimator performs well in small samples, while spline estimator performs well in large samples.*

**Keywords:** Kernel estimator, Correlated errors, Cubic spline, Nadaraya-Watson.

## YÜKSEK ÖĞRETİMDE ÖĞRENCİ İŞLERİ HİZMET KALİTESİNİN ÖLÇÜMÜ VE BİR ÖRNEK UYGULAMA

Mesut KUMRU\*

Pınar YILDIZ KUMRU\*\*

### ÖZET

*Soyutluk, ayrılmazlık, heterojenlik ve stoklanamama temel özelliklerinden dolayı hizmetin kalitesinin tanımlanması ve ölçümü genelde gayet zordur. Uygulamada hizmet kalitesinin ölçümü için pek çok yöntem geliştirilmiştir. Bunlardan bir tanesi de SERVQUAL yöntemidir. Bu çalışmada, yüksek öğretimde öğrenci işleri hizmet kalitesinin ölçümü için anılan yöntem kullanılmış, hizmetten yararlananların beklentileri ile algıladıkları hizmet performansı arasındaki açıklıklar incelenmiştir. Uygulama sırasında anket yönteminden yararlanılarak veri toplanmıştır. Yapılan ölçüm sonrasında öğrenci işleri müşterilerinin algılama düzeylerinin yüksek çıkmasına rağmen, beklentilerinin de yüksekliği öğrenci işleri bölümünün okulda verdiği hizmet seviyesinin hizmet alanlarca yeterli bulunmadığını göstermektedir.*

**Anahtar Kelimeler:** Hizmet kalitesi, SERVQUAL, Veri güvenirliliği, Temel bileşenler analizi.

### 1. GİRİŞ

Hizmet sektörünün önemi ve ekonomideki ağırlığı gün geçtikçe artmaktadır. Hizmet işletmelerinde kalite ve performans göstergelerinin belirlenmesi ve ölçülmesi üretim işletmelerine kıyasla daha zordur. Zira, hizmetin genelde üründen farklı özellikleri vardır. Hizmet soyuttur, heterojendir, üretildiği anda tüketilir ve stoklanamaz (Armstrong vd., 2000). Dolayısıyla, hizmetin kazandırdığı faydanın nicel ölçütlerle değerlendirilmesi kolay değildir. Bu yüzden, hizmetin değerlendirilmesinde, genellikle müşterilerin algılama düzeyleri üzerine ölçekler geliştirilmiştir. Hizmet kalitesi, verilen hizmetin müşteri beklentilerinin ne kadarını karşılayabildiğinin bir ölçüsü olarak ifade edilmiş, kaliteli hizmet vermek müşteri beklentilerinin karşılanabilmesi olarak tanımlanmıştır.

Özel sektörde olduğu kadar, kamu kurumlarının hizmetlerinde de kalite önemli bir yer tutmaktadır. Üniversitelerde sunulan hizmetlerin de bu açıdan değerlendirilmesi gerekir. Özellikle, öğrenci işleri hizmet kalitesinin belirlenmesi bu hizmetten yararlanan çok sayıdaki müşteriyi (öğretim elemanı, öğrenci ve veliler) ilgilendirmektedir. Bu husus, bu çalışmanın çıkış noktasını oluşturmuş, aşağıdaki bölümlerde konu ayrıntılı bir uygulamayla gözler önüne serilmiştir.

### 2. HİZMET KALİTESİNİN ÖLÇÜMÜ

Hizmet kalitesinin ölçümünde en yaygın kullanılan ölçek Zeithaml, Parasuraman ve Berry tarafından geliştirilen SERVQUAL yöntemidir (Zeithaml vd.,1990). Ölçek, genelde müşterinin hizmetten beklentilerini ve hizmeti algılama düzeyini belirlemek için tasarlanmıştır. Ayrıca, müşteri beklentilerinin ve algılamalarının öğrenilmesine olanak sağladığından dolayı bilginin müşteri ve yönetim arasında paylaşımında da

\*Doç. Dr., Doğu Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, e-posta: [mkumru@dogus.edu.tr](mailto:mkumru@dogus.edu.tr)

\*\*Yrd. Doç. Dr., Kocaeli Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, e-posta: [pinarki@kou.edu.tr](mailto:pinarki@kou.edu.tr)



kolaylık sunmaktadır. SERVQUAL sıkça kullanılan bir veri toplama aracı olarak hizmet kalitesini ölçmekle sınırlı kalmaz, hizmet kalitesini müşterinin hizmet elemanlarına verdiği önemle orantılı olarak ölçer. Bu ölçme yöntemi çeşitli önermelerden oluşan 5 bölüm (fiziksel özellikler, güvenilirlik, heveslilik, duyarlılık, yeterlilik) içerir. Sonuçta, hizmet kalitesi, “müşterilerin istek veya beklentileri ile algıları arasındaki *açıklık* ölçüsü” şeklinde tanımlanır. Bu açıklıklar aşağıdaki şekilde ifade edilmektedir (Zeithaml vd., 2006):

- a) Müşterinin bir hizmetten beklentisi ile hizmeti sunan kurumun yöneticilerinin müşteri beklentilerini algılamaları arasındaki açıklık.
- b) Yönetimin müşteri beklentilerini algılamaları ile bu algılanan beklentilere göre düzenlenen hizmet kalite şartnameleri arasındaki açıklık.
- c) Beklentiler doğrultusunda hazırlanan şartnameler ile sunulan hizmet arasındaki açıklık.
- d) Sunulan hizmet ile dış çevreden müşteriye iletilen mesajlar arasındaki açıklık.
- e) Müşteri beklentisi ile müşterinin aldığı hizmet arasındaki açıklık. (Çalışmada bu açıklık kullanılmıştır.)

Açıklık analizinde beş temel faktör dikkate alınmaktadır (Parasumaran vd., 1991). Bu faktörler *fiziksel özellikler (tangibles)*, *güvenilirlik (reliability)*, *heveslilik (responsiveness)*, *duyarlılık (emphaty)* ve *yeterlilik (assurance)*'dir. *Fiziksel özellikler*; işletmenin hizmet sunumundaki bina, araç- gereç ve personel görünümünü ele almakta, *güvenilirlik*; söz verilen bir hizmeti doğru ve güvenilir bir şekilde yerine getirebilme yeteneğini ifade etmekte, *heveslilik*; müşteriye yardım etme ve hızlı hizmet verme istekliliği, *duyarlılık*; çalışanların kendilerini müşterilerin yerine koyması ve müşterilere kişisel ilgi göstermesi, *yeterlilik* ise çalışanların bilgili ve nazik olmaları ve müşterilere güven duygusu uyandırabilme becerileri olarak tanımlanmaktadır (Parasumaran vd., 1991). Hizmet birimlerinin hedefi, en az müşteri beklentisi kadar performans ortaya koymaktır. Müşteri beklentisinin üzerinde bir performans ise müşteri sadakati oluşturmak açısından son derece önemlidir.

### 3. LİTERATÜR ARAŞTIRMASI

SERVQUAL yöntemi hizmet kalitesi ölçümünde yaygın olarak kullanılmaktadır. Farklı sektörler itibarıyla tipik uygulama örneklerini şöyle sıralayabiliriz: (Knutson vd., 1990) pansiyonculukta, (Stevens vd., 1995) otel ve restoran işletmeciliğinde, (Nitecki,1996) akademik kütüphanelerde, (Devebakan vd., 2003) sağlık işletmelerinde, (Altan vd., 2004) bankacılık sektöründe, (Atan vd., 2006) süpermarketlerde, (Ruzan vd., 2007) Ege ihracatçı birliklerinde, (Okumuş vd., 2007) havayolu işletmesinde, (Albacete-S vd., 2007) kırsal konaklamada, (Eleren vd., 2007) hazır yemek işletmesinde, (Aydın, 2007) seyahat işletmesinde, (Maddern vd., 2007) finansal hizmetlerde, (Yılmaz vd., 2007) yüksek öğretimde, (Lai vd., 2007) mobil iletişimde, (Ersöz vd., 2009) öğretmenevinde, (Zaim vd.,2009) hastanede, (Chen vd., 2009) gemi taşımacılığında, (Filiz vd., 2010) Eskişehir Belediyelerinde, (Filiz, 2011) yurt işletmesinde, SERVQUAL yöntemini kullanmışlardır.

Yerli ve yabancı literatürde, SERVQUAL yönteminin yüksek öğretimde öğrenci işleri hizmet kalitesinin ölçülmesinde kullanıldığına dair herhangi bir çalışmaya

rastlanmamıştır. (Abouchedid vd., 2002) özel bir üniversitede fakülteler itibariyle kayıt ve akademik danışmanlıkla ilgili davranışları ve öğrenci hizmet kalitesi algılamasını ölçmüştür. Bu çalışmanın bir uzantısı olarak, (Kitcharoen, 2004) öğrenciler ve akademik kadronun çeşitli idari hizmet bölümleri için (kayıt, finansman ve kütüphane) beklenti ve algılamalarını incelemiştir. Bu çalışmalarda öğrenci işleri ofisi doğrudan ele alınmamış, ayrıca farklı bir müşteri grubu olarak idari personel de çalışmaya dahil edilmemiştir. Bu bakımdan, bu çalışma bir özgünlük taşımaktadır.

Aşağıdaki bölümlerde, bir kamu yüksek okulunun öğrenci işleri hizmet kalitesi SERVQUAL yöntemi kullanılarak incelenmektedir.

### 3. UYGULAMA

#### 3.1 Uygulama Yapılan Kurum ve Uygulama Amacı

Çalışma, Kocaeli bölgesinde faaliyet gösteren bir kamu meslek yüksek okulunda gerçekleştirilmiş ve kalite ölçeği olarak sözü edilen SERVQUAL yönteminden yararlanılmıştır (Yaş, 2009). Araştırmanın amacı; öğrenci işleri hizmet kalitesinin değerlendirilmesi, kaliteyi oluşturan çeşitli faktörlerin hizmetten yararlananlar (müşteriler) gözetilerek analiz edilmesi ve böylece kalitenin iyileştirilmesine olanak sağlanmasıdır. Hizmetin müşterilerce nasıl algılandığı, hizmetin hangi boyutlarının nasıl değerlendirildiği de araştırmada ortaya konulması hedeflenen hususlardır.

Araştırmanın müşteri kitlesini öğrenciler, akademik ve idari personel oluşturmaktadır. Anılan okulda toplam öğrenci sayısı 3458, akademik personel sayısı 45 ve idari personel sayısı 12 olmak üzere toplam müşteri sayısı 3515'dir. Rastgele örnekleme yöntemiyle elde edilen veriler önce güvenilirlik analizine tabi tutulmuş, daha sonra yukarıda zikredilen beş temel faktör (boyut) açısından incelenerek müşterilerin beklenti ve algılama düzeyleri karşılaştırılmıştır. Önerme düzeyleri arasındaki farklar (SERVQUAL açıklık skorları) anlamlılık testine (t test) tabi tutulmuştur.

#### 3.2 Veri Toplanması

Araştırma için bir anket formu tasarlanmış, tasarlanan anket formu 2009 yılı bahar döneminde 3 haftalık bir süre boyunca yüksek okulun yerleşkesinde rastgele seçilen müşteriler üzerinde uygulanmıştır. Ankete katılanlar öğrenciler, idari ve akademik personelden oluşmaktadır ve bu kişilerin hepsi öğrenci işleri bölümünün hizmetlerinden faydalanmaktadır.  $P=0.85$ ,  $d=0.05$  ve  $0.95$  güven düzeyinde 196 örnek yeterli olmasına karşın  $(n=(3515(0.85)(0.15)1.96^2)/3514 (0.05)^2)=196$ , toplamda 220 kişi ile (ana kitlenin yaklaşık %6.3'ü) anket yapılmış; bunlardan 6 tanesi geçersiz kabul edilerek değerlendirme dışı tutulmuştur. 214 anket geçerli kabul edilmiştir.  $P=0.85$  değeri okulda yapılan bir ön çalışma sonrasında ortaya çıkmıştır.

Ankete katılanların müşteri grubu ve cinsiyet dağılımı Tablo 1'de verilmiştir.

**Tablo 1. Ankete katılanların müşteri grubu ve cinsiyet dağılımı**

	Toplam		Erkek		Kadın	
	Kişi Adedi	%	Kişi Adedi	%	Kişi Adedi	%
Öğrenci / 1. Öğretim	121	57	59	49	62	51
Öğrenci/ 2. Öğretim	82	38	48	59	34	41
İdari Personel	3	1	2	67	1	33
Akademik Personel	8	4	5	63	3	37
Toplam	214	100				

Anket 2 bölümden oluşmaktadır. Birinci bölüm beklenti önermelerini, ikinci bölüm ise algılama önermelerini içermektedir. Toplamda bölümlerde 22'ser önerme yer almaktadır. Her önermenin altında 1: Kesinlikle Katılmıyorum...7: Kesinlikle Katılıyorum şeklinde 7'li Likert ölçeği kullanılmıştır. Önermelerin SERVQUAL ölçeğine göre dağılımı şu şekildedir: Fiziksel Özellikler (1-4), Güvenilirlik (5-9), Heveslilik (10-13), Duyarlılık (14-17) ve Yeterlilik (18-22) (Bkz. Ek). Hizmet kalitesinin hesaplanmasında kullanılan yol basitçe (algılama skorları – beklenti skorları) şeklinde olup, ortaya çıkan fark *Servqual açıklık skoru* olarak ifade edilmektedir. Her önerme için skor +6 ile -6 arasında (sayı negatif yönde büyüdükçe hizmet kalitesinin azaldığını göstererek) değişebilir. Artı değerli skorlar hizmetin kaliteli, eksi değerler ise kalitesiz olduğunu göstermektedir.

Verilerin analizinde SPSS 17.0 paket programı kullanılmıştır.

### 3.3 Verilerin Güvenilirliği

Anket araştırması tamamlandıktan sonra elde edilen veriler güvenilirlik analizine tabi tutulmuştur. Faktörler bazında ve genelde bulunan Cronbach's Alpha değerleri aşağıdaki tabloda verilmektedir. Buna göre, hiçbir faktör kabul edilebilir seviye olan 0.70 değerinin altında çıkmamıştır. Faktörler bazında ve genel güvenilirlik mükemmel denebilecek seviyededir (Cronbach's, 2012) (Tablo 2).

**Tablo 2. Araştırmanın Cronbach's Alpha değerleri**

Faktörler	Alpha	Faktörler	Alpha
1. Fiziksel Özellikler	0.919	4. Duyarlılık	0.923
2. Güvenilirlik	0.919	5. Yeterlilik	0.921
3. Heveslilik	0.923	<b>Genel</b>	<b>0.923</b>

### 3.4. Önermelerin Gruplanması

Ankette yer alan 22 önermenin, hizmetin boyutları itibarıyla gruplanmasını belirlemek amacıyla, algılamaya göre temel bileşenler matrisi oluşturulmuştur (Tablo 3). Tablodan da görüldüğü gibi, uygulama başlangıcında öngörülen önerme bölünmeleri (hangi önermenin hangi faktöre ait olduğu), 22. önerme hariç, aynen gerçekleşmiştir. Yalnızca 22. soru 1. faktörün altında yer aldığından önceden yapılan boyutlandırmaya uymamaktadır. Bu nedenle, bu soru değerlendirilmeden çıkartılmıştır.

**Tablo 3. Temel bileşenler matrisi**

Önerme	Bileşen					Önerme	Bileşen				
	1	2	3	4	5		1	2	3	4	5
1	<b>0.620</b>	0.040	0.024	0.215	0.286	12	0.185	0.021	<b>0.811</b>	0.325	0.092
2	<b>0.635</b>	0.091	0.314	0.113	0.134	13	0.132	0.044	<b>0.779</b>	0.350	0.182
3	<b>0.692</b>	0.380	0.073	0.374	0.249	14	0.075	0.451	0.281	<b>0.490</b>	0.221
4	<b>0.564</b>	0.432	0.321	0.216	0.174	15	0.399	0.109	0.171	<b>0.482</b>	0.378
5	0.240	<b>0.524</b>	0.244	0.339	0.431	16	0.152	0.155	0.399	<b>0.596</b>	0.381
6	0.102	<b>0.543</b>	0.497	0.170	0.130	17	0.002	0.004	0.309	<b>0.625</b>	0.038
7	0.088	<b>0.684</b>	0.173	0.116	0.153	18	0.204	0.239	0.284	0.426	<b>0.602</b>
8	0.149	<b>0.650</b>	0.153	0.279	0.007	19	0.027	0.053	0.253	0.183	<b>0.732</b>
9	0.089	<b>0.742</b>	0.274	0.084	0.017	20	0.107	0.221	0.095	0.052	<b>0.749</b>
10	0.467	0.056	<b>0.652</b>	0.235	0.051	21	0.167	0.432	0.134	0.135	<b>0.664</b>
11	0.013	0.087	<b>0.761</b>	0.195	0.013	22	<b>0.476</b>	0.250	0.128	0.136	0.399

Faktörlerin varyans açıklama oranları, rotasyon da dahil edilerek Tablo 4’de verilmiştir. Çizelge incelendiğinde, beş faktörün toplamda varyans açıklama oranının %88.235’e ulaştığı görülmektedir. Birinci faktörün varyans açıklama oranı (%28.461) diğerlerine göre oldukça yüksektir. Böylece, anketin müşteri memnuniyetini ölçmede çok başarılı olduğu söylenebilir.

**Tablo 4. Toplam açıklanan varyans**

Bileşen	Başlangıç Öz değerleri			Kareli Yükler Toplamı (Çıkarım)			Kareli Yükler Toplamı (Rotasyon)		
	Toplam	Varyans %	Kümülatif %	Toplam	Varyans %	Kümülatif %	Toplam	Varyans %	Kümülatif %
1	6.264	28.461	28.461	6.254	28.461	28.461	5.003	24.191	24.191
2	4.022	18.274	46.755	4.022	18.174	46.735	3.619	16.357	40.548
3	3.234	14.693	61.428	3.234	14.693	61.428	3.557	16.162	56.710
4	3.041	13.817	75.245	3.041	13.817	75.245	3.497	15.890	72.600
5	2.859	12.990	88.235	2.859	12.990	88.235	3.441	15.635	88.235
6	0.528	2.399	90.634						
7	0.406	1.844	92.478						
8-21	.....	.....	.....						
22	0.009	0.040	100.000						

### 3.5 Servqual Açıklık Skorlarının Hesaplanması ve Değerlendirme

SERVQUAL açıklık skorları önce önerme bazında, daha sonra faktör bazında, en sonunda da genel ortalama bazında belirlenerek analiz edilmiştir. Önerme bazındaki SERVQUAL açıklık skorları Tablo 5’de verilmektedir.

Tablo 5. Önerme bazında Servqual açıklık skorları

N=214 Önerme	ALGILAMA		BEKLENTİ		SKOR (Fark) (c)=(a-b)	SKOR Std. Sapma $\sqrt{\frac{\sum c_i^2 - n\bar{c}^2}{n-1}}$	t	p
	Ortalama	Std. Sapma	Ortalama	Std. Sapma				
	(a)		(b)					
1	3.8084	0.6044	6.7336	1.5969	-2.9252	1.6114	26.5445	0.0000
2	4.3364	0.6428	6.6869	1.6836	-2.3505	1.6915	20.3330	0.0000
3	3.1638	0.6715	6.7103	1.6605	-3.5467	1.8173	28.5564	0.0000
4	2.8037	0.8308	6.6075	1.6583	-3.8037	1.8514	30.0450	0.0000
5	4.3972	0.5063	6.8271	1.8401	-2.4299	1.8733	18.9688	0.0000
6	4.7617	0.4571	6.8551	1.7561	-2.0935	1.7723	17.2731	0.0000
7	4.7103	0.6739	6.7243	1.8462	-2.0140	1.7839	16.5217	0.0000
8	4.6121	0.3634	6.9065	1.8599	-2.2944	1.9058	17.6086	0.0000
9	4.1963	0.4485	6.8785	1.7030	-2.6822	1.7588	22.3145	0.0000
10	3.9486	0.5363	6.8224	1.8695	-2.8738	1.8876	22.2775	0.0000
11	4.0280	0.6006	6.7617	1.7413	-2.7336	1.7556	22.7800	0.0000
12	3.8832	0.6006	6.7290	1.8412	-2.8458	1.8864	22.0776	0.0000
13	3.9953	0.6808	6.7243	1.8576	-2.7290	1.8863	21.1715	0.0000
14	4.6168	0.4711	6.8364	1.7770	-2.2196	1.7431	18.6208	0.0000
15	3.7383	0.7591	6.6776	1.6858	-2.9393	1.7543	24.5146	0.0000
16	3.6869	0.6684	6.7196	1.6979	-3.0327	1.7291	25.6574	0.0000
17	3.5514	0.7979	6.7150	1.7292	-3.1636	1.8405	25.1479	0.0000
18	4.4112	0.4905	6.8224	1.6527	-2.4112	1.6872	20.9124	0.0000
19	4.2850	0.4870	6.8131	1.8278	-2.5280	1.8119	20.4036	0.0000
20	4.1121	0.5445	6.8084	1.6320	-2.6963	1.6495	23.9034	0.0000
21	4.2944	0.5994	6.8131	1.5571	-2.5187	1.6032	22.9808	0.0000

Tabloda yer alan farklara baktığımız zaman, 4 önermede (4, 3, 17, 16) açık 3'ün üzerindedir. En büyük SERVQUAL açıklık skoru -3.807 ile 4. önermede olup, bu önermede yer alan “Öğrenci işleri binası ve çalışma (bekleme) salonu yeterli büyüklüktedir” ifadesine müşteriler en olumsuz cevapları vermiştir. Binanın eski, öğrenci işleri bürosunun ufak ve ergonomik olmaması bu açığın en büyük nedenidir. İyileştirme zordur, zaman alıcıdır. Kurumun devlet kurumu olması işi daha da zorlaştırmaktadır. İkinci en büyük açıklık skoru -3.5467 ile 3. önermededir. “Öğrenci işleri binası ve çalışma (bekleme) salonu göze hoş görünür” ifadesine müşteriler olumsuz cevap vermişlerdir. Bu durum, binanın mimarisi ve dekorasyonunun müşterilerce yetersiz görüldüğüne işaret etmektedir. İyileştirme zor ve zaman alıcıdır. Üçüncü büyük fark -3.1636 ile 17. önermededir. “Öğrenci işleri çalışanları müşterilerin özel isteklerini anlar” ifadesine müşteriler olumsuz yaklaşmıştır. Dördüncü büyük fark -3.0317 ile 16. önermededir. “Öğrenci işleri, müşterilerin menfaatlerini her şeyin üstünde tutar” ifadesi de oldukça olumsuz yanıtlanmıştır. 17. ve 16. önermeler çalışan personelin davranış biçimiyle (duyarlılıkla) ilgilidir. Bu açıklığın azaltılması nispeten daha kolaydır. Personele yapılacak telkinlerle müşteriye karşı duyarlılık artırılabilir.

Diğer önermelerin 10 tanesinin açıklık skoru 2.5-3.0 arasındadır. Bunlar sırasıyla 15, 1, 10, 12, 11, 13, 20, 9, 19 ve 21 no'lu önermelerdir. 10, 11, 12 ve 13 no'lu önermeler personelin heveslilik boyutuna aittir. Personelin eğitimi ve katılım ile bu açıklık azaltılabilir. Kalan önermelerin açıklık skorları 2.0 – 2.5 arasındadır. Öğrenci işlerinin en iyi olduğu kalite seviyesi kayıtların hatasız tutulmasıdır (7. önerme). Bunun sebebi, yüksek okulun bir devlet kurumu olması dolayısıyla her işin kendine göre bir

prosedürünün bulunması, çalışanların bunlara sıkı bir şekilde uyması veya uymak zorunda olmasıdır.

Önermelerin standart sapma değerlerine bakıldığında, 4. önermenin algılanmasında (öğrenci işleri binası ve çalışma salonunun yeterliliği) standart sapmanın diğer önermelerinkine göre en yüksek olduğu (0.8308) görülmektedir. Beklentiler bazında ise, en yüksek standart sapma 1.8695 ile 10. önermedir (Öğrenci İşleri müşterinin bir sorunu olduğu zaman, sorunu çözmek için samimi ilgi gösterir). SERVQUAL önermeleri arasında yer alan 8. önerme (Öğrenci işleri kayıtların tutulmasında çok titizdir) müşterilerin değerlendirmelerinde en yüksek sapmayı içermektedir (1.9058).

Faktör bazındaki SERVQUAL açıklık skorları Tablo 6'da verilmiştir.

**Tablo 6. Faktörlerin Servqual skorları**

N=214 Önerme	ALGILAMA		BEKLENTİ		SKOR (Fark) (c)=(a-b)	SKOR Std. Sapma $\sqrt{\frac{\sum c_i^2 - n\bar{c}^2}{n-1}}$	t	p
	Ortalama (a)	Std. Sapma	Ortalama (b)	Std. Sapma				
1. Fiziksel Öz.	3.5281	0.6765	6.6846	1.6497	-3.1565	1.7565	26.2823	0.0000
2. Güvenilirlik	4.5355	0.4910	6.8383	1.8009	-2.3028	1.8194	18.5113	0.0000
3. Heveslilik	3.9638	0.5745	6.7594	1.8274	-2.7956	1.8547	22.0473	0.0000
4. Duyarlılık	3.8970	0.6609	6.7372	1.7228	-2.8388	1.7694	23.4612	0.0000
5. Yeterlilik	4.2757	0.5214	6.8142	1.6674	-2.5386	1.6574	22.4086	0.0000
<b>Genel</b>	<b>4.0400</b>	<b>0.5790</b>	<b>6.7667</b>	<b>1.7096</b>	<b>-2.7265</b>	<b>1.7715</b>	<b>22.5145</b>	<b>0.0000</b>

SERVQUAL açıklık skorunun en fazla olduğu faktör fiziksel özellikler faktörüdür (-3.1565). Fiziksel özellikler faktöründe iyileştirme yapmak çeşitli koşullara bağlı ve maliyetli olsa da, bu konuya olabildiğince el atılmalıdır. Fiziksel özelliklerden sonra en fazla açıklık skorunu -2.8388 ile duyarlılık faktörü vermektedir. Burada, öğrenci işleri çalışanlarının, müşterilere karşı daha duyarlı olabilmelerini sağlamak için, teorik ve uygulamalı eğitime tabi tutulmaları gerekmektedir. Üçüncü sırada -2.7956 ile heveslilik faktörü vardır. Bu faktör, çalışanların davranış biçimi ve kurum kültürü ile de ilgilidir. Düzeltilmesi ve iyileştirme yapılması daha kolay olan faktördür. Yeterlilik faktörü -2.5386 açıklık skoru ile dördüncü sırada yer almaktadır. Burada da eğitimin önemi ortaya çıkmaktadır. Çalışanlar önce kendi aralarında bu açığı kapatmaya çalışmalı, sonra kurumun diğer çalışanları ile de işbirliği ile bu açığı kapatmalıdır. Kurumsal bir yapı olmasından dolayı güvenilirlik faktörü en küçük açıklık skoruna (-2.3028) sahiptir. Ancak, bu seviye yine de yüksek sayılabilir. Genel açıklık skoru -2.7265 olup, söz konusu yüksek okuldaki öğrenci işleri hizmet kalitesinin beklentilerin çok gerisinde kaldığına işaret etmektedir.

### 3.6 Farkların Eşleştirilmiş T Testi

Hizmet nitelikleriyle ilgili farkın anlamlılığının test edilmesi eşleştirilmiş t test (paired sample t test) yoluyla gerçekleştirilmiştir (Tablo 7). Burada,  $H_0$  hipotezi algılama ve beklenti skorları arasında anlamlı fark olmadığını;  $H_1$  hipotezi ise anlamlı fark bulunduğunu ifade etmiştir. Testin sonucunda, çizelgede de yer aldığı gibi, farkların anlamlılık düzeyinin  $p < 0.01$  olduğu görülmüştür. Böylece, öğrenci işleri bölümünün sunmuş olduğu hizmetin kalitesiyle ilgili müşteri beklenti ve algılamaları arasındaki

fark anlamlı bulunmuştur. Müşterilerin beklentileri mevcut hizmet kalitesinin üzerindedir.

**Tablo 7. Farkların eşleştirilmiş T Test sonuçları**

		0.95 Güven Aralığı					
Ortalama	Std. Sapma	Std. Hata (Ort.)	Alt Sınır	Üst Sınır	t	df	Anlamlılık (2- yönlü)
2.7265	1.7715	0.1211	2.4739	2.9791	22.5145	20	0.0000

#### 4. SONUÇ

Hizmet sunan kuruluşlarda hizmet kalitesi düzeyi kritik bir performans ölçütü olarak görülmekte, hizmet kalitesi ölçümü uygulamalarda öne çıkan konular arasında yer almaya devam etmektedir. Yüksek hizmet kalitesinin müşteri memnuniyeti, bağlılık, müşterilerin olumlu tavsiyeleri ve şikayetlerde azalmalar gibi arzu edilen sonuçları beraberinde getirmesi bu yöndeki gelişmeleri hızlandıran başlıca nedenlerdir. Bu konu yalnızca ticari kuruluşlar için değil, hizmet sunan ve ticari olmayan tüm kuruluş ve kurumlar için de önemlidir. Bu çalışmada bunun tipik bir uygulama örneği verilmiştir.

Araştırma sonuçlarına göre, öğrenci işleri müşterilerinin algılama düzeylerinin yüksek çıkmasına rağmen, beklentilerinin de yüksekliği SERVQUAL açıklık skorlarını negatif büyüklüğe taşımıştır. Tüm faktörlerde negatif skorlar ortaya çıkmıştır. Fiziksel özellikler faktörü en yüksek negatif SERVQUAL açıklık skoruna (-3.1565) sahip olduğu için öncelikle iyileştirilmesi gereken bir konudur. Duyarlılık faktörü -2.8388 açıklık skoru ile ikinci sırada önem arz etmektedir. Bu faktörde iyileştirme yapabilmek, öğrenci işleri çalışanlarının müşterilere karşı daha duyarlı olabilmelerini sağlamakla mümkündür. Bunun için, de teorik ve uygulamalı eğitim almalarına ihtiyaç vardır. Üçüncü sırada -2.7956 açıklık skoru ile yer alan heveslilik faktörü çalışanların davranış biçimi ve kurum kültürü ile ilgili olduğundan, kendilerine bu yönde bilgilendirme yapılması yeterli olacaktır. Yeterlilik faktörü -2.5386 açıklık skoru ile dördüncü sırada yer almaktadır. Burada da eğitimin önemi ortaya çıkmaktadır. Çalışanların kişisel ve grup çalışmaları ve diğer çalışanlarla işbirliği yaparak bu açığı kapatmaları mümkündür. Kurumsal bir yapı olmasından dolayı güvenilirlik faktörü en küçük açığa (-2.3028) sahip olmasına rağmen yine de yüksek seviyededir.

Toplam SERVQUAL açıklık skoru -2.7265 bulunmuştur. Bu değer oldukça büyük olup, öğrenci işleri bölümünün okulda verdiği hizmet seviyesinin hizmet alanlarca yeterli bulunmadığını göstermektedir.

Belirli aralıklarla bu tip anket çalışmalarının tekrarlanması ve yapılacak hizmet iyileştirmelerine bağlı olarak, zamanla müşteri memnuniyetinin de yükselmesi sağlanabilecektir.

## 5. KAYNAKLAR

- Armstrong, G., Kotler, P., 2000. Marketing: An Introduction, Prentice Hall, Inc., Upper Saddle River, New Jersey, 246.
- Zeithaml, V. A., Parasumaran, V., Berry, L. L., 1990. Delivering Quality Service: Balancing Customer Perceptions and Expectations, The Free Press, New York, 17.
- Zeithaml, V. A., Bitner, M.J. (2006), Services Marketing: Integrating Customer Focus Across the Firm, 4th ed., McGraw-Hill Co., Boston.
- Parasumaran, A., Zeithaml, V. A., Berry, L. L., 1991. Refinement and Reassessment of the Servqual Scale, Journal of Retailing, 67(4), 445.
- Knutson, B. J., Stevens, P., Wullaert, C., Patton, M., Yokoyama, F., 1990. LODGSERV: a Service Quality Index for the Lodging Industry, Hospitality Research Journal, 14, 2, 1-16.
- Stevens, P., Knutson, B. J., Patton, M., 1995. DINESERV: a Tool for Measuring Service Quality in Restaurants, Cornell Hotel and Restaurant Administration Quarterly, 56-60.
- Nitecki, D. A., 1996. Changing the Concept and Measure of Service Quality in Academic Libraries, The Journal of Academic Librarianship (May), 181-190.
- Devebakan, N., Aksaraylı, M., 2003. Sağlık İşletmelerinde Algılanan Hizmet Kalitesinin Ölçümünde SERVQUAL Skorlarının Kullanımı ve Özel Altınordu Hastanesi Uygulaması, Dokuz Eylül Üniversitesi, SBE Dergisi, 5(1). 38-54.
- Altan, Ş., Atan, M., 2004. Bankacılık Sektöründe Toplam Hizmet Kalitesinin SERVQUAL Analizi ile Ölçümü, Gazi Üniversitesi İ.İ.B.F.Dergisi, 1, 17-32.
- Atan, M., Baş, M., Tolon, M., 2006. "SERVQUAL Analizi ile Migros ve Gima Süpermarketlerinde Hizmet Kalitesinin Ölçülmesine Yönelik Bir Alan Çalışması", Gazi Üniversitesi İ.İ.B.F. Dergisi, 7(2), 256-270.
- Ruzan, E. F., Ruzan, A. R. F., Bakkurt, A., 2007. Ege İhracatçılar Birliklerinde Servqual Ölçeği ile Hizmet Kalitesinin Ölçülmesi, Yönetim ve Ekonomi, 14(2), Celal Bayar Üniversitesi, İİBF, 137-153.
- Okumuş, A., Asil, H., 2007. Hizmet Kalitesi Algılamasının Havayolu Yolcularının Genel Memnuniyet Düzeylerine Göre Olan Etkisinin İncelenmesi, İstanbul Üniversitesi İşletme Fakültesi İşletme Dergisi, 36(2) 7-29.
- Albacete-S, C. A., Fuentes-F, M. M., Lloréns, M. F. J., 2007. Service Quality Measurement in Rural Accommodation, Annals of Tourism Research, 34(1), 45-65.
- Eleren, A., Bektaş, Ç., Görmüş, A. Ç., 2007. Hizmet Sektöründe Hizmet Kalitesinin SERVQUAL Yöntemi ile Ölçülmesi ve Hazır Yemek İşletmesinde Bir Uygulama, Finans Politik & Ekonomik Yorumlar, 44(514), 75-88.



Aydın, K., 2007. Hizmet İşletmelerinde Servqual Yöntemi ile Hizmet Kalitesinin Ölçümü ve Kocaeli’ndeki Seyahat İşletmelerinden Efe Tur Uygulaması. Sosyal Siyaset Konferansları, 50. Kitap, 1101-1130.

Maddern, H., Maull, R., Smart, A., Baker, P., 2007. Customer Satisfaction and Service Quality in UK Financial Services, International Journal of Production and Operations Management, 27(9), 999-1019.

Yılmaz, V., Filiz, Z., Yaprak, B., 2007. Servqual Yöntemiyle Yüksek Öğretimde Hizmet Kalitesinin Ölçülmesi, Anadolu Üniversitesi, Sosyal Bilimler Dergisi, 7(1), 299-316.

Lai, F., Hutchinson, J., Li, D., Bai, C., 2007. An Empirical Assessment and Application of SERVQUAL in Mainland China’s Mobile Communications Industry, International Journal of Quality & Reliability Management, 24(3), .244 – 262.

Ersöz, S., Pınarbaşı, M., Türker, A. K., Yüzükırmızı, M., 2009. Hizmet Kalitesinin Servqual Metodu ile Ölçümü ve Sonuçların Yapısal Eşitlik Modelleri ile Analizi: Öğretmen Evi Uygulaması. International Journal of Eng. Research & Development, 1(1), 19-27.

Zaim, H., Bayburt, N., Zaim, S., 2009. Service Quality and Determinants of Customer Satisfaction in Hospitals: Turkish Experience.2009 EABR & TLC Conference Proceedings, June 11, Prague, Czech Republic.

Chen, K. K., Chang, C. T., Lai, C. S., 2009. Service Quality Gaps of Business Customers in the Shipping Industry, Transportation Research Part E, 45, 222-237.

Filiz, Z., Yılmaz, V., Yağız, C., 2010. Belediyelerde Hizmet Kalitesinin Servqual Analizi ile Ölçümü: Eskişehir Belediyelerinde Bir Uygulama, Anadolu Üniversitesi Sosyal Bilimler Dergisi, 10(3), 59-76.

Filiz, Z., 2011. Servqual Yöntemiyle Yurt İşletmesinde Hizmet Kalitesinin Ölçülmesi, International Journal of Research and Development, 3(1), 38-48.

Abouchedid, K., Nasser, R., 2002. Assuring Quality Service in Higher Education: Registration and Advising Attitudes in a Private University in Lebanon, Quality Assurance in Education, 10(4), 198-206.

Kitcharoen, K., 2004. The Importance-Performance Analysis of Service Quality in Administrative Departments of Private Universities in Thailand, ABAC Journal, 24(3), 20-46.

Yaş, O., 2009. Uzunyol MYO Öğrenci İşleri Departmanı Servqual Uygulaması, Kocaeli Üniversitesi Endüstri Mühendisliği Bölümü, Bitirme Tezi, 22-47.

Cronbach’s Alpha, 2012. Erişim: 14 Nisan 2006, [http://en.wikipedia.org/wiki/Cronbach's\\_alpha](http://en.wikipedia.org/wiki/Cronbach's_alpha).

## MEASURING THE SERVICE QUALITY OF STUDENT AFFAIRS AT HIGHER EDUCATION AND AN EXEMPLARY APPLICATION

### ABSTRACT

*Due to its main characteristics of intangibility, inseparability, heterogeneity and perishability of service, it is generally quite difficult to describe and measure the service quality. Several methods have been developed to measure service quality in application areas. SERVQUAL is one of these methods. In this study this method is used to measure the service quality of student affairs at higher education and the gaps between expected service and perceived performance of the service by the beneficiaries are derived. The data is collected by the survey method during the application. After the measurement, although the level of customer perceptions is high, the service quality level provided by the student affairs office is found to be insufficient by its customers due to their high expectations.*

**Keywords:** Service quality, SERVQUAL, Data reliability, Principle component analysis.

## Ek: Anket Soruları

	BEKLENTİ	ALGILAMA
<i>Fiziksel Özellikler</i>		
1. Öğrenci İşleri yeterli ve modern donanıma (masa, banko, ekipman, kırtasiye, vb.) sahiptir.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
2. Öğrenci İşleri çalışanları iyi görünümüne sahiptir.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
3. Öğrenci İşleri binası ve çalışma / bekleme salonu göze hoş görünür.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
4. Öğrenci İşleri binası ve çalışma / bekleme salonu yeterli büyüklüktedir.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
<i>Güvenilirlik</i>		
5. Öğrenci İşleri güvenilirdir. Hizmeti ilk anda ve zamanında yerine getirir.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
6. Öğrenci İşleri, işlemlerinde (öğrenci belgesi, not belgesi, askerlik belgesi verilmesi; staj takip işlemleri, mezuniyet işlemleri, vb.) hata yapmaz.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
7. Öğrenci İşleri bir şey yapmak için kesin bir zaman verdiğinde, söz verdiği gibi yapar.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
8. Öğrenci İşleri kayıtların tutulmasında çok titizdir.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
9. Öğrenci İşleri müşterileri öğrenci işleriyle ilgili işlemlerde kendilerini güvende hissederler.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
<i>Heveslilik</i>		
10. Öğrenci İşleri müşterinin bir sorunu olduğu zaman, sorunu çözmek için samimi ilgi gösterir.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
11. Öğrenci İşleri çalışanları hızlı hizmet verirler.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
12. Öğrenci İşleri çalışanları her zaman müşteriye yardım etmeye istekli ve gönüllüdürler.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
13. Öğrenci İşleri çalışanları asla müşterilerin ricalarını cevaplamayacak kadar meşgul değildirlir.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
<i>Duyarlılık</i>		
14. Öğrenci İşleri çalışma saatleri bütün müşterilere uygun olacak şekildedir.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
15. Öğrenci İşleri her müşteriyle kişisel olarak ilgilenecek çalışanlara sahiptir.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
16. Öğrenci İşleri müşterilerinin menfaatlerini her şeyin üstünde tutar.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
17. Öğrenci İşleri çalışanları müşterilerin özel isteklerini anlar.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
<i>Yeterlilik</i>		
18. Öğrenci İşleri çalışanları müşterilerin sorularına cevap verecek bilgiye sahiptir.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
19. Öğrenci İşleri çalışanları naziktir.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
20. Öğrenci İşleri çalışanlarının davranışları müşteride güven duygusu yaratır.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
21. Öğrenci İşlerinin müşterileri hizmetle ilgili işlemlerde kendilerini güvende hissederler.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
22. Öğrenci İşlerinin kullandığı teknolojik donanım yeterlidir.	1 2 3 4 5 6 7	1 2 3 4 5 6 7

Not: 22. önerme 5. faktör gruplamasına uymadığı için değerlendirilmeden çıkartılmıştır.

## NEYMAN, PÓLYA-AEPLI VE THOMAS DAĞILIMLARININ MOMENTLERİ ÜZERİNE BİR İNCELEME

Gamze ÖZEL\*

### ÖZET

*Neyman A, B tipi dağılımlar, Pólya-Aepli ve Thomas dağılımları, hem olasılık kuramında hem de biyoloji, sismoloji, risk kuramı, meteoroloji gibi birçok uygulama alanında önem taşımaktadır. Bu dağılımlar üzerine birçok çalışma yapılmasına karşın, olasılık fonksiyonlarının kapalı biçimlerine ulaşılabilmesi kullanımlarını da kısıtlamaktadır. Bu nedenle dağılımlara ait merkezsiz, merkezsiz olmayan, faktöriyel momentler ve kümülanlar gibi moment karakteristikleri önem kazanmaktadır. Bu çalışmada Neyman tipi dağılımlar, Pólya-Aepli dağılımı ve Thomas dağılımı açıklanmış; dağılımlara ait merkezsiz, merkezsiz olmayan, faktöriyel momentler ve kümülanlar elde edilmiştir.*

**Anahtar Kelimeler:** B tipi dağılımlar, Kümülanlar, Momentler, Neyman A, Pólya-Aepli dağılımı, Thomas dağılımı.

### 1. GİRİŞ

Neyman A, B tipi dağılımlar, (Neyman, 1939) tarafından belirli bir alandaki organizmaların dağılımını araştırmak; Pólya-Aepli (geometrik-Poisson) dağılımı bir DNA ipliğindeki hasarların dağılımı incelemek (Gudowska vd., 2007) ve Thomas dağılımı (Thomas, 1949) tarafından belirli bir bölgedeki bitkilerin yayılmasını belirlemek amacıyla tanımlanmıştır. Bu dağılımlar genellikle trafik kazaları, çevrebilimi, nöropsikoloji, radyobiyojoloji, kalite kontrolü ve telekomünikasyonda kullanılmaktadır (Chen vd., 2005), (Gudowska vd., 2007), (Meintanis, 2007), (Özel vd., 2010), (Randolph vd., 1995), (Robin, 2002).  $N$ ,  $\lambda$  parametresi ile Poisson dağılımına sahip bir raslantı değişkeni olsun ve ortaya çıkan her olaya  $Y_i$ ,  $i=1,2,\dots,N$  ile gösterilen aynı dağılımlı, bağımsız ve pozitif değerler alan raslantı değişkenleri bağlansın. Bu raslantı değişkenleri  $N$  raslantı değişkeninden de bağımsız olsunlar. Buna göre,

$$X = \sum_{i=1}^N Y_i \quad (1)$$

biçimindeki  $X$  raslantı değişkeni birleşik Poisson dağılımına sahiptir.

Özel olarak, (1) eşitliğinde  $Y_i$ ,  $i=1,2,3,\dots$ , raslantı değişkenleri Poisson dağılımlı ise,  $X$  raslantı değişkeni Neyman A tipi dağılıma;  $Y_i$ ,  $i=1,2,3,\dots$ , raslantı değişkenleri ikiterimli (binom) dağılımlı ise,  $X$  raslantı değişkeni Neyman B tipi dağılıma sahiptir. Benzer biçimde,  $Y_i$ ,  $i=1,2,\dots$ , raslantı değişkenleri geometrik dağılımlı ise,  $X$  raslantı değişkeni Pólya-Aepli dağılımına ve  $Y_i$ ,  $i=1,2,\dots$ , raslantı değişkenleri kaydırılmış (shifted) Poisson dağılımına sahip ise,  $X$  raslantı değişkeni Thomas dağılımına sahip olur. Neyman A, B tipi dağılımlar, Pólya-Aepli ve Thomas dağılımı üzerine birçok çalışma yapılmasına karşın, bu dağılımların olasılık fonksiyonlarının kapalı biçimlerine

\*Doç. Dr., Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara, e-posta: [gamzeozk@hacettepe.edu.tr](mailto:gamzeozk@hacettepe.edu.tr)

ulaşılamamıştır (Özel, 2012). Bu nedenle, dağılımların merkezsel, merkezsel olmayan, faktöriyel momentlerinin ve kümülanlarının kullanılması önem kazanmaktadır. Bu çalışmada Neyman tipi dağılımlar, Pólya-Aeppli dağılımı ve Thomas dağılımı açıklanmış; dağılımlara ait merkezsel, merkezsel olmayan, faktöriyel momentler ve kümülanlar elde edilmiştir.

## 2. NEYMAN, PÓLYA-AEPLI VE THOMAS DAĞILIMLARI

$N$  raslantı değişkeni  $\lambda$  ( $\lambda > 0$ ) parametresi ile Poisson dağılımına sahip olsun.  $Y_i^{(1)}$ ,  $i=1,2,3,\dots$ , raslantı değişkenleri  $\nu$  ( $\nu > 0$ ) parametresi ile Poisson dağılımlı ise, Neyman A tipi dağılımlı  $X_{(1)}$ 'in olasılık fonksiyonu için,

$$p_{X_{(1)}}(k) = P(X_{(1)} = k) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} e^{-\nu n} \frac{(\nu n)^k}{k!}, \quad k = 0, 1, 2, \dots \quad (2)$$

eşitliği;  $Y_i^{(2)}$ ,  $i=1,2,3,\dots$ , raslantı değişkenleri  $m$  ve  $p$  parametreleri ile ikiterimli dağılımlı ise, Neyman B tipi dağılımlı  $X_{(2)}$ 'nin olasılık fonksiyonu için,

$$p_{X_{(2)}}(k) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \binom{nm}{k} p^k (1-p)^{nm-k}, \quad k = 0, 1, 2, \dots \quad (3)$$

eşitliği ve  $Y_i^{(3)}$ ,  $i=1,2,3,\dots$ , raslantı değişkenleri  $\theta$  parametresi ile geometrik dağılıma sahip ise, Pólya-Aeppli dağılımlı  $X_{(3)}$ 'ün olasılık fonksiyonu aşağıdaki gibi yazılabilir:

$$p_{X_{(3)}}(k) = \sum_{n=1}^{\infty} \binom{k+n-1}{k} \theta^n (1-\theta)^k e^{-\lambda} \frac{\lambda^n}{n!}, \quad k = 0, 1, 2, \dots \quad (4)$$

$\alpha > 0$  için  $Y_i^{(4)}$ ,  $i=1,2,3,\dots$ , raslantı değişkenleri  $p_{Y_i^{(4)}}(j) = e^{-\alpha} \alpha^{j-1} / (j-1)!$ ,  $j=1,2,3,\dots$  biçiminde verilen kaydırılmış Poisson dağılımlı ise, Thomas dağılımlı  $X_{(4)}$ 'ün olasılık fonksiyonu için,

$$p_{X_{(4)}}(k) = \sum_{n=0}^k e^{-\lambda} \frac{\lambda^n}{n!} e^{-\alpha n} \frac{(\alpha n)^{k-n}}{(k-n)!}, \quad k = n, n+1, n+2, \dots \quad (5)$$

eşitliği yazılabilir. Ancak, (2), (3), (4) ve (5) eşitliklerinden olasılıklara ulaşmak güçtür. Panjer [8] tarafından (1) eşitliğindeki  $N$  raslantı değişkeninin olasılık fonksiyonunun  $p_N(n) = \frac{\lambda}{n} p_N(n-1)$ ,  $n=1,2,3,\dots$  ilişkisini sağlaması durumunda birleşik Poisson dağılımı için  $p_X(0) = e^{-\lambda(1-p_Y(0))}$  ve  $p_X(k) = \lambda \sum_{i=1}^k \frac{i}{k} p_Y(i) p_X(k-i)$ ,  $k=1,2,3,\dots$  eşitliği elde edilmiştir. Burada,  $p_Y(y)$ ,  $Y_i$ ,  $i=1,2,3,\dots$ , raslantı değişkenlerinin olasılık

fonksiyonunu göstermektedir. Neyman tipi dağılımlar, Pólya-Aeppli dağılımı ve Thomas dağılımı için Panjer eşitlikleri de kullanılabilir. Ayrıca  $p_X(k) = P(X = k)$ ,  $k = 0, 1, 2, \dots$ , olasılıkları, yinelemeli olasılıklara dayanmadan Özel ve İnal (Özel vd., 2008) tarafından ve Neyman A, B tipi dağılımlar, Pólya-Aeppli dağılımı ve Thomas dağılımının olasılık fonksiyonları Özel ve İnal (Özel vd., 2012) tarafından elde edilmiştir.

### 3. NEYMAN, PÓLYA-AEPLI VE THOMAS DAĞILIMLARININ MOMENT KARAKTERİSTİKLERİ

$Y_i$ ,  $i = 1, 2, \dots$ , aynı dağılımlı, bağımsız, kesikli raslantı değişkenleri  $j = 0, 1, 2, \dots$ , değerlerini  $P(Y_i = j) = p_j$  olasılıkları ile alsın ve  $\lambda_j = \lambda p_j$  olsun. Buna göre,  $Y_i$ ,  $i = 1, 2, \dots$ , raslantı değişkenlerinin moment çıkaran fonksiyonu  $M_Y(s) = p_0 + p_1 e^s + \dots + p_m e^{s^m}$  biçimindedir. (1) eşitliğindeki birleşik Poisson dağılımlı  $X$  r.d.'nin moment çıkaran fonksiyonu,

$$\begin{aligned} M_X(s) &= e^{\lambda[M_Y(s)-1]} \\ &= e^{\lambda[(p_0 + p_1 e^s + \dots + p_m e^{s^m}) - 1]} \\ &= e^{\lambda p_0 + \lambda p_1 e^s + \dots + \lambda p_m e^{s^m}} e^{-\lambda} \\ &= e^{-\lambda(1-p_0)} e^{\lambda_1 e^s + \lambda_2 (e^s)^2 + \dots + \lambda_m (e^s)^m} \end{aligned} \quad (6)$$

olur (Özel vd., 2008). Burada,  $p_0 = P(Y_i = 0)$  olarak tanımlıdır.

Özel olarak,  $Y_i^{(r)}$ ,  $i = 1, 2, 3, \dots$ ,  $r = 1, 2, 3, 4$ , raslantı değişkenleri sırasıyla Neyman A, B, Pólya-Aeppli ve Thomas dağılımına sahip olsun. (1) eşitliğindeki birleşik Poisson dağılımının merkezsel olmayan momentlerine ulaşmak için  $M_X(s)$ 'nin  $k$ . türevi alınrsa,

$$\mu'_k = E[X^k] = \left. \frac{\partial^k M_X(s)}{\partial s^k} \right|_{s=0}, \quad k = 1, 2, \dots \quad \text{elde edilir. Buna göre, } X \text{ raslantı}$$

değişkeninin merkezsel olmayan momentleri aşağıdaki gibidir:

$$\begin{aligned} \mu'_1 &= \lambda \mu'_Y(1), \\ \mu'_2 &= (\lambda \mu'_Y(1))^2 + (\lambda \mu'_Y(2)), \\ \mu'_3 &= (\lambda \mu'_Y(1))^3 + 3(\lambda \mu'_Y(2))(\lambda \mu'_Y(1)) + (\lambda \mu'_Y(3)), \\ &\vdots \end{aligned} \quad (7)$$

Burada,  $\mu'_Y(k) = E[Y^k]$ ,  $Y_i$ ,  $i = 1, 2, \dots$ , raslantı değişkenlerinin merkezsel olmayan momentleridir. Neyman A, B, Pólya-Aeppli ve Thomas dağılımlarına ait bazı merkezsel olmayan momentler Tablo 1'de verilmiştir:

**Tablo 1. Neyman A, B, Pólya-Aeppli ve Thomas Dağılımına ait bazı merkezel olmayan momentler**

	<b>Neyman A</b>	<b>Neyman B</b>	<b>Pólya-Aeppli</b>	<b>Thomas</b>
$\mu'_1$	$\lambda v$	$\lambda mp$	$\lambda(1-\theta)/\theta$	$\lambda(1+\alpha)$
$\mu'_2$	$(\lambda v)^2 + \lambda(v+v^2)$	$(\lambda mp)^2 + (\lambda mp) + [\lambda m(m-1)p^2]$	$[\lambda(1-\theta)/\theta]^2 + [\lambda(1-\theta)(2-\theta)/\theta^2]$	$[\lambda(1+\alpha)]^2 + \lambda\alpha(2+\alpha)$
$\mu'_3$	$(\lambda v)^3 + 3(\lambda v)[\lambda(v+v^2)] + \lambda(v+3v^2+v^3)$	$(\lambda mp)^3 + 3(\lambda mp)^2 + (\lambda mp) + 3[\lambda m(m-1)p^2](1+\lambda mp) + [\lambda m(m-1)(m-2)p^3]$	$[\lambda(1-\theta)/\theta]^3 + 3[\lambda(1-\theta)/\theta][\lambda(1-\theta)(2-\theta)/\theta^2] + [\lambda(1-\theta)[6+\theta(\theta-6)]/\theta^3]$	$\lambda\alpha(1+\alpha) + [\alpha + 2\lambda(2+\alpha)]$

(1) eşitliğindeki birleşik Poisson dağılımının merkezel momentlerine ulaşmak için (6) eşitliğinden yararlanarak elde edilen merkezel moment çıkararak fonksiyon

$$M_{X-\mu}(t) = e^{-\mu t} M_X(t) \text{ 'nin } k. \text{ türevi alınır, } \mu_k = E[(X-\mu)^k] = \left. \frac{\partial^k M_{X-\mu}(s)}{\partial s^k} \right|_{s=0}, \quad k = 1, 2, \dots$$

olur. Buna göre, X raslantı değişkeninin merkezel momentleri aşağıdaki gibidir:

$$\begin{aligned} \mu_1 &= [\mu + \lambda \mu'_Y(1)], \\ \mu_2 &= [\mu + \lambda \mu'_Y(1)]^2 + \lambda \mu'_Y(2), \\ \mu_3 &= [\mu + \lambda \mu'_Y(1)]^3 + 3\lambda \mu'_Y(2)[\mu + \lambda \mu'_Y(1)] + \lambda \mu'_Y(3), \\ &\vdots \end{aligned} \tag{8}$$

Neyman A, B, Pólya-Aeppli ve Thomas dağılımlarına ait bazı merkezel momentler Tablo 2’de verilmiştir:

**Tablo 2. Neyman A, B, Pólya-Aeppli ve Thomas Dağılımına ait bazı merkezel momentler**

	<b>Neyman A</b>	<b>Neyman B</b>	<b>Pólya-Aeppli</b>	<b>Thomas</b>
$\mu_1$	$(\mu + \lambda v)$	$(\mu + \lambda mp)$	$[\mu + \lambda(1-\theta)/\theta]$	$[\mu + \lambda(1+\alpha)]$
$\mu_2$	$(\mu + \lambda v)^2 + \lambda(v+v^2)$	$(\mu + \lambda mp)^2 + \lambda(mp + m(m-1)p^2)$	$[\mu + \lambda(1-\theta)/\theta]^2 + [\lambda(1-\theta)(2-\theta)/\theta^2]$	$[\mu + \lambda(1+\alpha)]^2 + \lambda\alpha(2+\alpha)$
$\mu_3$	$(\mu + \lambda v)^3 + \lambda(v+3v^2+v^3) + 3\lambda(\mu + \lambda v)(v+v^2)$	$(\mu + \lambda mp)^3 + 3\lambda(\mu + \lambda mp)(mp + m(m-1)p^2) + \lambda(mp + 3m(m-1)p^2) + m(m-1)(m-2)p^3$	$[\mu + \lambda(1-\theta)/\theta]^3 + 3[\mu + \lambda(1-\theta)/\theta][\lambda(1-\theta)(2-\theta)/\theta^2] + [\lambda(1-\theta)[6+\theta(\theta-6)]/\theta^3]$	$[\mu + \lambda(1+\alpha)]^3 + 3[\mu + \lambda(1+\alpha)]\lambda(\alpha^2 + \alpha - 1) + \lambda\alpha^3 + 5\alpha^2 + 3\alpha - 2$

$Y_i, i = 1, 2, \dots$ , raslantı değişkenlerinin olasılık çıkararak fonksiyonu  $g_Y(s)$  ise, (1) eşitliğindeki birleşik Poisson dağılımı için kümülan çıkararak fonksiyon aşağıdaki gibi elde edilmiştir:

$$C_X(s) = \lambda[(p_0 - 1)(p_1 e^s + p_2 e^{2s} + \dots + p_m e^{ms})] \tag{9}$$

Buradan  $\mu'_r = \mu'_r(Y) = E[Y^r]$ ,  $Y_i$ ,  $i = 1, 2, \dots$ , raslantı değişkenlerinin merkezsel olmayan momentleri olmak üzere, birleşik Poisson dağılımı için kümülanlar aşağıdaki gibi elde edilmiştir:

$$\kappa_r = \lambda \mu'_r \quad (10)$$

Neyman A, B dağılımları, Pólya-Aeppli ve Thomas dağılımına ait bazı kümülanlar Tablo 3'te yer almaktadır:

**Tablo 3. Neyman A, B, Pólya-Aeppli ve Thomas Dağılımına ait bazı kümülanlar**

	<b>Neyman A</b>	<b>Neyman B</b>	<b>Pólya-Aeppli</b>	<b>Thomas</b>
$\kappa_1$	$\lambda v$	$\lambda mp$	$\lambda(1-\theta)/\theta$	$\lambda(\alpha-1)$
$\kappa_2$	$\lambda[v+v^2]$	$\lambda[mp+m(m-1)p^2]$	$\lambda(2-\theta)(1-\theta)/\theta^2$	$\lambda(\alpha^2+\alpha-1)$
$\kappa_3$	$\lambda[v+3v^2+v^3]$	$\lambda[mp+3m(m-1)p^2+m(m-1)(m-2)p^3]$	$\lambda(1-\theta)[6+(\theta-6)\theta]/\theta^3$	$\lambda(\alpha^3+5\alpha^2+3\alpha-2)$

(1) eşitliğindeki birleşik Poisson dağılımının faktöriyel momentlerine ulaşmak için

$$g_X(s) \text{ 'nin } k. \text{ türevi alınırsa, } \mu'_X[k] = E[X(X-1)\dots(X-(k-1))] = \left. \frac{\partial^k g_X(s)}{\partial s^k} \right|_{s=1}, \quad k = 1, 2, \dots$$

elde edilir. Buna göre, X raslantı değişkeninin faktöriyel momentleri aşağıdaki gibi elde edilmiştir:

$$\begin{aligned} \mu'_X[1] &= \lambda \mu'_Y[1], \\ \mu'_X[2] &= (\lambda \mu'_Y[1])^2 + (\lambda \mu'_Y[2]), \\ \mu'_X[3] &= (\lambda \mu'_Y[1])^3 + 3(\lambda \mu'_Y[2])(\lambda \mu'_Y[1]) + (\lambda \mu'_Y[3]), \\ &\vdots \end{aligned} \quad (11)$$

Burada,  $\mu'_Y[r] = E[Y(Y-1)\dots(Y-(r-1))]$ ,  $Y_i$ ,  $i = 1, 2, \dots$ , raslantı değişkenlerinin faktöriyel momentleridir. Neyman A, B dağılımları, Pólya-Aeppli ve Thomas dağılımına ait bazı faktöriyel momentler Tablo 4'te yer almaktadır:

**Tablo 4. Neyman A, B, Pólya-Aeppli ve Thomas Dağılımına ait bazı faktöriyel momentler**

	<b>Neyman A</b>	<b>Neyman B</b>	<b>Pólya-Aeppli</b>	<b>Thomas</b>
$\mu'_X[1]$	$\lambda v$	$\lambda mp$	$\lambda(1-\theta)/\theta$	$\lambda(1+\alpha)$
$\mu'_X[2]$	$(\lambda v)^2 + \lambda v$	$(\lambda mp)^2 + [\lambda m(m-1)p^2]$	$[\lambda(1-\theta)/\theta]^2 + [\lambda(1-\theta)/\theta^2]$	$[\lambda(1+\alpha)]^2 + \lambda(1+\alpha)(2+\alpha)$
$\mu'_X[3]$	$(\lambda v)^3 + 3(\lambda v)^2 + \lambda v$	$(\lambda mp)^3 + 3[\lambda m(m-1)p^2] + (\lambda mp) + [\lambda m(m-1)(m-2)p^3]$	$[\lambda(1-\theta)/\theta]^3 + 3[\lambda(1-\theta)/\theta^2] + [\lambda(1-\theta)/\theta] + [\lambda(1-\theta) + \lambda(1-\theta)^2]/\theta^4$	$[\lambda(1+\alpha)]^3 + [\lambda(1+\alpha)(2+\alpha)] + [\lambda(1+\alpha)(2+\alpha)^2 + (3+\alpha)]$



#### 4. SONUÇ

Neyman A, B tipi dağılımların, Pólya-Aeppli ve Thomas dağılımının istatistiksel önemi gerçek hayatta kullanılabilirliklerinden kaynaklanmaktadır. Neyman A, B tipi dağılımlar, Pólya-Aeppli ve Thomas dağılımı kullanılarak birçok çalışma yapılmasına karşın, bu dağılımların momentleri üzerine yapılan çalışmalar sınırlıdır. Bu çalışmada Neyman tipi dağılımlar, açıklanmış; dağılımlara ait merkezsiz, merkezsiz olmayan, faktöriyel momentler ve kümülanlar elde edilmiştir. Böylece bu dağılımların çevrebilimi, sismoloji, risk kuramı, biyoloji vb. birçok alanda etkin biçimde kullanılabilirliği sağlanabilir.

#### 5. KAYNAKLAR

Chen, C. W., Randolph, P., Tian-Shy, L., 2005. Using CUSUM Control Schemes for Monitoring Quality Levels in Compound Poisson Production Environment: the Geometric Poisson Process. *Quality Engineering*, 17. 2. 207-217.

Gudowska-Nowak, E., Lee, R., Nasonova, E., Ritter, S., Scholz, M., 2007. Effect of LET and Track Structure on the Statistical Distribution of Chromosome Aberrations. *Advances in Space Research*. 39. 1070–1075.

Meintanis, S. G., 2007. A New Goodness of Fit Test for Certain Bivariate Distributions Applicable to Traffic Accidents. *Statistical Methodology*. 4. 22-34.

Neyman, J., 1939. On a New Class of Contagious Distributions Applicable in Entomology and Bacteriology, *Annals of Mathematical Statistics*. 10. 35-57.

Özel G., İnal C., 2008. The Probability Function of the Compound Poisson Process and an Application to Aftershock Sequences. *Environmetrics*. 19. 79–85.

Özel, G., İnal, C., 2010. The Probability Function of a Geometric Poisson Distribution. *Journal of Statistical Computation and Simulation*, 80. 5. 479-487.

Özel, G., İnal, C., 2012. On the Probability Function of the First Exit Time for Generalized Poisson Processes, *Pakistan Journal of Statistics*. 28. 4. Basımda.

Panjer, H., 1981. Recursive Evaluation of a Family of Compound Distributions. *ASTIN Bulletin*. 12. 22-26.

Randolph, P., Sahinoglu, M., 1995. A Stopping Rule for a Compound Poisson Random Variable, *Applied Stochastic Models and Data Analysis*. 11. 2. 135-143.

Robin, S., 2002. A Compound Poisson Model for Word Occurrences in DNA Sequences, *Applied Statistics*. 51. 4. 437-451.

Thomas, M., 1949. A Generalization of Poisson's Binomial Limit for Use in Ecology, *Biometrika*, 36. 18-25.

## AN INVESTIGATION ON THE MOMENTS OF NEYMAN, PÓLYA-AEPPLI AND THOMAS DISTRIBUTIONS

### ABSTRACT

*Neyman type A, B distributions, Polya-Aeppli and Thomas distributions play important roles both in probability theory itself and its applications in such as biology, seismology, risk theory, and meteorology. Although there have been many studies on these distributions, the non-existence of closed forms for their probability functions restrict their usage. Hence, the moment characteristics of these distributions, such as central, non-central, factorial moments and cumulants, play an important role. In this study, Neyman type distributions, Polya-Aeppli distribution and Thomas distribution are explained; their central, non-central, factorial moments and cumulants are derived.*

**Keywords:** B distributions, Cumulants, Moments, Neyman type A, Polya-Aeppli distribution, Thomas distribution.

# BİREYLERİN KREDİ KARTLARINI İPTAL ETTİRME NİYETLERİNİN İNCELENMESİNDE LOJİSTİK REGRESYON ANALİZİ

Gültekin ATALIK\*

Y. Murat BULUT\*\*

Zeki YILDIZ\*\*\*

## ÖZET

*Lojistik regresyon; bağımlı değişkenin kategorik ikili, üçlü, çoklu kategorilerde gözlemlendiğinde bağımsız değişken ya da değişkenlerle sonuç ilişkisini belirlemede kullanılan bir yöntemdir. Bu çalışmada, bireylerin kredi kartlarını iptal ettirmesi bakımından sınıflandırılmasında lojistik regresyon analizinden yararlanılmıştır.*

**Anahtar Kelimeler:** Kredi kartı kullanımı, Lojistik regresyon analizi, Sınıflandırma.

## 1. GİRİŞ

Kredi kartı, belirlenmiş limitler içerisinde sisteme dahil olan işyerlerinde gereksinim duyulan mal ve hizmeti nakit ödeme yapmaksızın satın alınmasına ve bedelinin belirli bir tarihte ödenmesine olanak sağlayan popüler bir ödeme aracıdır. Teknoloji alanındaki hızlı gelişmeler dünya ülkelerine paralel olarak ülkemizde de kredi kartı kullanımını arttırmıştır (Cengiz, 2009).

Bu çalışmada, bireylerin kredi kartlarını iptal ettirme niyetleri lojistik regresyon analizi ile incelenecek daha sonra araştırma sonucunda elde edilen bulgular yorumlanacaktır.

## 2. YÖNTEM

Lojistik regresyon, çok değişkenli istatistiksel tekniklerden biri olup, bu tekniklerin sınıflama özelliğine hizmet eden bir yöntemdir. Bu yöntem sonucunda elde edilecek denklem sayesinde yeni atanan bireylerin hangi grupta yer alacağına karar verilebilir (Özdamar, 2004). Lojistik regresyon analizi belirli varsayımların sağlanmadığı durumlarda diskriminant analizi ve çapraz tablolara alternatif olarak kullanılan bir yöntemdir. Kümeleme analizinde bireylerin atanacağı gruplar bilinmezken lojistik regresyon ve diskriminant analizinde ise bireylerin atanacağı gruplar bilinmektedir. Lojistik regresyonu, doğrusal regresyondan ayıran özellikler aşağıda verildiği gibidir.

- i. Doğrusal regresyon analizinde tahmin edilecek bağımlı değişken sürekli iken lojistik regresyon analizinde kesiklidir.
- ii. Doğrusal regresyon analizinde bağımlı değişkenin değeri, lojistik regresyon analizinde ise bağımlı değişkenin düzeylerinden birinin gerçekleşmesi olasılığı tahmin edilir.

\*Arş. Gör., Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [gultekinatalik@anadolu.edu.tr](mailto:gultekinatalik@anadolu.edu.tr)

\*\*Arş. Gör., Eskişehir Osmangazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [ybulut@ogu.edu.tr](mailto:ybulut@ogu.edu.tr)

\*\*\*Doç. Dr., Eskişehir Osmangazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [zyildiz@ogu.edu.tr](mailto:zyildiz@ogu.edu.tr)

## BİREYLERİN KREDİ KARTLARINI İPTAL ETTİRME NİYETLERİNİN İNCELENMESİNDE LOJİSTİK REGRESYON ANALİZİ

Gültekin ATALIK\*

Y. Murat BULUT\*\*

Zeki YILDIZ\*\*\*

### ÖZET

*Lojistik regresyon; bağımlı değişkenin kategorik ikili, üçlü, çoklu kategorilerde gözlemlendiğinde bağımsız değişken ya da değişkenlerle sonuç ilişkisini belirlemede kullanılan bir yöntemdir. Bu çalışmada, bireylerin kredi kartlarını iptal ettirmesi bakımından sınıflandırılmasında lojistik regresyon analizinden yararlanılmıştır.*

**Anahtar Kelimeler:** Kredi kartı kullanımı, Lojistik regresyon analizi, Sınıflandırma.

### 1. GİRİŞ

Kredi kartı, belirlenmiş limitler içerisinde sisteme dahil olan işyerlerinde gereksinim duyulan mal ve hizmeti nakit ödeme yapmaksızın satın alınmasına ve bedelinin belirli bir tarihte ödenmesine olanak sağlayan popüler bir ödeme aracıdır. Teknoloji alanındaki hızlı gelişmeler dünya ülkelerine paralel olarak ülkemizde de kredi kartı kullanımını arttırmıştır (Cengiz, 2009).

Bu çalışmada, bireylerin kredi kartlarını iptal ettirme niyetleri lojistik regresyon analizi ile incelenecek daha sonra araştırma sonucunda elde edilen bulgular yorumlanacaktır.

### 2. YÖNTEM

Lojistik regresyon, çok değişkenli istatistiksel tekniklerden biri olup, bu tekniklerin sınıflama özelliğine hizmet eden bir yöntemdir. Bu yöntem sonucunda elde edilecek denklem sayesinde yeni atanan bireylerin hangi grupta yer alacağına karar verilebilir (Özdamar, 2004). Lojistik regresyon analizi belirli varsayımların sağlanmadığı durumlarda diskriminant analizi ve çapraz tablolara alternatif olarak kullanılan bir yöntemdir. Kümeleme analizinde bireylerin atanacağı gruplar bilinmezken lojistik regresyon ve diskriminant analizinde ise bireylerin atanacağı gruplar bilinmektedir. Lojistik regresyonu, doğrusal regresyondan ayıran özellikler aşağıda verildiği gibidir.

- i. Doğrusal regresyon analizinde tahmin edilecek bağımlı değişken sürekli iken lojistik regresyon analizinde kesiklidir.
- ii. Doğrusal regresyon analizinde bağımlı değişkenin değeri, lojistik regresyon analizinde ise bağımlı değişkenin düzeylerinden birinin gerçekleşmesi olasılığı tahmin edilir.

\*Arş. Gör., Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [gultekinatalik@anadolu.edu.tr](mailto:gultekinatalik@anadolu.edu.tr)

\*\*Arş. Gör., Eskişehir Osmangazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [ybulut@ogu.edu.tr](mailto:ybulut@ogu.edu.tr)

\*\*\*Doç. Dr., Eskişehir Osmangazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [zyildiz@ogu.edu.tr](mailto:zyildiz@ogu.edu.tr)

- iii. Lojistik regresyon analizinde, doğrusal regresyon analizinde söz konusu olan bağımsız değişkenin çok değişkenli normal dağılım göstermesi şartı aranmaz (Cankuş, 2008).
- iv. Lojistik regresyon analizinde, regresyon analizindeki ortak kovaryansa sahip olma şartı aranmaz.

Bağımlı değişkenin alabileceği değerlerin 0-1 arasında olduğu durumlarda kullanılan

$$\pi = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)} \quad (1)$$

formülündeki bu fonksiyona “Lojistik Fonksiyon” adı verilir.

Burada;

$\beta_0$ :Sabiti

$\beta_1, \beta_2, \dots, \beta_k$ :Regresyon parametrelerini

$X_1, X_2, \dots, X_k$ :Bağımsız değişkenleri

$\pi$ : İncelenen olayın gözlenme olasılığını göstermektedir.

### 3. ÇALIŞMADA KULLANILAN DEĞİŞKENLER

Çalışmada kullanılan bağımsız değişkenler 2008 yılında yapılmış olan bir kredi kartı anketinde yer alan değişkenlerden oluşmaktadır. Ele alınan değişkenler kredi kartı iptal ettirme niyetini etkileyebilecek değişkenlerden oluşmaktadır. Çalışmada kullanılan bağımsız değişkenlerin aldığı değerler 0 ve 1’dir. Örneğin;  $X_1$ :Ö.Kolaylığı değişkeni için “0 Kredi kartı alışverişlerde ödeme kolaylığı sağlamaz” ve “1 Kredi kartı alışverişlerde ödeme kolaylığı sağlar” olarak kodlanmıştır. Değişkenler aşağıdaki gibidir.

$X_1$ :Ö.Kolaylığı(Kredi kartı alışverişlerde ödeme kolaylığı sağlar)

$X_2$ :G.Faiz( Kredi kartları gereksiz faiz ödemektir)

$X_3$ :E.Problem(Kredi kartları ekonomik problemleri çözmeye yardımcıdır)

$X_4$ :A.Davranış(Kredi kartı kullanmak akıllıca davranıştır)

$X_5$ :K.Kullanmak(Kontrollü bir şekilde kredi kartı kullanmak bir zarar getirmez)

$X_6$ :Fayda(Kredi kartına ödenen yıllık ücret düşünülduğünde kredi kartlarından sağlanan fayda yeterlidir)

$X_7$ :Avantaj(Bilinçli tüketiciler için kredi kartı kullanmak avantajlıdır)

$X_8$ :F.Harcama(Kredi kartı ihtiyaçtan daha fazla harcama yaptırır)

$X_9$ :Sakinca(Kredi kartlarını internette alışveriş yaparken kullanmanın hiçbir sakıncası yoktur)

$X_{10}$ :Tehlike(İnternet, telefon üzerinden yapılan alışverişlerde kredi kartını kullanmak tehlikelidir)

- X<sub>11</sub>**:Mağduriyet(Kredi kartının çalınması veya kaybolması durumunda tüketiciler mağdur edilmez)
- X<sub>13</sub>**:Dürüstlük(Bankalar kredi kartı işlemlerinde tüketicilere dürüst davranır)
- X<sub>14</sub>**:Hata(Bankalar kredi kartı hizmetlerini(faiz hesabı, geri ödeme günü, hesap özeti vb.)hatasız yapar.)
- X<sub>15</sub>**:Tasarruf(Kredi kartları tasarruf(birikim) yapmayı engeller.)
- X<sub>16</sub>**:Mutluluk(Kredi kartına sahip olmak insanı mutlu eder)
- X<sub>17</sub>**:Memnuniyet(Kullandığım kredi kartından tamamıyla memnunum.)
- X<sub>18</sub>**:Tavsiye(Kullandığım kredi kartımı çevremdeki insanlara tavsiye ederim.)
- X<sub>19</sub>**:G.Kullanım(Gelecekte kredi kartı kullanmaya devam edeceğim.)
- X<sub>20</sub>**:Gurur(Kredi kartımın ismini çevremdekilere söylemekten gurur duyarım.)
- X<sub>21</sub>**:K.Devam(Bankanın sağladığı koşullar ne şekilde değişirse değişsin bundan sonrada aynı kredi kartını kullanmaya devam edeceğim.)
- X<sub>22</sub>**:Yenileme(Kredi kartımın kullanım süresi dolduğunda yenilemeyi düşünüyorum.)
- X<sub>23</sub>**:K.Karar(Yakın gelecekte kredi kartı kullanmama kararı alabilirim.)
- X<sub>24</sub>**:Kullanım S.(Kredi kartımın kullanım sıklığını azaltacağım.)
- X<sub>25</sub>**:Dolandırıcılık(Kredi kartı dolandırıcılıkları çok yaygın olduğundan kredi kartımı kullanamıyorum.)
- X<sub>26</sub>**:Seçim(Tekrar seçim yapmak zorunda kalsam yine bu kredi kartını seçerim.)
- X<sub>27</sub>**:K.Ücreti(Kredi kartının yıllık sağladığı fayda düşünüldüğünde yıllık ödenen kart ücreti önemsizdir.)

Açıklanan değişken (İptal) ise, “Kredi kartımı iptal ettirmeyi düşünmüyorum” (0) ve “Kredi kartımı iptal ettirmeyi düşünüyorum” (1) şeklinde iki kategoriden oluşmaktadır.

Çalışmada kullanılan veriler 2008 yılında anket yoluyla İstanbul, Ankara, İzmir, Bursa, Eskişehir ve Denizli illerinden bu illerde oturan Osmangazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü öğrencileri aracılığıyla elde edilmiştir. Anket 384 kişiye uygulanmış, fakat 64 anket hatalı olduğundan dolayı 320 anket değerlendirmeye alınmıştır.

#### 4. VERİLERİN ÇÖZÜMLENMESİ

Çalışmada kredi kartı kullanıcılarının kredi kartlarını iptal ettirmeyi düşünen ve iptal ettirmeyi düşünmeyen biçiminde sınıflandırılması ve bu sınıflandırma için kullanılacak değişkenlerin belirlenmesine çalışılmıştır. Bu çalışma iki grup için söz konusu olan bir ayırma problemidir. Bağımsız değişkenin iki sıklı ve bağımsız değişkenlerinde kesikli olması nedeniyle bu amacı gerçekleştirebilmek için kullanılacak teknik Lojistik Regresyon Analizi'dir. Sınıflama işleminin yapılabilmesi için “SPSS 15.0 for Windows Release 15.0.0” paket programı kullanılmıştır.

İleriye Doğru Değişken Seçme Tekniği sonucu oluşturulan lojistik regresyon katsayıları ve bunlara ilişkin test sonuçları Tablo 1’de verilmiştir.

**Tablo 1. İleriye doğru değişken seçme tekniğiyle çözümlene sonuçları**

Değişken	B	S.E	Wald	p	Exp( $\hat{\beta}$ )
G.Faiz(Katılıyorum)( $X_2$ )	0,830	0,333	6,220	0,013	2,293
Tehlike(Katılıyorum) ( $X_{10}$ )	-0,849	0,435	3,811	0,051	0,428
G.Kullanım(Katılıyorum) ( $X_{19}$ )	-4,202	1,083	15,046	0,000	0,015
K.Karar(Katılıyorum) ( $X_{23}$ )	2,900	0,346	70,104	0,000	18,175
Kullanım S.(Katılıyorum) ( $X_{24}$ )	1,185	0,426	7,740	0,005	3,270
Dolandırıcılık(Katılıyorum) ( $X_{25}$ )	1,148	0,365	9,895	0,002	3,152
Sabit	1,266	1,153	1,207	0,272	3,547

İleriye doğru değişken seçme (Forward stepwise selection) tekniğiyle elde edilen lojistik regresyon modeli,

$$g(x) = 1,266 + 0,830 \times (\text{G. Faiz(Katılıyorum)}) - 0,849 \times (\text{Tehlike(Katılıyorum)}) \\ + 2,900 \times (\text{K. Karar(Katılıyorum)}) + 1,185 \times (\text{Kullanım. S(Katılıyorum)}) + 1,148 \\ \times (\text{Dolandırıcılık(Katılıyorum)})$$

şeklindedir.

Model sonucu oluşturulan sınıflama tablosu Tablo 2'de verilmiştir.

**Tablo 2. Modelin sınıflandırma sonuçları**

Gözlemlenen	Kestirim		Doğru Oranları(%)	Atama
	0	1		
0	138	32	81,2	
1	21	147	87,5	
Toplam Doğru Sınıflama Oranı (%)			84,3	

Tablodan da görülebileceği gibi oluşturulan modelin sınıflama yüzdesi oldukça yüksektir. Bu yüzde aynı zamanda modelin anlamlı olduğunun da bir göstergesi olmasına rağmen ayrıca ki-kare testiyle de anlamlılık sınaması yapılacaktır.

Çoklu doğrusal regresyonda katsayıların anlamlılığına ilişkin genel anlamlılık sınaması F testine karşılık gelebilecek benzer bir test lojistik regresyon analizi için geliştirilmiştir.  $L_0$  sadece sabit terimden oluşan modelin olabilirlik değeri,  $L_1$  elde edilen modelin olabilirlik değeri olmak üzere

$$C = -2 \times \log(L_0/L_1) = -2 \times (\log L_0 - \log L_1) \quad (2)$$

olarak tanımlanan ölçüt (k-1) serbestlik dereceli Ki-kare dağılımı göstermektedir (Tatlıdil, 1992).

Denklemin anlamlılığı için  $C=219,822$  olarak bulunmuştur.  $\alpha = 0,05$  ve 5 serbestlik dereceli Ki-kare tablo değeri 11,07'dan daha büyük olduğu için model anlamlı bulunmuştur.

Lojistik regresyonda katsayıların yorumu Odds (olasılık) oranı ile yapılabilmektedir. Odds oranı bir olayın meydana gelmesi olasılığının, gelmemesi olasılığına oranı olarak tanımlanabilir. Tablo 1'deki  $\text{Exp}(\hat{\beta})$  değerleri odds oranlarıdır. Tablo 1'deki odds oranlarının bir kısmı şu şekilde yorumlanabilir. K.Karar( $x_{23}$ ) faktörüne katılan bireylerin, mevcut kredi kartını iptal ettirme olasılığı bu faktöre katılmayan bireylere göre 18,175 kat daha fazladır. Benzer şekilde Kullanım S. ( $x_{24}$ ) faktörüne katılan bireylerin, mevcut kredi kartını iptal ettirme olasılığı bu faktöre katılmayan bireylere göre 3,270 kat daha fazladır.

Regresyon analizindeki  $R^2$  istatistiği ile lojistik regresyon analizindeki  $R^2$  değerini karşılaştırmak uygun değildir. Bununla birlikte literatürde lojistik regresyon analizi için birkaç  $R^2$  istatistiğine yer verilmektedir. Bunlardan ikisi Cox ve Snell  $R^2$  ve Nagelkerke  $R^2$  istatistikleridir. Cox ve Snell  $R^2$  olabilirlik esasına göre çoklu  $R^2$  istatistiğine benzemektedir. İstatistiğin maksimum değeri 1'den küçük olması bu istatistiğin yorumunu güçleştirmektedir. Cox ve Snell  $R^2$  istatistiğinin 0-1 arasında değer almasını sağlamak amacıyla Nagelkerke  $R^2$  istatistiği geliştirilmiştir (Cengiz, 2009). Çalışmamızdaki Cox ve Snell  $R^2$  değeri 0,478 bulunmuştur. Nagelkerke  $R^2$  değeri ise 0,638 olarak bulunmuştur. Bu değer anlamı bağımlı değişken ile bağımsız değişkenler arasında %63,8'lik bir ilişki olduğudur. Yani ele alınan bağımsız değişkenler bağımlı değişkendeki değişimin %63,8'ini açıklayabilmektedir.

## 5. SONUÇ

Çeşitli illerde yapılan ankete katılan bireylerin anketteki bazı soruları dikkate alarak, bu bireylerin kredi kartlarını iptal ettirip ettirmemesi üzerine sınıflandırılmaları söz konusudur. Çalışmada, incelenen değişkenlerle binary lojistik regresyon modeli oluşturulmuş ve bu modelin doğru sınıflandırma oranı % 84,3 olarak bulunmuştur. Modelin istatistiksel olarak ayırıcı özelliğe sahip olduğu C istatistiği yardımıyla da ortaya konulmuştur.

Günümüzde kredi kartı kullanımı bankalar arasında ciddi bir rekabete sebep olduğundan, bankalar yakın gelecekte kredi kartı kullanmama kararı alabilecek olan bireylere yönelik çeşitli kampanyalar (promosyonlar) düzenleyerek bu bireylerin kredi kartlarını iptal ettirmelerinin önüne geçebilirler. Benzer şekilde, bankalar kredi kartı dolandırıcılığının çok olduğunu düşünen müşterilerin güvenini kazanmaya yönelik adımlar atarlarsa bu müşterilerin kartlarının iptal edilmesinin önüne geçilebilir.

Sonuç olarak, ele alınan değişkenlerin kredi kartı kullanıcılarını "Kredi kartını iptal ettirebilir" ve "Kredi kartını iptal ettirmez" biçiminde sınıflandırılmasında kullanılabilmesi belirlenmiştir. Başka bir ifadeyle, ele alınan değişkenlerin sınıflandırmada uygun değişkenler olduğu belirlenmiştir.



## 6. KAYNAKLAR

Cengiz, E., 2009. Bireylerin Kredi Kartlarını Deęiřtirme Tutumları, Fırat Üniversitesi Sosyal Bilimler Dergisi, 19(2), 179-196.

Özdamar, K., 2004. Paket Programlar ile İstatistiksel Veri Analizi, Kaan Kitabevi, Eskişehir.

Cankuş, B., 2008. Belediye Hizmetlerine İliřkin Algı ve Memnuniyetin İkili Lojistik Regresyon Analizi ile Ölçümü (Eskişehir İli Örneęi), Eskişehir Osmangazi Üniversitesi, Eskişehir (yayımlanmamıř).

Tatlıdil, H., 1992. Çok Deęiřkenli İstatistiksel Analiz, Engin Yayınları, Ankara.

## LOGISTIC REGRESSION ANALYSIS IN EXAMINING THE USERS' INTENTION FOR THE CANCELLATION OF CREDIT CARDS

### ABSTRACT

*Logistic regression is a method used to determine the relationship between a dependent variable, when it is observed in binary, triple or multiple categories, with independent variable or variables. In this study, logistic regression analysis is utilized to classify the users in terms of cancelling their credit cards.*

**Keywords:** Credit card using, Logistic regression analysis, Classification.

## LOJİSTİK REGRESYON VE DOĞRUSAL DİSKRİMİNANT ANALİZLERİNDE KULLANILAN BAZI İNDEKSLERİN KARŞILAŞTIRILMASI

Atilla GÖKTAŞ\*

Barış KESKİN\*\*

Selen ÇAKMAKYAPAN\*\*\*

### ÖZET

*Lojistik regresyon ve doğrusal diskriminant analizi, bireylerin ya da gözlemlerin sınıflandırılmasında yaygın olarak kullanılan iki yöntemdir. Bu analizlerin aynı amaçla kullanılabilmesi, hangisinin daha iyi sonuçlar elde ettiği sorusunu akla getirmektedir. Bu konu üzerine çalışmalar yapılmış ve bu iki analiz karşılaştırılmıştır. Diskriminant analizi için gerekli varsayımların, lojistik regresyon için gerekli olmaması bu iki analizin farklı koşullarda tercih edilebilirliğini de değiştirmiştir. Bu çalışmada ise, değişkenlerin normal dağılım varsayımını sağlamadığı durumda bu iki analizin ve analizleri değerlendirmede kullanılan indekslerin karşılaştırılması amacıyla bir benzetim çalışması gerçekleştirilmiştir. Verilerin çok değişkenli normal dağılım varsayımını sağlamayıp, farklı dağılımlar gösterdiği durumda lojistik regresyon analizinin diskriminant analizine göre genel olarak daha iyi sonuç verdiği görülmüştür. Örneklem büyüklüğü arttıkça iki analizden elde edilen sonuçlar arasındaki farklılık azalmıştır. Örneklem büyüklüğü ne olursa olsun, tüm indeks ölçütlerine göre lojistik regresyon analizinin doğrusal diskriminant analizine göre sınıflandırmada daha başarılı olduğu görülmüştür. Cohen'in Kappa katsayısı yeni bir indeks olarak kullanılmıştır. Ayrıca, hangi modelin iyi olduğu bilindiğinde elde edilen indeks değerleri sayesinde, indekslerin iki yöntemin tahmindeki doğruluklarını karşılaştırmadaki başarısı değerlendirmeye alınmıştır.*

**Anahtar Kelimeler:** Diskriminant analizi, Cohen'in Kappası, Lojistik regresyon.

### 1. GİRİŞ

Lojistik regresyon (LR) ve doğrusal diskriminant analizi (DA), bireylerin ya da gözlemlerin sınıflandırılmasında yaygın olarak kullanılan iki yöntemdir. Aynı amaçla kullanılmalarına rağmen, aralarındaki en önemli farklılık varsayımlarından ve parametre tahminlerinde kullandıkları yöntemlerden kaynaklanmaktadır. DA'da açıklayıcı değişkenlerin normal dağılımdan gelmiş olması ve grupların aynı kovaryans matrisine sahip olması gerekirken, LR'de bu varsayımlara gerek yoktur.

Literatürde bu iki analizin karşılaştırılmasına yönelik çalışmalar yapılmış ve bu çalışmalarda varsayımlar sağlandığında DA'nın daha iyi sonuçlar verebileceği, ancak diğer tüm durumlarda LR'nin daha başarılı olacağı ifade edilmiştir (Press vd., 1978).

Bu analizlerin başarısını değerlendirmede ya da iki analizin karşılaştırılması için önerilen ölçütler vardır. Bu ölçütlerden bazıları B, Q ve C indeksleridir. Bu çalışmada, LR ve DA ile ilgili elde edilen bilgiler ışığında, bu indekslerin ve ayrıca Cohen'in

\*Yrd. Doç. Dr., Muğla Sıtkı Koçman Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Muğla, e-posta: [gatilla@mu.edu.tr](mailto:gatilla@mu.edu.tr)

\*\*Arş. Gör., Muğla Sıtkı Koçman Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Muğla, e-posta: [bariskeskin@mu.edu.tr](mailto:bariskeskin@mu.edu.tr)

\*\*\*Arş. Gör., Muğla Sıtkı Koçman Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Muğla, e-posta: [selencak@mu.edu.tr](mailto:selencak@mu.edu.tr)

Kappa katsayısının bu bilgiye ne kadar paralel sonuçlar verdiği bakılarak aralarında bir değerlendirme yapılmıştır. Çalışmada ilk olarak, LR ve DA'dan genel hatlarıyla kısaca bahsedilmiş, iki analize ilişkin tahmin yöntemleri karşılaştırılmış ve kullanılan indekslere ilişkin bilgiler verilmiştir. Sonrasında, yapılan benzetim çalışması anlatılmış ve elde edilen sonuçlar değerlendirilmiştir.

## 2. LOJİSTİK REGRESYON VE DOĞRUSAL DİSKRİMİNANT ANALİZİ

### 2.1 Lojistik Regresyon

Lojistik regresyon analizi; çok değişkenli verilerin sınıflandırılmasında, gerek bu verilere uygulanabilecek çeşitli istatistiksel yöntemler için gerekli bir ön analiz olarak, gerekse başlı başına bir analiz olarak sıkça kullanılan bir analizdir. Bu analizde bağımlı değişken kategorik, bağımsız değişkenler ise kategorik değişken, sürekli değişken veya kategorik ve sürekli değişkenlerin bir karması olabilir. Lojistik regresyon analizi normallik, ortak kovaryansa sahip olma, süreklilik varsayımı gibi ön koşullara gerek duymadığından, bu varsayımları gerektiren yöntemlere alternatif olarak geliştirilmiştir.

Lojistik regresyon analizinde kullanılan model, bağımlı değişkenin 0, 1 gibi iki ya da ikiden çok düzey içeren kesikli değişken olması durumunda normallik varsayımı gerektirmemesi nedeniyle kullanım rahatlığı sağlamaktadır (Gürcan, 1998). Yorum kolaylığı ve kestirim güçlülüğü nedeniyle yaygın olarak kullanılmaktadır.

Bu analizde çok değişkenli istatistiksel verilerin sınıflandırılması, bağımlı değişkenin hesaplanan olasılık tahminleri yardımıyla yapılır.  $y$  bağımlı değişkenin sadece 0 ve 1 gibi iki değer aldığı durumda,  $P(y_i = 1)$   $i$ . gözlemin 1 değeri olması olasılığı,  $P(y_i = 0)$   $i$ . gözlemin 0 değeri olması olasılığı olmak üzere  $i$ . gözlemin beklenen değeri eşitlik 1 ile verilir.

$$E(y_i) = 1 \times P(y_i = 1) + 0 \times P(y_i = 0) = P(y_i = 1) \quad (1)$$

Analizde  $p$  adet bağımsız değişkenli  $n$  adet gözlem için kullanılan lojistik regresyon modeli ise  $\beta_k$   $k$ . bağımsız değişkene ait katsayıyı göstermek üzere, gözlemin grup 1'de olma olasılığı ya da başka bir ifadeyle bağımlı değişkenin 1 değerini alması olasılığı eşitlik 2 ile verilmiştir.

$$P(y_i = 1) = \frac{1}{1 + e^{-(\sum_{k=0}^p \beta_k x_{ik})}} \quad i = 1, 2, \dots, n \quad (2)$$

Modelde yer alan parametre tahminleri, en çok olabilirlik, yeniden ağırlıklandırılmış en küçük kareler ve minimum lojit ki-kare yöntemleri ile hesaplanabilir (Gürcan, 1998). Bu çalışma kapsamında, en çok olabilirlik tahmin edicileri dikkate alınmıştır.

Bu analizin en önemli avantajı, bağımlı değişkeni etkileyebilecek önemli değişkenlerin belirlenebilmesi ve belirli karakteristiklere göre bir gözlemin bağımlı değişken kategorilerine düşmesi olasılıklarının hesaplanabilmesidir.

## 2.2 Doğrusal Diskriminant Analizi

DA, kelime ve genel anlamı itibarıyla noktaları, bireyleri ya da gözlemleri ayırma ile ilgili bir analizdir. Bu analiz bireylere ait  $p$  tane özellikten yararlanarak ait oldukları grupları belirlemede veya mevcut grupları birbirinden ayıracak en iyi fonksiyonu bulmada kullanılan çok değişkenli istatistik tekniklerinden biridir. Hatalı sınıflandırma olasılığını en aza indirgeyerek, gözlemleri ait oldukları gruplara ayırma amacına yönelik olarak kullanılan istatistiksel bir karar verme yöntemi olarak da tanımlanmaktadır. Birden çok bağımsız değişkene göre gözlemleri gruplara atamada yaygın olarak kullanılan ve kabul görmüş olan bu yöntem, değişkenlerin çok değişkenli normal dağılım göstermesi, aralarında önemli ilişkilerin olmaması ve grup varyanslarının homojenliği varsayımlarına dayanır. İlk olarak Fisher tarafından gözlemleri iki farklı gruba atamak için önerilen yöntem, sonrasında daha fazla grup için de geliştirilmiş, geliştirilmiştir. Yöntemin geliştirilmesi ile DA'nın kullanımı giderek yaygınlaşmıştır.

Bu analizde gözlemlerin sınıflandırılması işlemi, elde edilen diskriminant fonksiyonları ve kesim noktaları yardımıyla yapılır. Yöntemin başarısı, veri kümesindeki kaç tane gözlemi doğru biçimde sınıflandırdığı (kendi grubuna atadığı) ve bazı indeks ölçütleri ile ölçülmektedir.

DA'nın en basit hali iki grup söz konusu olduğunda elde edilir. Gruplara ait gözlemlerin birbirinden ayrılmasında ya da yeni bir gözlemin bir gruba atanmasında, bu grupların merkezleri arasından geçen doğrusal bir ayırma fonksiyonu kullanılır. DA modeli, LR'nin aksine, iki ayrı grubun sırasıyla  $\mu_1$  ve  $\mu_2$  ortalamaları ve aynı  $\Sigma$  kovaryans matrisi ile normal dağıldıklarını varsayar. Bu varsayım altında, bir bireyin grup 1'e atanması olasılığı ise aşağıdaki gibi ifade edilebilir.

$$P(y_i = 1) = \frac{1}{1 + e^{-\left(\sum_{k=0}^p \beta_k x_{ik}\right)}} \quad i = 1, 2, \dots, n \quad (3)$$

Burada dikkat çeken nokta, LR ve DA modellerinin aynı fonksiyonel formda ifade edilebildikleridir. Aralarındaki fark ise katsayıların tahmininden kaynaklanmaktadır.

DA'da,  $\pi_1$  ve  $\pi_2$  bir gözlemin sırasıyla grup bir ve grup ikide olma önsel olasılıklarını temsil etmek üzere,  $\beta_0$  ve  $\beta' = (\beta_1, \beta_2, \dots, \beta_p)$  aşağıdaki gibi hesaplanır.

$$\beta_0 = -\log \frac{\pi_2}{\pi_1} + \frac{1}{2}(\mu_1 + \mu_0)' \Sigma^{-1} (\mu_1 - \mu_0)$$

$$\beta = (\mu_1 - \mu_0)' \Sigma^{-1}$$

Eğer  $\pi_1, \pi_2, \mu_1, \mu_2, \Sigma$  parametreleri bilinmiyorsa örneklemden elde edilen tahmin değerleri kullanılır.  $n_1$  grup 1'deki gözlem sayısı,  $n_2$  grup 2'deki gözlem sayısı ve  $n = n_1 + n_2$  toplam gözlem sayısı olmak üzere tahmin değerleri aşağıdaki eşitlikler ile elde edilir.

$$\begin{aligned}\hat{\pi}_1 &= n_1/n, & \hat{\pi}_2 &= n_2/n \\ \bar{x}_{11} &= \frac{1}{n_1} \sum_{y_i=0} x_{i1}, & \bar{x}_{12} &= \frac{1}{n_1} \sum_{y_i=0} x_{i2}, \dots, & \bar{x}_{1p} &= \frac{1}{n_1} \sum_{y_i=0} x_{ip} \\ \bar{x}_{21} &= \frac{1}{n_1} \sum_{y_i=0} x_{i1}, & \bar{x}_{22} &= \frac{1}{n_1} \sum_{y_i=0} x_{i2}, \dots, & \bar{x}_{2p} &= \frac{1}{n_1} \sum_{y_i=0} x_{ip} \\ \hat{\mu}'_1 &= \bar{x}_1 = (\bar{x}_{11}, \bar{x}_{12}, \dots, \bar{x}_{1p}), & \hat{\mu}'_2 &= \bar{x}_2 = (\bar{x}_{21}, \bar{x}_{22}, \dots, \bar{x}_{2p}) \\ \hat{\sigma}^2 &= \frac{(n_1 - 1)\Sigma_1 + (n_2 - 1)\Sigma_2}{n_1 + n_2 - 2}\end{aligned}$$

### 2.3 Lojistik Regresyon ve Diskriminant Analizde Kullanılan Tahmin Yöntemlerinin Karşılaştırılması

- Aynı kovaryans matrisiyle normal dağılıma sahip olma varsayımı sağlanmadığı zaman DA yöntemi ile elde edilen eğim katsayısı tahminleri tutarlı olmayabilir. Bu nedenle, örneklem büyük olsa dahi bu yöntemle elde edilen tahminlerin iyi tahminler olacağı ya da iyi bir uyum elde edileceği konusunda bir garanti yoktur. Pratikte bağımsız değişkenlerin sürekli olmadıkları, iki veya daha fazla kategoriye sahip kesikli değişken oldukları durumlarla sıklıkla karşılaşmaktadır. Böyle durumlarda, DA ile, veri büyüklüğü sonsuz olsa dahi, bağımlı değişken için elde edilen olasılığın doğru olması beklenemez. Böyle durumlarda, LR ve parametre tahminleri için tutarlı yöntemlerden biri olan en çok olabilirlik yöntemi kullanılabilir.

- Diskriminant tahminleri, normallik koşulu ihlal edildiğinde, model katsayılarının önemi ile ilgili hatalı sonuçlar verebilir. Bu durumda, gerçekte sıfır olan eğim katsayısının büyük örneklerde en çok olabilirlik yöntemi ile sıfır olarak tahmin edilme eğilimi olmasına rağmen, diskriminant fonksiyonlarıyla tahmin edilen parametrelerle anlamsız değişkenlerde modelde yer alma eğiliminde olacaktır.

- Halperin, Blackwelder ve Verter (1971) en çok olabilirlik yöntemiyle elde edilen LR ve diskriminant fonksiyonlarıyla elde edilen DA tahminlerinin sayısal karşılaştırmasını yapmışlardır. Normal dağılım varsayımının sağlanmadığı koşullarda genellikle en çok olabilirlik yönteminin modele biraz daha iyi uyum sağladığını bulmuşlardır. Ayrıca, eşitlik 2 ile ifade edilen model geçerli olsa bile diskriminant tahminlerinin çok zayıf bir uyum vermesi ihtimali için teorik bir taban olduğunu bulmuşlardır (Press vd., 1978).

- Lojistik regresyon modeli yeterli istatistiklerin elde edilebildiği bir modeldir. En çok olabilirlik tahminleri yeterli istatistik fonksiyonlarıdır ve her zaman hata kareler ortalamasını daha küçük yapar. Ancak diskriminant fonksiyon tahmini yeterli istatistikler elde edemez.

- Lojistik regresyon modelin en çok olabilirlik kestirimleri, olayların gözlenen ve beklenen sayılarının eşit olmasını gerektirir ( $\sum y_i = \sum P(x_{j1}, \dots, x_{ki})$ ). Bu özellik, herhangi bir düzleştirme sürecinde de öncelikli olarak olması istenen bir özelliktir. Diskriminant fonksiyonu yaklaşımında ise tahmin toplamları gerçek gözlem sayılarından büyük olabilmektedir.

- Bazı uygulamalarda, diskriminant fonksiyonu tahmin edicilerinin önemli derecede yanlı olma eğilimi gösterdiğine dair kanıtlar vardır. McFadden (1996), Bayesci analiz çalışmasında, açıklayıcı değişkenlerin tipik önsel dağılımı için, diskriminant analizine dayanan seçim olasılıklarının tahminlerinin önemli derecede yanlı olacağını söylemiştir (Press vd., 1978).

### 3. LOJİSTİK REGRESYON (LR) VE DOĞRUSAL DİSKRİMİNANT ANALİZİNİN (DA) KARŞILAŞTIRILMASINDA KULLANILAN İNDEKSLER

LR ve DA yöntemlerinin karşılaştırılmasında kullanılan en basit ölçüt Doğrusal Sınıflandırma Oranı (DSO)'dır. Buna rağmen, yeterince hassas ve istatistiksel olarak etkin bir ölçüt değildir. Değişkenlerin kategorik olması durumunda doğrusal sınıflandırma oranı ile minimum düzeyde bilgi elde edilebilir. DSO'nun değeri 0 ile 1 arasında değişmekle birlikte genel olarak yorumlama aralığı 0.5 ile 1 arasındadır ve 1'e yakın olması istenir. 0.5'in altında elde edilen bir DSO değeri genellikle "şansa bağlı uyum"a atfedilir. DSO'nun zayıf yönü ise iki gruba ait toplam doğru tahmin sayılarının hangisinde yoğunlaşma olduğu hakkında bilgi vermemesidir.

Harell ve Lee (1985), yöntemlerin tahminlerdeki doğruluğunu kıyaslamak amacıyla, doğrusal sınıflandırma oranından daha güvenilir ölçüt olarak A, B, C ve Q indekslerini önermiştir (Pohar vd., 2004). Önerilen bu indeksler, yöntemlerin tahmin etmede ya da bireyleri gruplara ayırmada ne kadar iyi olduğunu daha etkili şekilde karşılaştıran kriterlerdir. C indeksi eşitlik 4 ile verilmiştir.

$$C = \sum_{i=1}^n \sum_{\substack{j=1 \\ y_i=0, y_j=1}}^n [I(P_j > P_i) + \frac{1}{2}I(P_j = P_i)] / n_1 n_2 \quad (4)$$

Burada,  $P_i$  i. gözleme ait olasılık değeri,  $P_j$  j. gözleme ait olasılık değerini ve I işaret fonksiyonunu göstermektedir. Görüldüğü gibi bu indeks, gerçek grup üye değerlerinden bağımsızdır. Sadece gruplar arası ayırmanın bir ölçüsüdür. Tahminin doğruluğuna dair bilgi vermez. C indeksi 1 değerini aldığı anda "mükemmel ayırma"ya, 0.5 değerini aldığı anda "şansa bağlı ayırma"ya atfedilir.

C indeksinden farklı olarak, B ve Q indeksleri tahminlerin doğruluğunu değerlendirmede kullanılabilir. B indeksi, tahmini olasılık değeri ile gerçek değerler arasındaki farkların karelerinin ortalamasının bir ölçüsüdür.

$$B = 1 - \sum_{i=1}^n (P_i - y_i)^2 / n \quad (5)$$

Burada,  $P_i$  i. gözleme ait olasılık değeri,  $y_i$  ise gerçek değeridir (1 ya da 0). B indeksi 0 ile 1 aralığında değerler alır. B indeksinin 1 değeri alması "mükemmel tahmin"e atfedilir. Örnekleme büyüklükleri eşit iken rastgele tahmin durumunda, B indeksinin değeri 0.75 civarındadır.

B indeksi ile benzer başka bir indeks ise Q indeksidir. Tahminlerin doğruluğunun bir ölçüsü olarak aşağıdaki eşitlik 6 ile bulunur.

$$Q = \sum_{i=1}^n [1 + \log_2(P_i^{Y_i} (1 - P_i)^{1 - Y_i})] / n \quad (6)$$

Q indeksinin 1 değerini göstermesi “mükemmel tahmin”e, 0 değerini göstermesi “rastgele tahmin”e atfedilir. 0 değerinden küçük çıkması ise rastgele tahminden de kötü olduğunu gösterir.

İki düzeyli verilerde, gözlemciler ya da aynı gözlemci tarafından yapılan iki ölçüm arasındaki uyum genellikle Cohen’in kappa katsayısı ( $\kappa$ ) ile incelenir. DSO’ya göre  $\kappa$  katsayısının avantajı, uyumun şans ile ortaya çıkma durumunu düzeltmektedir. Tablo 1 yardımıyla hesaplanan Cohen’in kappa katsayısı, eşitlik 7 ile verilir (Alpar vd., 2010).

**Tablo 1. 2x2 uyum tablosu**

Gözlemci B	Gözlemci A		
	Var	Yok	Toplam
Var	a	b	(a+b)
Yok	c	d	(c+d)
	(a+c)	(b+d)	n

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \quad (7)$$

Burada,  $P_0$ , DSO’yu göstermektedir.  $P_e$  şansa bağlı uyumu gösterir ve eşitlik 8 ile verilmektedir:

$$P_e = \frac{[(a+c) \times (a+b)] + [(b+d) \times (c+d)]}{n^2} \quad (8)$$

Kappa katsayısı -1 ile +1 arasında değerler alır.  $\kappa = 1$  olması durumunda tam uyum söz konusudur.  $\kappa \geq 0$  ise gözlenen uyum şansa bağlı uyuma eşit ya da büyüktür. Gözlenen uyumun şansa bağlı uyumdan küçük olması durumunda  $\kappa < 0$  olur. Kappa katsayısının negatif değerleri güvenilirlik açısından anlamlı değildir. Bu yüzden sadece pozitif değerleri dikkate alınır.

Landis ve Koch uyumun derecesini elde edilen kappa katsayısı 0.20’ye eşit yada küçük ise “zayıf uyum”, 0.21-0.40 aralığında ise “ortanın altında uyum”, 0.41-0.60 aralığında ise “orta düzeyde uyum”, 0.61-0.80 aralığında ise “iyi düzeyde uyum” ve 0.81-1.00 aralığında ise “çok iyi düzeyde uyum” olarak tanımlamışlardır (Subhash, 1996).

İki gözlemci arasındaki uyumun şansa bağlı kısmını düzelten bir uyum ölçüsü olarak tanımlanan kappa katsayısının iki sorunu vardır. Birinci sorun; DSO yüksek olmasına rağmen, satır-sütun toplamlarındaki büyük dengesizlikler nedeniyle küçük bir  $\kappa$  değerinin elde edilmesidir. İkinci sorun ise, satır-sütun toplamlarında asimetrik

dengeşizlik olduđunda  $\kappa$  'nın simetrik dengeşizlik olması durumuna göre daha yüksek bulunabilmesidir.  $(a+c)$  ile  $(b+d)$  ve  $(a+b)$  ile  $(c+d)$  toplamlarının eşit ya da birbirine yakın olması durumunda satır-sütun toplamlarının dengeli olduđu söylenir. Bu toplamlar birbirinden uzaklaştıkça satır-sütun toplamları dengeşizleşmeye başlar.  $(a+c)$  ile  $(a+b)$  ve  $(b+d)$  ile  $(c+d)$  toplamlarının eşit ya da birbirine yakın olması durumunda ise satır-sütun toplamlarının simetrik olduđu söylenir. Yine bu toplamların birbirinden uzaklaşması satır-sütun toplamlarının asimetrik olmasına yol açmaktadır (Subhash, 1996).

Bu çalışmada, yukarıda tanımlanan Cohen'in Kappa'sı, gerçek grup üye değerleri ile tahmin edilen üye değerlerinin uyumunu ölçmek için kullanılmıştır. Kappa'nın kullanımında karşılaşılabilecek iki sorun da grup büyüklüklerinin eşit alınması sayesinde ortadan kaldırılmıştır. Burada, "gözlemciler" yerine "gerçek grup üye değerleri" ve "tahmin edilen grup üye değerleri" kullanılmıştır. "Evet-Hayır" yerine ise grupları temsilen "0-1" gösterimi kullanılmıştır.

#### 4. UYGULAMA

Verilerin normal dağılım göstermesi varsayımı, diskriminant analizinin temel varsayımlarından biridir. Bu varsayımın ihlal edilmesi durumunda, böyle bir varsayım gerektirmeyen lojistik regresyon analizinin daha iyi sonuçlar vermesi beklenir. Bu beklenti literatürde daha önce yapılan çalışmalarla da desteklenmiştir. Bu çalışmada ise, bir benzetim çalışmasıyla farklı örneklem büyüklüğündeki normal dağılmayan veriler üretilmiş ve sonrasında bu verilere uygulanan lojistik regresyon ve diskriminant analizlerinden elde edilen sonuçlara ilişkin indeks değerleri hesaplanmıştır. Normal dağılım varsayımının ihlali durumunda, lojistik regresyon analizinin diskriminant analizinden iyi sonuçlar vereceđi bilgisinden yola çıkarak, hesaplanan bu indekslerin başarısı karşılaştırılmıştır.

Bu çalışmada, dağılımları sırasıyla üstel, poisson ve tekdüze olan üç adet bağımsız değişken için veri üretimi yapılmıştır. Bu dağılımlar normal dağılım gösterme varsayımını bozmak amacıyla, keyfi olarak tercih edilmiştir. Açıklayıcı değişkenlerin dağılımlarına ilişkin parametre değerlerinin değiştirilmesi sayesinde iki farklı grup oluşturulmuştur. Gruplar arasındaki farklılık, Mahalanobis uzaklığından faydalanılarak, "yok", "az", "orta" ve "çok" olarak düzeylere ayrılmıştır. Bu düzeylere ilişkin parametre değerleri Tablo 2'de yer almaktadır. Her iki grubun örneklem büyüklükleri aynı olmak üzere, örneklem büyüklüğünün indekslere etkisini ölçmek amacıyla, altı farklı örneklem büyüklüğü için veriler üretilmiştir. Örneklem büyüklüğü seçimi için ön çalışma yapıp, yeterince büyük ve yeterince küçük olan örneklem büyüklüklerini de içerecek şekilde keyfi olarak belirlenmiştir. Bu büyüklükler 20, 40, 60, 80, 100 ve 120'dir. Böylece, düzeyler ve örneklem büyüklükleri dikkate alınarak 24 farklı kombinasyon oluşturulmuştur.



Tablo 2. Gruplara ilişkin parametre değerleri

BAĞIMSIZ DEĞİŞKENLERİN DAĞILIMI	GRUPLAR	PARAMETRELER	GRUPLAR ARASI FARK			
			YOK	AZ	ORTA	ÇOK
ÜSTEL DAĞILIM	Grup 1	Ortalama	1	1	1	1
	Grup 2		1.1	1.5	2	3
POISSON DAĞILIMI	Grup 1	Ortalama	1	2	3	2
	Grup 2		1.1	2.5	4	4
TEKDÜZE DAĞILIM	Grup 1	Alt Sınır	0	0	0	0
		Üst Sınır	1	1	1	1
	Grup 2	Alt Sınır	0	0	0	0
		Üst Sınır	1.1	1.5	2	3

Bu çalışmada her bir örneklem büyüklüğü için 1000 tekrar yapılmış ve elde edilen sonuçların ortalamaları hesaplanmıştır. Üretilen verilere lojistik regresyon ve diskriminant analizleri uygulanarak, bireylerin hangi gruba ait olduklarına ilişkin olasılık tahminleri yapılarak, gruplara atama işlemi gerçekleştirilmiştir. Sonraki aşamada, elde edilen tahmini grup üyelikleri ile gerçek grup üyelikleri yardımıyla her iki analize ait B, C ve Q indeksleri ile Kappa katsayısı hesaplanarak karşılaştırılmıştır.

## 5. SONUÇLAR VE TARTIŞMA

İki grup arasında fark olmadığı durumda lojistik regresyon ve diskriminant analizlerinin her ikisi de atama işlemini rasgele olarak yapacaklardır. Bu nedenle Tablo E1’de yer alan indeks değerlerinin birbirine çok yakın olduğu görülmektedir.

Literatürde rasgele tahmin değerinin 0.5 olarak verildiği C indeksinin, fark “yok” düzeyi için aldığı değerler 0.65 civarındadır ve bu değer örneklem büyüklüğü arttıkça azalma eğilimi göstermektedir. Buna rağmen örneklem büyüklüğü 120 için aldığı en düşük 0.58 değeri, 0.5 değerinden büyük bir değerdir. Bu nedenle, C indeksinin rasgele tahmini belirlemede zayıf kaldığı söylenebilir. B indeksi için rasgele tahmin değeri 0.75, Q için 0 olarak verilir. Kappa için 0.20’den az değerler kötü uyumu ifade eder. Bu çalışmada, bu değer rasgele tahmine karşılık gelmektedir. Tüm indeks değerlerinde, C indeksinde olduğu gibi örneklem büyüklüğü arttıkça azalma söz konusudur.

Gruplar arası fark olmadığında, rastgele tahmin yapan lojistik regresyon ve diskriminant analizinden elde edilen sonuçlar çok benzerdir. Ancak gruplar arasındaki fark düzeyi arttıkça lojistik regresyonun diskriminant analizinden daha başarılı olduğu dikkat çeker. Bunun en önemli sebebi; LR’nin DA gibi açıklayıcı değişkenlere normal dağılma şartı koymamasıdır.

Tüm farklılık düzeylerinde en dikkat çeken nokta, indeks değerlerindeki en büyük değişimin örneklem büyüklüğü 20 ile 40 arasında iken olmasıdır. Bu nedenle indekslerin küçük örneklem büyüklüklerine karşı hassas olduğu söylenebilir. Örneklem büyüklüğü arttıkça tüm indeks değerlerindeki değişim de azalmaktadır. Örneklem büyüklüğünden en fazla etkilenen indeks, fonksiyonel yapısı sebebiyle C indeksi olmuştur. Çünkü C indeksinin payda kısmında iki grubun örneklem büyüklüklerinin çarpımı yer alır. Q indeksinde ise örneklem büyüklüğü 40’tan büyük olduğunda değişim azdır.

Fark düzeylerinin yüksek olması durumunda, beklendiği gibi tüm indeks değerleri iyi bir atama yapıldığını işaret etmektedir. Kappa'nın 0.61-0.80 arasında aldığı değerler iyi bir grup atamasının ifade eder. Lojistik regresyonun, diskriminant analizinden üstünlüğünü ifade eden en iyi indeks ise Q indeksidir. Bu indekslerle aynı amaçla kullanılabilmesi iddia edilen Kappa katsayısı, gruplar arasındaki fark "orta" düzeydeyken (Tablo E3), diskriminant analizi için lojistik regresyon analizi değerlerinden daha yüksek değerler vermiştir. Beklenenin aksi yönünde verdiği bu sonuca rağmen, diğer indekslere nazaran daha kolay hesaplanabilir olması ve çoğu zaman iyi sonuçlar elde etmesi nedeniyle, iki yöntemi karşılaştıran bir indeks olarak kullanılabilmesi düşünülmektedir.

Sonuç olarak, tahmin edilen olasılık değerlerini dikkate alınarak hesaplanan Q indeksinin dikkate alınan indeksler arasında en iyi indeks olduğu görülmüştür. Q indeksinin ardından, yine olasılıkları kullanan B indeksi iyi bir indekstir denilebilir.

## 6. KAYNAKLAR

Efron, B., 1975. The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis, Journal of the American Statistical Association, Vol. 70, No. 352, 892-898.

Pohar, M., Blas, M., Turk, S., 2004. Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study, Metodoloski Zvezki, Vol. 1, No. 1, 143-161.

Press, S. J., Wilson, S., 1978. Choosing Between Logistic Regression and Discriminant Analysis, Journal of the American Statistical Association, Vol. 73, No. 364, 699-705.

Gürçan, M., 1998. Lojistik Regresyon ve Bir Uygulama, Yüksek Lisans Tezi, On Dokuz Mayıs Üniversitesi 63s., Samsun.

McFadden, D., 1976. "A Comment on Discriminant Analysis 'Versus' Logit Analysis", Annals of Economic, and Social Measurement, 5, 511-523.

Özdamar, K., 1997. Paket Programlar ile İstatistiksel Veri Analizi, Anadolu Üniversitesi Yayınları 512 s., Eskişehir.

Subhash, S., 1996. Applied Multivariate Techniques, John Wiley & Son's, USA, p. 287-317.

Alpar, C. R., Gözükara, Bağ, H. G., Karabulut, E., 2010. 2x2 Tablolarda Gözlemciler / Gözlemler Arası Uyumun Değerlendirilmesi, Hacettepe Dış Hekimliği Fakültesi Dergisi Cilt: 34, Sayı: 1-2, Sayfa: 46-52.

## A COMPARISON OF SOME INDEXES USED IN LOGISTIC REGRESSION AND LINEAR DISCRIMINANT ANALYSIS

### ABSTRACT

*Logistic regression and linear discriminant analysis are two widely used methods to classify individuals or observations. The use those analyses for the same goal brings in mind the question of which analysis present better results. A comparison has been made and a study has been presented on this matter. Assumptions that are necessary for discriminant analysis, which are not necessary for logistic regression, made preferences switch under different conditions. In this paper, a simulation study has been carried out to make a comparison of these two methods and the indexes that are used to evaluate these analyses when the variables do not satisfy the normal distribution assumption. It is found that in general logistic regression analysis presents better results in comparison with the discriminant analysis method for data generated from multivariate non-normal distribution. As the sample size increases, the diversity of the results obtained from both analyses are considerably decreased. It is found that no matter what the sample size is logistic regression analysis has always been better in classification than the discriminant analysis method according to any index criteria. Cohen's Kappa coefficient has been used as a new index. In addition, when the better model is known, the indexes used are evaluated in terms of their success of estimating the true model.*

**Keywords:** Discriminant analysis, Cohen's Kappa, Logistic regression.

## Ek : Tablolar

Tablo E1. İki Grup Arası Fark Yokken İndeks Sonuçları

Örneklem Büyüklüğü	Analiz	C	B	Q	Kappa
20	LR	0.652690	0.772009	0.068327	0.222400
	DA	0.652700	0.771857	0.067631	0.221850
40	LR	0.612874	0.761998	0.036343	0.162175
	DA	0.612838	0.761970	0.036181	0.162200
60	LR	0.595916	0.758619	0.025678	0.135533
	DA	0.595866	0.758606	0.025627	0.135267
80	LR	0.591763	0.757949	0.023716	0.127813
	DA	0.591769	0.757943	0.023681	0.128375
100	LR	0.583337	0.756516	0.019278	0.115680
	DA	0.583335	0.756513	0.019262	0.115210
120	LR	0.580103	0.755996	0.017722	0.111783
	LR	0.580086	0.755994	0.017710	0.111633

Tablo E3. İki Grup Arası Fark Ortayken İndeks Sonuçları

Örneklem Büyüklüğü	Analiz	C	B	Q	Kappa
20	LR	0.845927	0.850054	0.346129	0.556700
	DA	0.842250	0.847358	0.330723	0.553800
40	LR	0.838228	0.843362	0.316924	0.541350
	DA	0.835191	0.841994	0.307992	0.543100
60	LR	0.833851	0.839913	0.303627	0.530433
	DA	0.831148	0.838865	0.296234	0.533233
80	LR	0.832706	0.839230	0.300703	0.528350
	DA	0.830234	0.838325	0.294164	0.531500
100	LR	0.831722	0.838650	0.298142	0.527240
	DA	0.829292	0.837861	0.292004	0.530770
120	LR	0.830432	0.837653	0.294297	0.524600
	LR	0.827956	0.836888	0.294297	0.528933

Tablo E2. İki Grup Arası Fark Azken İndeks Sonuçları

Örneklem Büyüklüğü	Analiz	C	B	Q	Kappa
20	LR	0.753422	0.804738	0.179261	0.380450
	DA	0.752115	0.804025	0.175499	0.378400
40	LR	0.734282	0.795829	0.146955	0.345100
	DA	0.733442	0.795514	0.145377	0.345300
60	LR	0.728986	0.793478	0.138392	0.337217
	DA	0.728134	0.793280	0.137380	0.336867
80	LR	0.725708	0.791935	0.133141	0.330775
	DA	0.724951	0.791810	0.132410	0.329913
100	LR	0.726404	0.792047	0.133451	0.331050
	DA	0.725529	0.791904	0.132687	0.331550
120	LR	0.723959	0.790927	0.129502	0.327800
	LR	0.723090	0.790803	0.128847	0.327300

Tablo E4. İki Grup Arası Fark Çokken İndeks Sonuçları

Örneklem Büyüklüğü	Analiz	C	B	Q	Kappa
20	LR	0.948777	0.923619	0.644286	0.791550
	DA	0.943107	0.910164	0.574960	0.758850
40	LR	0.942987	0.914602	0.601025	0.765050
	DA	0.939308	0.905190	0.553419	0.744700
60	LR	0.942759	0.913193	0.593044	0.763300
	DA	0.939907	0.904643	0.550469	0.742717
80	LR	0.941085	0.911116	0.583672	0.757188
	DA	0.938178	0.902791	0.542595	0.738738
100	LR	0.941284	0.911257	0.583109	0.759110
	DA	0.938671	0.903337	0.543702	0.738910
120	LR	0.941504	0.911213	0.583334	0.757825
	LR	0.938975	0.903327	0.544222	0.739225

# TAMAM OLMAYAN PİYASADA TEK PERİYOT KARESEL RİSKTEN KORUMA PROBLEMİNDE PARAMETRELERİN BELİRSİZLİĞİ ALTINDA DAYANIKLI OPTİMİZASYON YAKLAŞIMI

Gültaç EROĞLU İNAN\* Aysen APAYDIN\*\* Mustafa Ç. PINAR\*\*\*

## ÖZET

*Bu çalışmada, öncelikle tamam olmayan piyasada tek periyot için karesel riskten korunma problemi ele alındı. Problemin Föllmer-Schweizer (1989) tarafından elde edilen optimal çözümü verildi. Sonrasında, varlık fiyatı için Pinar (2006) tarafından önerilen bir stokastik model ile yazarlar tarafından problemin optimal çözümü elde edildi. Yine aynı çalışmada elde edilen bazı özel hesaplamalara yer verildi. Bir diğer aşamada, model parametrelerinin belirsizlik gösterdiği durum için, ilgili çalışmada önerilen dayanıklı optimizasyon yaklaşımı ele alındı. Son aşamada, sadece oynaklığın bilinmediği durum için bu yaklaşımın bir uygulamasına yer verildi. Sayısal sonuçlar, oynaklığın bilindiği durumda elde edilen klasik çözüm ile belirsizlik durumunda elde edilen çözümün, amaç fonksiyonuna yakın değerler verdiğini göstermiştir. Sonuç olarak, belirsizlik durumunda elde edilen çözümün, oynaklık değişimine karşı dayanıklı olduğunu söyleyebiliriz.*

**Anahtar Kelimeler:** Belirsizlik, Dayanıklı optimizasyon, Karesel riskten korunma.

## 1. GİRİŞ

Opsiyon, iki taraf arasında yapılan, bir finansal varlığın, gelecekte belirlenen bir tarihte (vade tarihi) bir prim karşılığında (opsiyonun fiyatı) önceden belirlenmiş bir fiyattan (uygulama fiyatı  $K$ ) satma veya satın alma hakkını veren bir sözleşmedir. Opsiyon sözleşmesi, taraflardan birine anlaşmadaki hakları kullanmak veya bu haklardan vazgeçmek seçeneği tanırken, diğer taraftan hak sahibinin sözleşmedeki haklarını kullanmayı seçmesi durumunda, anlaşmanın gereklerini yerine getirme sorumluluğu yüklemektedir. Bir opsiyonu satan tarafa, opsiyonu satan ya da opsiyonu yazan denilmektedir. Bir opsiyonu alan tarafa ise opsiyonu alan denilmektedir. Opsiyonlar, alıcı açısından alım opsiyonu ve satım opsiyonu olmak üzere ikiye ayrılmaktadır. Vade tarihi açısından ise opsiyonlar, Amerikan Opsiyonu ve Avrupa Opsiyonu olarak ayrılmaktadır. Avrupa Opsiyonları sadece vade tarihinde kullanılabilir. Amerikan Opsiyonları ise vade tarihinden önce kullanılabilir.

Bugün pek çok mal ve finansal varlık üzerine yazılmış opsiyonlar mevcuttur. Opsiyonlar gelecekte bir zamanda olası, fiyat dalgalanma beklentilerine karşı çeşitli stratejiler geliştirme imkanı vermektedir. Opsiyon anlaşmalarının, bir finansal varlık olarak yararları, risk aktarma, kar sağlama olarak sayılabilir. Opsiyon almanın hak sahibine sağladığı en büyük fayda, sınırlı bir riskle, sınırsız kar sağlama olanağı elde

\*Araş. Gör., Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü e-posta: [geroglu@science.ankara.edu.tr](mailto:geroglu@science.ankara.edu.tr)

\*\*Prof. Dr., Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü e-posta: [apaydin@science.ankara.edu.tr](mailto:apaydin@science.ankara.edu.tr)

\*\*\*Prof. Dr., Bilkent Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü e-posta: [mustafap@bilkent.edu.tr](mailto:mustafap@bilkent.edu.tr)

etmektedir. Opsiyon sözleşmesi satın alınan riski veya uğrayabileceği zarar ise ödediği prim ile sınırlıdır. Opsiyon sözleşmesi satan kişinin riski veya uğrayabileceği zararı ise sınırsız olabilmektedir (Alpan, 1999).

Bu risk, varlığın dönem sonundaki fiyatına bağlı olarak oluşabilecek kayıptır. Bu kayıp opsiyonun bedeli olarak tanımlanmaktadır ve ileriki bölümlerde  $H$  olarak gösterilecektir. Opsiyonu satan, kendisini bu riske karşı korumak için  $H$  opsiyonun bedelini dönem sonunda geri sağlayacak bir ticari strateji oluşturmaya çalışır.

Eğer içinde bulunulan piyasa tamam olmayan (complete market) bir piyasa ise, opsiyonun bedeli  $H$  geri kazanılabilir,  $V_M = H$  olacak şekilde kendi kendini finanse eden (self-financing strategy) bir ticari strateji vardır. Eğer piyasa tamam olmayan (incomplete market) bir piyasa ise  $H$  geri sağlanamaz, riskten korunma stratejisi bazı optimalite kriterlerine bağlı olarak seçilmelidir. Bu durumda, riskten korunma stratejisini oluşturmak için genellikle karesel risk minimizasyonu (quadratic risk minimization) kullanılmaktadır (Coleman vd., 2003).

## 2. YÖNTEM

Bu çalışmada, tamam olmayan piyasada tek periyotlu durumda varlık fiyatının hareketi için tanımlanan modelin parametrelerinde belirsizlik olduğu varsayımı altında karesel riskten korunma problemi ele alındı. Çalışmada belirsizlik durumu ile mücadele için optimizasyon yöntemi olarak, dayanıklı optimizasyon yöntemi kullanılmıştır.

### 2.1 Tek Periyot için Karesel Riskten Korunma Problemi

Bu bölümde, tamam olmayan piyasada, tek periyot için karesel riskten korunma problemi ele alındı. Burada ekonomi,  $k=0,1$  zamanlarında  $X_k$  maliyeti ile bir  $X$  riskli varlığı (stok) ve bir risksiz varlığı (bond or account) içermektedir.  $k=1$  zamanında  $X_1$ , stok maliyeti  $k=1$  zamanına kadar biriktirilmiş bilgi ile ölçülebilen bir rasgele değişkendir.  $k=0$  zamanında  $X_0$  değeri bilinmektedir yani bir  $x_0 > 0$  için  $P(X_0 = x_0) = 1$  dir.  $k=1$  zamanında, opsiyonun bedeli,  $H$  rasgele değişkeni ile tanımlanır. Bir alım opsiyonu için,

$$H = (X_1 - K)_+ = \begin{cases} X_1 - K & , \quad X_1 > K \\ 0 & , \quad X_1 \leq K \end{cases} \quad (1)$$

olarak tanımlanmaktadır. Burada  $K$ , uygulama (anlaşma) fiyatı olarak tanımlanan önceden belirlenen bir fiyattır.  $X_1 > K$  olması durumunda  $H = X_1 - K$  pozitif bir değer alacaktır.  $(X_1 - K)_+$  gösteriminde yer alan  $+$ , pozitiflik durumu için kullanılmıştır.

Opsiyonu satan taraf,  $X_1 > K$  olması durumunda, pazarda daha pahalı olan bir malı  $K$  anlaşma fiyatından karşı tarafa satarak  $H = X_1 - K$  zarar etmiş olacaktır. Bu olaya karşı, opsiyonu satan kendini garantiye almak ve opsiyonu korumak ister.

Bunun için kendisine, bir portföy oluşturur. Bu portföyün  $k=0$  zamanındaki değeri;

$$V_0 = \xi X_0 + \eta_0 \quad (2)$$

dır.

Burada;

$$\begin{aligned} \xi &: k=0 \text{ zamanında } X \text{ riskli varlığından satın alınan miktar} \\ \eta_0 &: k=0 \text{ zamanında risksiz varlığa yatırılan miktar} \end{aligned}$$

olarak tanımlanmaktadır. Risksiz varlığın maliyeti 1 olarak alınmaktadır. Açıklamak gerekirse, örneğin,  $\eta_0=1$  miktarda İsviçre Frankına yatırım yapıldığı düşünülebilir (Föllmer vd., 1989).

$k=1$  zamanında sonuçlanan portföyün değerinin  $V_1 = H$  olması istenir. Buradan;

$$\begin{aligned} V_1 = \xi X_1 + \eta_1 &\rightarrow H = \xi X_1 + \eta_1 \\ &\rightarrow \eta_1 = H - \xi X_1 \end{aligned} \quad (3)$$

olarak hesaplanır.

Verilen bir  $H$  için, bir strateji,  $\xi$  ve  $V_0$  sabitlerinin ilk baştaki seçimleriyle belirlenebilir.

Bir  $(\xi, V_0)$  stratejisinin neden olduğu maliyet incelenmek istensin.

$$C_k : "k. zamana kadar birikmiş maliyet"$$

olmak üzere

$k=0$  zamanından  $k=1$  zamanına kadar geçen sürede yapılan ilave maliyet

$$\begin{aligned} C_1 - C_0 &= \eta_1 - \eta_0 \\ &= (V_1 - \xi X_1) - (V_0 - \xi X_0) \\ &= H - V_0 - \xi(X_1 - X_0) \\ &= H - V_0 - \xi \Delta X \end{aligned} \quad (4)$$

olarak verilir. Burada problem, ticari stratejinin beklenen karesel ilave maliyetini en küçükleyecek,  $V_0$  ve  $\xi$  değerlerini belirlemektir.

Problem matematiksel olarak,

$$\min_{V_0, \xi} = E(C_1 - C_0)^2 = E(H - V_0 - \xi \Delta X)^2 \quad (5)$$

biçiminde ifade edilebilir.  
Problemin çözümü ise,

$$\xi = \frac{Cov(H, X_1)}{Var(X_1)} \quad (6)$$

$$V_0 = E(H) - \xi E(\Delta X) \quad (7)$$

olarak elde edilir (Föllmer vd., 1989). Föllmer-Schweizer çalışmalarında, çözüm için Doğrusal Regresyonda En Küçük Kareler yaklaşımını kullanmışlardır.

Burada özetlemek gerekirse;

$C_0$ ,  $k=0$  başlangıç zamanındaki nakite durumu,  $C_1$ ,  $k=1$  zamanındaki nakite durumu olmak üzere, başlangıç zamanından  $k=1$  zamanına,  $C_1 - C_0$  ilave maliyetinin, karesel beklenen değeri  $E(C_1 - C_0)^2 = E(H - V_0 - \xi \Delta X)^2$  en küçük olacak şekilde,

$$\xi = \frac{Cov(H, X_1)}{Var(X_1)}$$

$$V_0 = E(H) - \xi E(\Delta X)$$

olarak elde edilmiştir.

Karesel riskten korunma probleminin optimizasyonu alanında yapılan çalışmaların detaylı incelemesi için okuyucular (Schweizer, 2001) çalışmasına yönelebilir.

(Avellaneda vd., 1995), (Ahn vd., 1997) ve (Ahn vd., 1999) varlık fiyatının oynaklığının (volatility) tam olarak bilinmediği fakat bir belirsizlik aralığında yer aldığı durumda, opsiyon fiyatlama için dayanıklı riskten korunma stratejileri geliştirmişlerdir.

İlk çalışmada oynaklık tam olarak bilinmemekte, sadece  $\sigma_{\min}$  ve  $\sigma_{\max}$  uç değerleri arasında değer aldığı bilinmektedir. Bu çalışmada, oynaklık için aralık belirsizliği altında, Black-Scholes yapısında, varlık fiyatının, Black-Scholes-Barenblatt Eşitliği olarak tanımlanan bir lineer olmayan PDE'nin (Partial Differential Equations) çözümü olarak tanımlanabildiği gösterilmiştir. Eşitliğin çözümü için, sonlu fark algoritması (finite-differencing) ve üçgensel ağaç (trinomial tree) sunulmuştur.

İkinci çalışmada, Black-Scholes yaklaşımında, dayanıklı riskten koruma stratejisi geliştirmek için, üstel yarar fonksiyonu kullanılmıştır. En kötü oynaklık senaryosu için (worst-case volatility scenario) yatırımcının faydası en iyilenmeye çalışılmıştır. Çalışmada Avrupa Opsiyonları (European Option) ele alınmıştır.



Üçüncü çalışmada ise, bir önceki çalışmanın aksine yatırımcı sadece Black-Scholes stratejisi ile sınırlandırılmamaktadır. Burada Dinamik Programlama Eşitliği (Dynamic Programming Equation-DPE) yaklaşımı geliştirilmiştir. Bu yaklaşım sadece Avrupa Opsiyonlarında değil, Barrier Opsiyonları gibi opsiyonlarda da kullanılabilir. DPE yaklaşımının sonuçları, Black-Scholes yaklaşımına göre daha iyi çalışan etkili stratejiler elde edildiğini göstermiştir.

## 2.2 Varlık Fiyatının Hareketi için Temel Bir Stokastik Model

(Pınar, 2006), çalışmasında varlık fiyatının hareketini  $X_1 = X_0(1 + \mu + \sigma Z)$  stokastik modeli olarak tanımlamıştır. Burada

$\mu$ : Varlık fiyatının eğilimi (drift)  
 $\sigma$ : Varlık fiyatının sapması (volatility)  
 $Z \sim N(0,1)$ : Standart Normal Dağılım

biçiminde tanımlıdır.  $\Delta X = X_1 - X_0 = X_0(\mu + \sigma Z)$ , Avrupa alım opsiyonu için,  $K$  uygulama fiyatı ile,  $H = (X_1 - K)_+ = (X_0(1 + \mu + \sigma Z) - K)_+$  dir. Çalışmada  $E(H - V_0 - \xi \Delta X)^2$  değeri

$$A + B(-2V_0 - 2\xi X_0) + C(-2\xi X_0 \sigma) + V_0^2 + 2V_0 \xi X_0 \mu + \xi^2 (X_0^2 \mu^2 + X_0^2 \sigma^2) \quad (8)$$

olarak elde edilmiştir.

Burada,

$$\begin{aligned} A &= [X_0^2(1 + 2\mu + \mu^2) - 2K(1 + \mu) + K^2](1 - F_Z(b)) \\ &\quad + [X_0^2(2\sigma + 2\mu\sigma) - 2KX_0\sigma]f_Z(b) + X_0^2\sigma^2(1 - F_Z(b) + bf_Z(b)) \\ B &= [X_0(1 + \mu) - K](1 - F_Z(b)) + X_0\sigma f_Z(b) \\ C &= [X_0(1 + \mu) - K]f_Z(b) + X_0\sigma(1 - F_Z(b) + bf_Z(b)) \end{aligned}$$

biçiminde elde edildi.

Burada  $f_Z$ ,  $Z$  rasgele değişkeninin olasılık yoğunluk fonksiyonu,  $F_Z$   $Z$  rasgele değişkeninin dağılım fonksiyonu,

$$b = \frac{K}{X_0\sigma} - \frac{1 + \mu}{\sigma} \text{ dir (Pınar, 2006).}$$

Bu çalışmamızda yazarlar tarafından,

$k = 0$  zamanında  $E(H - V_0 - \xi \Delta X)^2$  karesel beklenen değeri minimum olacak şekilde

$$\xi = \left( \frac{\begin{pmatrix} X_0^2 \sigma^2 (1 - F_Z(b) + b f_Z(b)) + X_0^2 \sigma (1 + \mu) f_Z(b) \\ -K X_0 \sigma (1 - F_Z(b)) X_0 (1 + \mu) \end{pmatrix}}{X_0^2 \sigma^2} \right) \quad (9)$$

olarak, portföyün değeri ise

$$V_0 = X_0 (1 + \mu) (1 - F_Z(b)) + X_0 \sigma f_Z(b) - K (1 - F_Z(b)) - \xi X_0 \mu \quad (10)$$

olarak elde edilmiştir. Çıkarımlar, parçalı momentler (Winkler, 1972) yardımıyla elde edilmiştir.

### 2.3 Dayanıklı Optimizasyon

Optimizasyon algoritmaları ve yazılımlar kullanıcıya, kompleks optimizasyon problemlerinin geniş ölçüde ele alınmasına olanak vermesine rağmen, optimal çözümler problem girdilerindeki ufak değişikliklere oldukça duyarlı olabilmektedir. Bu duyarlılığın nedeni ise gerçek dünya verilerinin nadiren kesin değerli olmasıdır.

Klasik doğrusal optimizasyon yönteminde, küçük veri belirsizlikleri (%1 veya daha az) göz ardı edilir. Verilen (sözde) veriler kesin ise problem çözülür ve sonuçlanan nominal optimal çözümdür. Küçük veri belirsizliklerinin çözümün uygunluk ve optimalite özelliklerini önemli ölçüde etkilemeyeceği umulur. Fakat bazen küçük veri belirsizlikler, dikkat edilmesi gereken önemli sonuçlara neden olmaktadır.

Dayanıklı optimizasyon belirsizliğin kümeler yoluyla bir tarifine dayanan, en son geliştirilen bir tekniktir. Belirsiz parametrelerin sadece bilinen kümelere ait olduğu bilinmektedir. Belirsiz problemin dayanıklı karşılığı üzerinde durulmaktadır. Bu dayanıklı karşılıklar aslında, orijinal problemin en kötü durum formülasyonlarıdır.

Bir belirsiz doğrusal optimizasyon problemi,

$$\left\{ \min_x \{c^t x + d : Ax \leq b\} \right\}_{(c, d, A, b) \in U} \quad (11)$$

ifade edilsin. Burada,  $U$  bir belirsizlik kümesidir. Klasik optimizasyon probleminin dayanıklı karşılığı,

$$\min_x \left[ \max_{(c, d, A, b) \in U} \{[c^t x + d]\} : Ax \leq b \quad \forall (c, d, A, b) \in U \right] \quad (12)$$

olarak tanımlıdır. Problem

$$\min_{t, x} \{t / c^t x + d \leq t, Ax - b \geq 0 \quad \forall (c, d, A, b) \in U\} \quad (13)$$

olarak da verilebilir (Ben Tal vd., 2009, 9).

## 2.4 Parametrelerin Belirsizliği Altında Dayanıklı Optimizasyon Yaklaşımı

Bu bölümde,  $\mu$  ve  $\sigma$  parametrelerinin her ikisinde belirsizlik olduğu durumda,  $V_0$  ve  $\xi$  değerlerinin elde edilmesine ilişkin Pınar (2006) dayanıklı yaklaşımı ele alındı. Çalışmada dayanıklı karesel opsiyon koruma problemi

$$\min_{V_0, \varepsilon} \max_{(\mu, \sigma)} \phi(V_0, \varepsilon, \mu, \sigma) = E_Z \left\{ \left[ (X_0(1 + \mu + \sigma Z) - K)_+ - V_0 - \varepsilon X_0(\mu + \sigma Z) \right]^2 \right\} \quad (14)$$

olarak, belirsizlik kümesi ise

$$\begin{pmatrix} \mu \\ \sigma \end{pmatrix} \in U = \left\{ \begin{pmatrix} \mu \\ \sigma \end{pmatrix} : \begin{pmatrix} \mu \\ \sigma \end{pmatrix} = \begin{pmatrix} - \\ - \\ \sigma \end{pmatrix} + P_u : \|u\|_2 \leq 1 \right\} \quad (15)$$

biçiminde tanımlıdır (Pınar, 2006, 5).

**Önerme :**  $V_0^*$ ,  $\xi^*$  çifti, (14) probleminde optimaldir; eğer ve sadece,

$$\max_{(\mu, \sigma)} \phi(V_0^*, \varepsilon^*, \mu, \sigma) = \phi(V_0^*, \varepsilon^*, \mu_j, \sigma_j), \quad \text{olacak ve } j=1, \dots, r \text{ } r \text{ negatif olmayan}$$

sayılar  $\alpha_j \geq 0$ ,  $\sum_{j=1}^r \alpha_j = 1$  olacak şekilde  $1 \leq r \leq 3$  için  $(\mu_1, \sigma_1), \dots, (\mu_r, \sigma_r)$  çiftleri varsa;

öyle ki

$$\xi^* = \frac{\sum_{j=1}^r \alpha_j \text{Cov}_j(H, X_1)}{\sum_{j=1}^r \alpha_j \text{Var}_j(H, X_1)} \quad (16)$$

$$V_0^* = \sum_{j=1}^r \alpha_j E_j(H) - \xi^* \sum_{j=1}^r \alpha_j E_j(\Delta X) \quad (17)$$

dır. Burada;

$$E_j(H) = E(H(\mu_j, \sigma_j)), \quad E_j(\Delta X) = E(\Delta X(\mu_j, \sigma_j))$$

$$\text{Cov}_j(H, X_1) = \text{Cov}(H(\mu_j, \sigma_j), X_1(\mu_j, \sigma_j)) \quad j = 1, \dots, r$$

dır.

**İspat:** Sabit  $\mu, \sigma$  değerleri için,  $V_0, \xi$  'ın fonksiyonu olan

$$g_Z = \left\{ \left[ (X_0(1 + \mu + \sigma Z) - K)_+ - V_0 - \varepsilon X_0(\mu + \sigma Z) \right]^2 \right\}, \quad V_0, \xi \text{ da birlikte konvektir ve}$$

$Z$  etrafında beklenen değer konveksliği korumaktadır. Bir konveks fonksiyonun noktasal supremumu aynı zamanda konveks bir fonksiyon olduğundan,  $\mu, \sigma$  etrafında maksimizasyon  $V_0, \xi$  da fonksiyonun konveks olmasını sağlar. Demyanov ve Malozemov (1990, 196-197)'e göre  $V_0^*, \xi^*$ , (14) probleminde optimaldir, öyle ki;

$$\sum_{j=1}^r \alpha_j \left( \frac{\partial E_z [g_z(\mu_j, \sigma_j, V_0^*, \xi^*)]}{\partial \xi^*} \right) = 0 \quad (18)$$

$$\sum_{j=1}^r \alpha_j \left( \frac{\partial E_z [g_z(\mu_j, \sigma_j, V_0^*, \xi^*)]}{\partial V_0^*} \right) = 0 \quad (19)$$

formülleri çözüldüğünde (16) ve (17) kolaylıkla elde edilir.

### 3. BULGULAR

Bu bölümde öncelikle, varlık fiyatının hareketinin  $X_1 = X_0(1 + \mu + \sigma Z)$  stokastik modeline dayandığı varsayımı altında tek periyot için klasik riskten korunma probleminin optimizasyonu yapılacaktır.  $\sigma$  oynaklık parametresinin bazı değerleri için çalışmanın ikinci bölümünde elde edilen optimal  $\xi, V_0$  değerleri ve  $E(H - V_0 - \xi \Delta X)^2$  amaç fonksiyonu değerleri verilecektir.

Sonraki aşamada;  $\sigma$  oynaklık parametresinin kesin olarak bilinmediği; sadece hangi aralıkta değer aldığı bilindiği durumda dayanıklı riskten korunma probleminin optimizasyonu ele alınacaktır.

Son olarak elde edilen klasik ve dayanıklı çözümler amaç fonksiyonuna verdikleri değerler açısından karşılaştırılacaktır.

İlk olarak, çalışmanın ikinci bölümü, 2.2 kısmında, varlık fiyatının hareketinin  $X_1 = X_0(1 + \mu + \sigma Z)$  modeline dayandığı varsayımı altında,  $\min_{V_0, \xi} = E(C_1 - C_0)^2 = E(H - V_0 - \xi \Delta X)^2$  probleminin optimal çözümü yazarlar tarafından,

$$\xi = \left( \frac{\left( X_0^2 \sigma^2 (1 - F_z(b) + b f_z(b)) + X_0^2 \sigma (1 + \mu) f_z(b) \right)}{-K X_0 \sigma (1 - F_z(b)) X_0 (1 + \mu)} \right) \bigg/ X_0^2 \sigma^2 \quad (20)$$

olarak, portföyün değeri ise

$$V_0 = X_0 (1 + \mu) (1 - F_z(b)) + X_0 \sigma f_z(b) - K (1 - F_z(b)) - \xi X_0 \mu \quad (21)$$

olarak elde edildiği hatırlatılsın. Çalışmada;  $X_0 = 10, K = 9, \mu = 0.1, 0.1 \leq \sigma \leq 0.2$  değerleri keyfi olarak alınmıştır. Bu varsayımlar altında, bazı  $\sigma$  değerleri için optimal çözümler  $\xi, V_0$  ve  $Z$  amaç fonksiyonu değerleri Tablo 1’de verilmiştir.

**Tablo 1.**  $k=0$  zamanında optimal riskli varlık miktarı, portföyün değeri ve amaç fonksiyonu değerleri

$\sigma$	$\xi$	$V_0$	$Z$
0.10	0.9772	1.0312	0.0052
0.11	0.9655	1.0495	0.0101
0.12	0.9522	1.0716	0.0174
0.13	0.9380	1.0968	0.0273
0.14	0.9234	1.1248	0.0399
0.15	0.9088	1.1548	0.0553
0.16	0.8944	1.1866	0.0735
0.17	0.8803	1.2198	0.0944
0.18	0.8667	1.2541	0.1180
0.19	0.8537	1.2893	0.1442
0.20	0.8413	1.3253	0.1729

İkinci aşamada, kabul edilen değerler altında 4. Bölümde (14) problemi olarak tanımlı problemde, sadece oynaklığın bilinmediği varsayılarak,

$$\min_{V_0, \xi} \max_{\sigma} E_Z \left\{ \left[ \left( H = (X_1 - K)_+ \right) - V_0 - \xi X_0 (\mu + \sigma Z) \right]^2 \right\} \quad (22)$$

dayanıklı optimizasyon probleminin optimal çözümü ele alındı. (Pinar, 2006) yaklaşımı olan ilgili önermenin uygulanmasıyla elde edilen optimal çözüm Tablo 2’de verilmiştir.

**Tablo 2.** Problem (22) için optimal çözüm

$\sigma$	$(\xi^*, V_0^*)$	$Z$
0.20	(0.8413, 1.3253)	0.1729

Optimal çözümün, oynaklığın bazı farklı değerleri için amaç fonksiyonuna verdiği değerler Tablo 3’de verilmiştir.

**Tablo 3.** Optimal çözümün, oynaklığın bazı farklı değerleri için amaç fonksiyonu değerleri

$\sigma$	$(\xi^* = 0.8413, V_0^* = 1.3253)$ için
$Z$ amaç fonksiyonu değerleri	
0.10	0.0306
0.11	0.0322
0.12	0.0346
0.13	0.0381
0.14	0.0431
0.15	0.05
0.16	0.0591
0.17	0.0796
0.18	0.0847
0.19	0.1016
0.20	$\max_{0.1 \leq \sigma \leq 0.2} E_Z \left\{ [H - V_0 - \xi \Delta X]^2 \right\} = 0.1729$

Bazı  $(\xi, V_0)$  değerleri için amaç fonksiyonunun en büyük değerleri Tablo 4’de verilmiştir.

**Tablo 4. Bazı riskli varlık miktarı, portföyün başlangıç değerleri için, amaç fonksiyonunun aldığı en büyük değerler**

$(\xi, V_0)$	$\max_{0.1 \leq \sigma \leq 0.2} E_Z \left\{ [H - V_0 - \xi \Delta X]^2 \right\}$
(0.9862,1.0217)	0.2718
(0.9665,1.0495)	0.2583
(0.9522,1.0716)	0.2425
(0.9380,1.0968)	0.2277
(0.9234,1.1248)	0.2139
(0.9088,1.1548)	0.2017
(0.8944,1.1866)	0.1915
(0.8803,1.2541)	0.1834
(0.8667,1.2541)	0.1776
(0.8537,1.2893)	0.1741
(0.8413,1.3253)	$\min_{V_0, \xi} \max_{0.1 \leq \sigma \leq 0.2} E_Z \left\{ [H - V_0^* - \xi^* \Delta X]^2 \right\} = 0.1729$

Her bir  $\sigma$  için amaç fonksiyonun klasik ve dayanıklı çözümde aldığı değerler ve farkları Tablo 5’de verilmiştir.

**Tablo 5. Her bir oynaklık değeri için klasik ve dayanıklı çözümlerin amaç fonksiyonu değerleri ve farkları**

$\sigma$	$\xi$	$V_0$	$Z_0$	$Z_1$	$Z_1 - Z_0$
(0.8413,1.3253) için					
0.10	0.9772	1.0312	0.0052	0.0487	0.0435
0.11	0.9655	1.0495	0.0101	0.0517	0.0416
0.12	0.9522	1.0716	0.0174	0.0555	0.0381
0.13	0.9380	1.0968	0.0273	0.0604	0.0331
0.14	0.9234	1.1248	0.0399	0.0671	0.0272
0.15	0.9088	1.1548	0.0553	0.0761	0.0208
0.16	0.8944	1.1866	0.0735	0.0880	0.0145
0.17	0.8803	1.2198	0.0944	0.1032	0.0088
0.18	0.8667	1.2541	0.1180	0.1222	0.0042
0.19	0.8537	1.2893	0.1442	0.1453	0.0011
0.20	0.8413	1.3253	0.1729	0.1729	0

#### 4. TARTIŞMA VE SONUÇ

Tablo 1'e göre,  $\sigma$  oynaklık değeri arttıkça portföyün değeri  $V_0$  değeri de artmaktadır. Tablo 5, bazı  $0.1 \leq \sigma \leq 0.2$  değerleri için, elde edilen ( $\xi^* = 0.8413, V_0^* = 1.3253$ ) dayanıklı çözümünün ve klasik çözümlerin amaç fonksiyonuna verdiği değerleri içermektedir.  $\sigma$  oynaklığının kesin olarak bilindiđi durumda amaç fonksiyonunun aldığı değer  $Z_0$ , belirsiz olduđu durumda, ( $\xi^* = 0.8413, V_0^* = 1.3253$ ) için amaç fonksiyonunun değeri  $Z_1$  olarak tanımlanmıştır. Tablo 5'e göre,  $Z_1 - Z_0$  farkları incelendiğinde; elde edilen değerlerin küçük olduđu gözlemlenmiştir. Burada,  $\sigma$  oynaklığının kesin olarak bilindiđi durumda ve belirsizlik durumunda elde edilen klasik ve dayanıklı çözümlerin birbirine yakın olduđu söylenebilir. Böylelikle elde edilen ( $\xi^* = 0.8413, V_0^* = 1.3253$ ) çözümünün, parametre belirsizliğinden etkilenmediđi, parametre belirsizliğine karşı dayanıklı bir çözüm olduđu rahatlıkla söylenebilir. Sonraki çalışmalarda, sapma ve oynaklığın aynı anda belirsizlik gösterdiđi durum incelenebilir. Farklı stokastik modeller ele alınabilir. Tek periyot durumunda yapılan çalışma, çoklu periyot durumu için genelleştirilebilir.

#### 5. KAYNAKLAR

- Ahn, H., Muni, A., Swindle, G., 1997. Misspecified Asset Price Models and Robust Hedging Strategies, Applied Mathematical Finance Vol. 4, 21-36.
- Ahn, H., Muni, A., Swindle, G., 1999. Optimal Hedging Strategies for Misspecified Asset Price Models, Applied Mathematical Finance, Vol.6, 197-208.
- Avellaneda, M., Levy., A., Paras, A., 1995. Pricing and Hedging Derivative Securities in Markets with Uncertain Volatilities, Applied Mathematical Finance, Vol. 2, 73-88.
- Alpan, F., 1999. Örneklerle Futures Anlaşmalar ve Opsiyonlar. Literatür Yayınları, İstanbul.
- Ben-Tal, A., El Ghaoui, L., Nemirovski, A., 2009. Robust Optimization, Princeton University Press, USA.
- Coleman, T. F., Li, Y., Patron, M. C., 2003. Discrete Hedging Under Piecewise Linear Risk Minimization, Journal of Risk, Vol. 5, 39-65.
- Dem'yanov, V. F., Malozemov, V. N., 1990. Introduction to Minimax. Dover Publications, New York.
- Föllmer, B. Y., Schweizer, M., 1989. Hedging by Sequential Regression: An Introduction to the Mathematics of Option Trading, ASTIN Bulletin, Vol.18, No.2; 147-160.

Pınar, M. Ç., 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 (Bilkent University 2006-Princeton University 2008).

Schweizer, M., 2001. Variance-optimal Hedging in Discrete Time. Mathematics of Operation Research, Vol. 20, 1-32.

Winkler R. L., Roodman G., Britney, R., 1972. The Determination of Partial Moments, Management Science, Vol. 19, No. 3, 290-296.

## ROBUST OPTIMIZATION APPROACH FOR ONE PERIOD QUADRATIC OPTION HEDGING PROBLEM UNDER THE UNCERTAINTY OF PARAMETERS IN INCOMPLETE MARKET

### ABSTRACT

*In this study, first the quadratic hedging problem for one period is considered in an incomplete market. Optimal solution of the problem obtained by Föllmer-Schweizer (1989) is given. Next, the optimal solution of the problem is obtained by the authors with a stochastic model proposed by Pınar (2006). Some special calculations obtained in the same study were included. In another stage, the robust optimization approach, which is suggested in the related work is discussed for the situation where the model parameters are subject to uncertainty. In the last stage, an application of this approach is included for the situation where only the volatility is unknown. Numerical results show that classical solution obtained for the situation when volatility is known and the solution obtained under the uncertainty situation give similar values at the objective function. As a result, we can say that, the solution obtained under the uncertainty situation is robust to volatility variation.*

**Keywords: Uncertainty, Robust optimization, Quadratic option hedging.**



## BULANIK ZAMAN SERİLERİ İLE ANKARA HAVA KALİTESİ VERİSİNİN ÇÖZÜMLENMESİ

Sibel ALADAĞ\* Çağdaş Hakan ALADAĞ\*\* Erol EĞRİOĞLU\*\*\*

### ÖZET

*Son yıllarda, hava sıcaklığı ve borsa verileri gibi belirsizlik içeren zaman serilerinin çözümlenmesinde bulanık zaman serileri başarıyla kullanılmaktadır. Literatürde, daha güvenilir öngörüler elde etmek amacıyla, bulanık zaman serisi yaklaşımları üzerinde çalışılmaktadır. Güvenilir öngörüler elde edilmesinde anahtar noktalardan biri, aralık uzunluğunun doğru olarak belirlenebilmesidir. Bu problemi çözmek amacıyla literatürde önerilen bazı yaklaşımlar mevcuttur. Hava kalitesi verileri de yapılarından dolayı belirsizlik içeren zaman serileridir ve bu tür zaman serilerinin çözümlenmesinde de bulanık zaman serilerinin kullanılması gerekmektedir. Yapılan bu çalışmada, Ankara hava kalitesi verisi ilk defa bulanık zaman serileri yöntemiyle çözümlenmiştir. Çözümlemede aralık uzunluğu için farklı teknikler kullanan altı farklı bulanık zaman serisi öngörü yöntemi kullanılmış ve elde edilen sonuçlar karşılaştırılarak yorumlanmıştır.*

**Anahtar Kelimeler:** Ankara hava kalitesi, Aralık uzunluğu, Bulanık zaman serileri, Optimizasyon, Öngörü.

### 1. GİRİŞ

Borsa verileri, hava sıcaklığı değerleri ya da hava kirliliği ölçümleri gibi zaman serileri belirsizlik içeren serilerdir. Bu çeşit verilerde gözlem değerleri aynı gün hatta aynı saat içinde bile değişiklik gösterebilmektedir. Bu tarz serilerin çözümlenmesinde klasik zaman serisi yaklaşımları yerine bulanık zaman serilerinin kullanılması gerekmektedir (Aladağ vd., 2009).

Bulanık zaman serisi yaklaşımı ilk olarak (Song vd., 1993a), (Song vd., 1993b), (Song vd., 1994) tarafından bulanık küme teorisine (Zadeh, 1965) dayalı olarak önerilmiştir. Çok karmaşık matris işlemleri içeren bu yöntem, (Chen, 1996) tarafından daha basit bir şekle getirilmiştir. Chen çalışmasında, Song ve Chissom tarafından geliştirilen yöntemdeki karmaşık matris işlemleri yerine, bulanık ilişki tablosundan yararlanma fikrini ortaya koymuştur (Chen, 1996).

Bulanık zaman serileri yaklaşımına olan ilgi giderek artmakta ve çeşitli alanlarda birçok uygulama alanı bulmaktadır (Aladağ vd., 2010). Daha güvenilir öngörülere ulaşmak amacıyla literatürde önerilen çeşitli bulanık zaman serisi yaklaşımları bulunmaktadır (Eğrioğlu vd., 2009). Uygulamalarda daha iyi öngörüler elde etmek amacıyla bulanık zaman serileri üzerinde çalışmalar devam etmektedir. Güvenilir öngörülere ulaşılmasında anahtar noktalardan biri aralık uzunluğunun doğru belirlenmesidir (Eğrioğlu vd., 2011). Seçilen aralık uzunluğu, kullanılacak bulanık zaman serisi modelinin öngörü performansını doğrudan etkileyen bir etkidir. Aralık uzunluğunun

\*Sos. Gv. Uzm. Yrd., Sosyal Gvenlik Kurumu, HSGM, Elektronik Yayın ve Bilgi Dağıtım Daire Başkanlığı, Ankara, e-posta: [saladag@sgk.gov.tr](mailto:saladag@sgk.gov.tr)

\*\*Doç. Dr., Hacettepe niversitesi, Fen Fakltesi, İstatistik Blm, Beytepe, Ankara, e-posta: [aladag@hacettepe.edu.tr](mailto:aladag@hacettepe.edu.tr)

\*\*\*Doç. Dr., Ondokuz Mayıs niversitesi, Fen Edebiyat Fakltesi, İstatistik Blm, Samsun, e-posta: [erole@omu.edu.tr](mailto:erole@omu.edu.tr)

belirlenmesi problemini çözmek amacıyla literatürde önerilen çeşitli yaklaşımlar mevcuttur. Aralık uzunluğunun belirlenmesi için, Huarng (2001) ortalamaya ve dağılıma dayalı yaklaşımlar önermiştir. Huarng ve Yu (2006) aralık uzunluğunun belirlenmesi için orana dayalı yeni bir yaklaşım geliştirmiştir. Literatürde aralık uzunluğunun belirlenmesi için bir başka yaklaşım ise Eğrioğlu vd. (2011) tarafından önerilmiştir. Eğrioğlu vd. (2011) çalışmalarında, en iyi aralık uzunluğunu belirlemek için kısıtlı optimizasyona dayalı bir yaklaşım kullanarak, Chen (1996) tarafından önerilen yöntemi geliştirmişlerdir.

Özellikle büyük şehirlerde hava kalitesi verilerinin değerlendirilmesi, hem insan sağlığı, hem çevre koşulları, hem de yapılacak yatırımlar açısından büyük önem taşımaktadır. Hava kalitesi verileri yapılarından ötürü belirsizlik içermektedirler ve klasik zaman serisi yöntemleri bu tür serilerin içerdiği belirsizliği modellemede yetersiz kalmaktadır. Bu nedenle, şehirlerin hava kalitesi verilerinin çözümlenmesinde bulanık zaman serileri yaklaşımlarından yararlanmak çok daha akıllıca olacaktır. Yapılan çalışmada, Ankara hava kalitesi verisi ilk kez bulanık zaman serileri yöntemiyle çözümlenmiştir. Çözümlemede kullanılan bulanık zaman serisi yöntemleri Song ve Chissom (1993a), Chen (1996), Huarng (2001), Huarng ve Yu (2006) ve Eğrioğlu vd. (2011) tarafından önerilen yaklaşımlardır. Huarng (2001) çalışmasında hem ortalamaya hem dağılıma dayalı olmak üzere iki farklı yöntem önermiştir ve bu iki yöntemde yapılan çalışmada kullanılmıştır. Buna göre, Ankara hava kalitesi verisi toplamda altı farklı bulanık zaman serisi yaklaşımıyla öngörülmüştür. Uygulama sonucunda elde edilen sonuçlar incelenmiş ve yorumlanmıştır.

Makalenin ikinci bölümünde, bulanık zaman serileri ve temel tanımları kısaca verilmiştir. 2.1'de Chen (1996) tarafından önerilen bulanık zaman serisi öngörü yöntemi özetlenmiştir. 2.2'de Eğrioğlu vd. (2011) tarafından önerilen optimizasyona dayalı aralık uzunluğu belirleme yaklaşımını tanıtılmıştır. Yapılan uygulama ve elde edilen sonuçlar üçüncü bölümde verilmiştir. Son bölümde ise elde edilen bulgular ışığında ulaşılan sonuçlar değerlendirilmiştir.

## 2. BULANIK ZAMAN SERİLERİ

Aladağ vd. (2010) tarafından bulanık zaman serileri ve temel bazı tanımları aşağıdaki gibi verilmiştir.

$U = \{u_1, \dots, u_b\}$  evrensel küme olsun.  $U$ 'nun elemanları aralıklardır. Bu aralıklar zaman serisinin tüm değerlerini kapsayan evrensel kümenin parçalanması ile elde edilir.

$U$ 'nun elemanlarına bağlı olarak  $A_i$  bulanık kümeleri aşağıdaki gibi tanımlanabilir.

$$A_i = f_{A_i}(u_1)/u_1 + \dots + f_{A_i}(u_b)/u_b$$

Burada  $f_{A_i}$ ,  $A_i$  bulanık kümesinin üyelik fonksiyonudur ve  $f_{A_i} : U \rightarrow [0,1]$  olmaktadır.  $f_{A_i}(u_a)$ ,  $u_a$ 'nın  $A_i$  bulanık kümesine ait olmasının üyelik derecesidir.

**Tanım 1.**  $Y(t)$  ( $t = \dots, 0, 1, 2, \dots$ ) reel sayıların bir alt kümesi olsun ve  $A_i$  bulanık kümelerinden oluşan evrensel kümeyi temsil etsin.  $A_j$  ( $j = 1, 2, \dots$ ) bulanık kümelerinden

oluşan  $F(t)$ ,  $Y(t)$  ( $t = \dots, 0, 1, 2, \dots$ ) evrensel küme üzerindeki bulanık zaman serisi olarak adlandırılır.

**Tanım 2.** \* herhangi bir operatörü göstermek üzere, eğer  $F(t) = F(t-1) * R(t, t-1)$  olacak şekilde bir  $R(t, t-1)$  bulanık ilişkisi varsa,  $F(t)$  sadece  $F(t-1)$ 'den etkilenmektedir. Bu durumda  $F(t)$  birinci dereceden bulanık zaman serisi olarak da adlandırılabilir.  $F(t-1) = A_i$  ve  $F(t) = A_j$  olduğunda,  $F(t-1)$  ile  $F(t)$  arasındaki bulanık ilişki  $A_i \rightarrow A_j$  şeklinde de gösterilebilir.  $A_i$  bulanık ilişkinin sol yanı ve  $A_j$  bulanık ilişkinin sağ yanı olarak isimlendirilir.

## 2.1 Chen'in Bulanık Zaman Serisi Öngörü Yöntemi

Tanım 1 ve Tanım 2'ye bağlı olarak, Chen (1996) tarafından önerilen birinci dereceden bulanık zaman serisi yaklaşımı aşağıda verildiği gibi altı adımda özetlenebilir.

**Adım 1.**  $U = \{u_1, u_2, \dots, u_b\}$  evrensel küme ve evrensel kümenin elemanları olan alt aralıklar  $(u_i, i = 1, 2, \dots, b)$  tanımlanır. Evrensel kümenin başlangıç ve bitiş noktaları belirlenir. Buna göre evrensel küme

$$U = [\text{Başlangıç}, \text{Bitiş}]$$

şeklinde ifade edilebilir. Daha sonra, belirlenen uygun aralık uzunluğuna göre evrensel küme eşit alt aralıklara bölünür.

**Adım 2.** Evrensel küme ve belirlenen alt aralıklara bağlı olarak  $A_i$  bulanık kümeler

$$A_i = f_{A_i}(u_1)/u_1 + \dots + f_{A_i}(u_b)/u_b$$

şeklinde tanımlanır.

**Adım 3.** Gözlemler bulanıklaştırılır. Örneğin, eğer bir gözlemin en büyük üyelik derecesi  $A_j$  bulanık kümesinde ise, bu gözlemin bulanık değeri  $A_j$  olarak belirlenir.

**Adım 4.** Bulanık ilişki ve grup ilişki tablosu oluşturulur. Gruplandırma için örnek verilirse, örneğin  $A_1 \rightarrow A_2$ ,  $A_1 \rightarrow A_1$  ve  $A_1 \rightarrow A_3$  ilişkileri bulunsun, bu ilişkiler

$$A_1 \rightarrow A_1, A_2, A_3$$

biçiminde gruplandırılabilir.

**Adım 5.** Öngörüler elde edilir. Öngörü elde edilirken birkaç durum söz konusudur.  $F(t-1) = A_j$  olsun. Buna göre, öngörü elde ederken oluşabilecek durumlar aşağıda verilmiştir.

*Durum 1.* Bulanık mantık ilişki tablosunda sadece bir ilişki olabilir. Örneğin, sadece  $A_j \rightarrow A_i$  ilişkisi varsa,  $F(t)$  öngörü değeri  $A_i$ 'dir.

*Durum 2.* Bulanık grup ilişki tablosunda  $A_j \rightarrow A_i, A_j, \dots, A_k$  ilişkisi varsa,  $F(t)$  öngörü değeri  $A_i, A_j, \dots, A_k$  'dir.

*Durum 3.* Bulanık grup ilişki tablosunda  $A_j \rightarrow Boş$  ise  $F(t)$  öngörü değeri  $A_j$  'dir.

**Adım 6.** Berraklaştırma işlemi uygulanır. Bir önceki adım sonunda elde edilen bulanık sayılar, berraklaştırma işlemi kullanılarak gerçel sayılara dönüştürülür. Berraklaştırma işlemi için merkezileştirme yöntemi kullanılır.

## 2.2 Eğrioğlu vd. Yöntemi

Eğrioğlu vd. (2011) yaptıkları çalışmada, aralık uzunluğunu belirlemek için kısıtlı optimizasyona dayalı bir yöntem önermişlerdir. Aralık uzunluğu optimizasyona dayalı yaklaşımla bulunduktan sonra, Chen (1996) tarafından önerilen yöntemi kullanarak öngörülerini hesaplamışlardır. Eğrioğlu vd. (2011) tarafından kullanılan optimizasyon yöntemi altın arama yöntemine dayalı bir süreçtir. Belirtilen süreçte,

$\text{Min } f(x)$

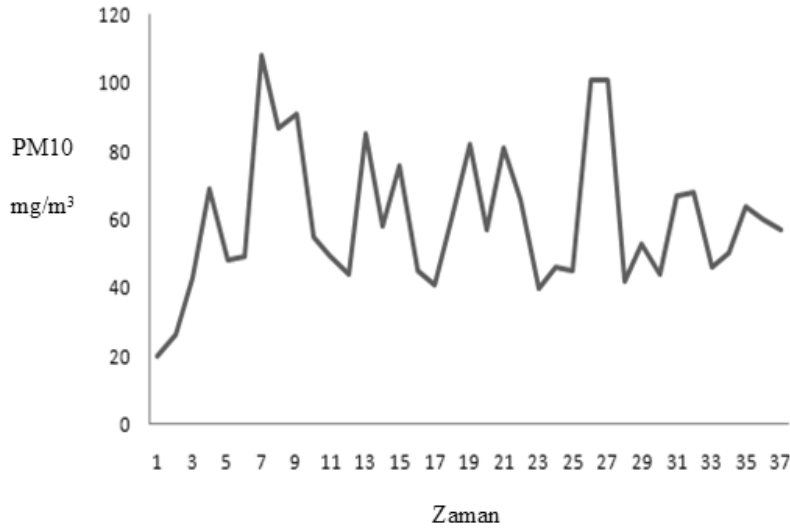
Kısıtlar:

$$x_1 < x < x_2$$

şeklinde verilen optimizasyon problemi çözümlenir. Verilen optimizasyon probleminin çözümü sonucunda,  $f(x)$  fonksiyonunu minimum yapan ve  $(x_1, x_2)$  aralığında yer alan  $x$  değeri bulunur. Verilen minimum problemi, aralık uzunluğunun en iyi değerinin bulunması problemi olarak düşünüldüğünde,  $f(x)$  fonksiyonu,  $x$  değeri için Chen (1996) yöntemiyle bulunan öngörülerden hesaplanan hata kareler ortalaması karekök (HKOK) değerini vermektedir. Buna göre, belirtilen optimizasyon problemi çözümlendiğinde,  $(x_1, x_2)$  aralığında yer alan ve en küçük hata değerini veren aralık uzunluğu bulunmaktadır. Eğrioğlu vd. (2011) belirtilen optimizasyon probleminin çözümünde MATLAB paket programından yararlanmışlardır. Yöntem hakkında detaylı bilgi Eğrioğlu vd. (2011) çalışmasında bulunabilir.

## 3. UYGULAMA

Uygulamada, 2007 Kasım ile 2010 Ağustos ayları arasındaki, aylık gözlemlerden (PM10 maddesi,  $\text{mg/m}^3$ ) oluşan Ankara hava kalitesi zaman serisi çözümlenmiştir. Çözümlenen serinin grafiği Şekil 1'de verilmiştir.



Şekil 1. Ankara hava kalitesi zaman serisi

Şekil 1’de grafiği görülen zaman serisi 6 farklı bulanık zaman serisi yaklaşımıyla çözümlenmiştir. Kullanılan bulanık zaman serisi yaklaşımlarının uygulanış şekilleri aşağıda özetlenmiştir:

Song ve Chissom (1993b) ve Chen (1996) yaklaşımları kullanırken, aralık uzunluğu olarak 5, 10, 15 ve 20 değerleri denenmiştir. İki yaklaşım içinde 10 aralık uzunluğunun en iyi sonucu ürettiği görülmüştür.

Huang’ın (2001) ortalamaya dayalı ve dağılıma dayalı yaklaşımlarıyla en iyi aralık uzunluğu sırasıyla 20 ve 2 olarak bulunmuştur. Huang ve Yu (2006) önerdikleri yaklaşım için  $\alpha$  parametresinin 0.50 değerine yakın alınması gerektiğini belirtmişlerdir. Buna göre, Huang ve Yu (2006) tarafından önerilen yaklaşım kullanılırken  $\alpha$  değeri için 0.40, 0.45, 0.50, 0.55 ve 0.60 değerleri denenmiş ve en iyi sonuçlar 0.45 değeri için elde edilmiştir.

Eğrioğlu vd. (2011) tarafından önerilen yaklaşımda, optimizasyon sürecinde en iyi aralık değerinin arandığı aralığın sınırları, veri yapısına uygun olarak 1 ve 20 olarak alınmıştır. Optimizasyon süreci sonunda, en iyi aralık uzunluğu 1.4057 olarak belirlenmiştir. Yapılan tüm işlemlerde MATLAB paket programı kullanılmıştır.

Elde edilen öngörü sonuçlarını daha iyi değerlendirebilmek amacıyla, HKOK, ortalama mutlak hata yüzdesi (OMHY) ve yön sağlamlığı (YS) olmak üzere üç farklı performans ölçütü kullanılmıştır. Kullanılan altı farklı bulanık zaman serisi yönteminden elde edilen tüm sonuçlar Tablo 1’de özetlenmiştir. Tablo 1’de, Ort. ve Dağ. sırasıyla, Huang (2001) tarafından önerilen ortalamaya ve dağılıma dayalı yaklaşımları ve Oran ise Huang ve Yu (2006) tarafından önerilen orana dayalı yaklaşımı göstermektedir. Tablo 1 incelendiğinde, hesaplanan tüm performans ölçütlerine göre, en iyi öngörüler Eğrioğlu vd. (2011) tarafından önerilen bulanık zaman serisi yaklaşımının verdiği görülmüştür. Bir başka ifadeyle, Eğrioğlu vd. (2011) tarafından önerilen yöntem kullanıldığında, en küçük HKOK ve OMHY değerine sahip ve en büyük YS değerine sahip, en güvenilir öngörüler elde edilmiştir.

**Tablo1. Tüm yöntemlerde test kümesi üzerinden elde edilen performans ölçütleri değerleri**

Ölçütler	Song ve Chissom (1993b)	Chen (1996)	Huarng Ort.(2001)	Huarng ve Yu Dağ(2001)	Oran (2006)	Eğrioğlu vd. (2011)
HKOK	18.168	18.442	22.665	14.164	18.361	13.127
OMHY	0.272	0.283	0.380	0.149	0.274	0.117
YS	0.743	0.743	0.657	0.829	0.743	0.857

#### 4. SONUÇLAR

Belirsizlik içeren zaman serilerinin çözülmesinde klasik zaman serisi yaklaşımları yetersiz kalmaktadır. Son yıllarda, bu tür zaman serilerinin çözümlenebilmesi için bulanık zaman serisi yaklaşımları geliştirilmekte ve kullanılmaktadır. Bulanık zaman serileri yaklaşımlarında, güvenilir öngörüler elde edilmesinde, en önemli kararlardan biri aralık uzunluğunun doğru olarak belirlenebilmesidir. Bu nedenle, literatürde bu problemin çözümü için önerilen çeşitli yaklaşımlar bulunmaktadır. Yapılan çalışmada, yapısı gereği belirsizlik içeren, 2007 Kasım ile 2010 Ağustos ayları arasındaki, aylık gözlemlerden oluşan Ankara hava kalitesi zaman serisi farklı bulanık zaman serisi öngörü modelleri ile çözümlenmiştir. Çözümlemede, literatürde iyi bilinen Song ve Chissom (1993b) ve Chen (1996) tarafından önerilen bulanık zaman serileri yaklaşımları kullanılmıştır. Ek olarak, aralık uzunluğunun belirlenmesinde, Huarng (2001) tarafından önerilen ortalama ve dağılıma dayalı yaklaşımlar ve Huarng ve Yu (2006) tarafından önerilen orana dayalı yaklaşımın uygulandığı yöntemler de kullanılmıştır. Son olarak, Eğrioğlu vd. (2011) tarafından önerilen, aralık uzunluğunun belirlenmesinde kısıtlı optimizasyona dayalı bir yaklaşımın kullanıldığı, bulanık zaman serisi yöntemiyle de çözümleme yapılmıştır. Ankara hava kalitesi verilerinin ilk kez bulanık zaman serisiyle çözümlendiği bu çalışmada, hem literatürde en çok bilinen iki bulanık öngörü modeli hem de farklı aralık uzunluğu belirleme tekniklerinin kullanıldığı dört farklı bulanık öngörü modeli uygulanarak çözümleme yapılmıştır. Tüm yöntemlerden hesaplanan öngörüler, HKOK, OMHY ve YS olmak üzere üç farklı performans ölçütüne göre değerlendirilmiştir. Elde edilen sonuçlara göre, belirsizlik içeren Ankara hava kalitesi zaman serisi için en güvenilir öngörülerin, kullanılan üç performans ölçütüne göre, Eğrioğlu vd. (2011) tarafından önerilen yöntemden elde edildiği açıkça görülmüştür.

#### 5. KAYNAKLAR

Aladağ, C. H., Başaran, M. A., Eğrioğlu, E., Yolcu, U., Uslu, V. R., 2009. Forecasting in High Order Fuzzy Times Series by Using Neural Networks to Define Fuzzy Relations. *Expert Systems with Applications*, 36 (3), 4228-4231.

Aladağ, C. H., Yolcu, U., Eğrioğlu, E., 2010. A High Order Fuzzy Time Series Forecasting Model Based on Adaptive Expectation and Artificial Neural Networks. *Mathematics and Computers in Simulation*, 81, 875-882.

Chen, S. M., 1996. Forecasting Enrollments Based on Fuzzy Time-series. *Fuzzy Sets and Systems*, 81, 311-319.

Eğrioglu, E., Aladağ, C. H., Uslu, V. R., Başaran, M. A., Yolcu, U., 2009. A New Hybrid Approach Based on SARIMA and Partial High Order Bivariate Fuzzy Time Series Forecasting Model. *Expert Systems with Applications*, 36 (4), 7424-7434.

Eğrioglu, E., Aladağ, C. H., Başaran, M. A., Yolcu, U., Uslu, V. R., 2011. A New Approach Based on the Optimization of the Length of Intervals in Fuzzy Time Series. *Journal of Intelligent and Fuzzy Systems*, 22, 15-19.

Huang, K., 2001. Effective Length of Intervals to Improve Forecasting in Fuzzy Time-Series. *Fuzzy Sets and Systems*, 123, 387-394.

Huang, K., Yu, H. K., 2006. The Application of Neural Networks to Forecast Fuzzy Time Series. *Physica A*, 363, 481-491.

Song, Q., Chissom, B. S., 1993a. Fuzzy Time Series and its Models. *Fuzzy Sets and Systems*, 54, 269-277.

Song, Q., Chissom, B. S., 1993b. Forecasting Enrollments with Fuzzy Time Series - Part I. *Fuzzy Sets and Systems*, 54, 1-10.

Song, Q., Chissom, B. S., 1994. Forecasting Enrollments with Fuzzy Time Series - Part II. *Fuzzy Sets and Systems*, 62 (1), 1-8.

Zadeh, L. A., 1965. Fuzzy Sets. *Inform and Control*, 8, 338-353.

## **ANALYSIS OF ANKARA AIR QUALITY DATA WITH FUZZY TIME SERIES**

### *ABSTRACT*

*Fuzzy time series have been successfully used in recent years to analyze time series which include uncertainty such as air temperature and stock market data. In the literature, fuzzy time series approaches have been studied in order to reach more accurate forecasts. Determining the length of interval correctly is one of the crucial points in obtaining accurate forecasts. Some methods have been proposed in the literature to solve this problem. Air quality data also include uncertainty due to its nature, and such time series should be analyzed by using fuzzy time series. In this study, Ankara air quality time series are analyzed by using the fuzzy time series method for the first time. Six different fuzzy time series approaches, which employ different techniques for determining the length of interval, are used in the implementation, and the results are compared and discussed.*

**Keywords:** Ankara air quality, Length of interval, Fuzzy time series, Optimization, Forecasting.

## BULANIK YEREL POLİNOMİYAL PARAMETRİK OLMAYAN REGRESYON MODELLERİ

Memmedağa MEMMEDLİ\* Münevvere YILDIZ\*\* Özer ÖZDEMİR\*\*\*

### ÖZET

*Bu çalışmada parametrik olmayan bulanık regresyon modelleri göz önüne alınmıştır. Mevcut bulanık regresyon modelleri, bulanık versiyonu geliştirilen parametrik olmayan yerel kübik regresyon modeli ile karşılaştırılmıştır. Bu yeni yaklaşımda, önerilen modelin performansının diğerlerine göre daha iyi sonuçlara sahip olduğu bulunmuştur.*

**Anahtar Kelimeler:** Bulanık parametrik olmayan regresyon, LR bulanık sayısı, Yerel polinomiyal düzeltme.

### 1. GİRİŞ

Kesin durumlara ait bilginin yer aldığı çalışmalarda klasik regresyon modelleri yaygın olarak çeşitli alanlarda kullanılmaktadır. Ancak bazı veri gruplarını bu modellerle ifade etmek mümkün olamamaktadır. Bununla birlikte, bazen de bulanık yapıya sahip verilerin kullanılmasının gerekliliği, bu klasik regresyon modellerinin kullanımını olanaksız hale getirmektedir. Bu durumda, belirsizliğin olduğu ve klasik regresyon modellerinin yetersiz kaldığı durumlarda yeni arayışlar başlamış ve bulanık regresyon modelleri ileri sürülmüştür.

Bulanık regresyon modelleri konusunda ilk çalışmalar Tanaka vd. (1982) ile başlamıştır ve bu çalışmaların çoğu bulanık parametrik regresyon modelleri üzerine odaklanmıştır. Bulanık regresyon modelleri için farklı yaklaşımlar ileri sürülerek; matematiksel programlama temelli yöntemler (Tanaka vd., 1982; Nasrabadi ve Nasrabadi, 2004), en küçük kareler yöntemi (Diamond, 1988; Hong vd., 2004) ve sinir ağı yaklaşımları (Cheng ve Lee, 2001) gibi çeşitli yöntemler kullanılmıştır.

Cheng ve Lee (2001) bulanık regresyon analizinde, radyal tabanlı fonksiyonları kullanan sinir ağları üzerine çalışma yapmışlardır. Bulanık parametrik regresyon analizi konusunda çeşitli pratik çalışmalar da literatürde yer almaktadır (Hong vd., 2004; Kao ve Chyu, 2002; Sanchez ve Gomez, 2003). Pratik çalışmalarda yanıt değişkeni ile açıklayıcı değişkenler arasındaki ilişkinin belirli bir parametrik modelle ifade edilmesi çok kısıtlayıcı olmaktadır. Ayrıca yanlış sonuçlara da neden olabilmektedir. Bu açıdan, son yıllarda parametrik olmayan regresyon modellerine olan ilgi giderek artmıştır (Fan ve Gijbels, 1996; Loadar, 1999).

Ishibuchi ve Tanaka (1992), geri yayılım sinir ağı algoritmasını kullanarak, çeşitli parametrik olmayan bulanık regresyon modellerini önermişlerdir. Parametrik olmayan regresyon modellerine ilginin giderek artmasıyla bu modellerin bulanık versiyonları da geliştirilmeye başlanmıştır (Fan ve Gijbels, 1996; Wood, 2006). Cheng ve Lee (1999) k-

\*Prof. Dr., Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [mmammadov@anadolu.edu.tr](mailto:mmammadov@anadolu.edu.tr)

\*\*Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [munevvere@hotmail.com](mailto:munevvere@hotmail.com)

\*\*\*Öğr. Gör., Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [ozerozdemir@anadolu.edu.tr](mailto:ozerozdemir@anadolu.edu.tr)



en yakın komşu ve çekirdek tahminleme yöntemlerinin bulanık versiyonlarını incelemişlerdir. Wang vd. (2007) bulanık yerel doğrusal regresyon modelini ele almışlar ve deneysel çalışmalar ile bu yaklaşımın (Cheng ve Lee, 1999)'deki bulanık modellere göre avantajlarını göstermişlerdir. Bu çalışmalarda düzeltme parametresinin seçimi için çapraz-geçerlilik kriteri kullanılmıştır. Ancak, Fan ve Gijbels (1996) yerel regresyon uyumlaması (fitting) için daha etkin seçim yöntemleri önermişlerdir.

Klasik teoride parametrik olmayan regresyon modelleri üzerine yapılmış pek çok çalışmanın henüz bulanık versiyonları geliştirilmemiştir, bu açıdan parametrik olmayan regresyon modelleri teorisinde kullanılan yöntemlerin bulanık versiyonlarının geliştirilmesi önemli bir konudur.

Çalışmada parametrik olmayan bulanık yerel polinomial regresyon modelleri, bulanık yerel doğrusal regresyon modeli ve literatürde bulanık versiyonu üzerine çalışma yapılmamış olan bulanık yerel kübik regresyon modelleri incelenmiştir. Gerçekleştirilen simülasyon çalışması ile farklı çekirdek fonksiyonları için uygun düzeltme parametresi belirlenerek, bu modellerin performansları ASE (ortalama hata kareler) performans kriteri yardımıyla değerlendirilmiştir. Bulanık yerel kübik regresyon modelinin, bulanık yerel doğrusal regresyon modeline göre daha iyi performans gösterdiği gözlemlenmiştir.

## 2. BULANIK PARAMETRİK OLMAYAN REGRESYON MODELİ İÇİN YEREL POLİNOMİYAL YÖNTEM

Tek değişkenli bulanık parametrik olmayan regresyon modeli eşitlik 1'deki gibi ele alınır:

$$y = F(x) \{+\} \varepsilon \quad (1)$$

Burada,

$$F(x) = (m(x), \alpha(x), \beta(x)) \quad (2)$$

kesin  $\mathcal{D}$  kümesinde tanımlanmış  $LR$  bulanık değerler alan bulanık fonksiyon,  $\varepsilon$  bulanık hata terimi,  $\{+\}$  - kullanılan bulanık sıralama yöntemine bağlı olarak tanımlanan operatör'dür (Cheng ve Lee, 1999; Wang vd., 2007).

$(x_i, y_i)$ 'lerin ( $i = 1, 2, \dots, n$ ), kesin girdilere ve  $LR$  bulanık çıktılara sahip eşitlik 1 modeli için gözlem verileri oldukları varsayımı altında, her bir  $y$  çıktısı,  $c_y$ - merkez,  $\alpha_y$  ve  $\beta_y$  sol ve sağ yayılım bileşenleri ile  $(c_y, \alpha_y, \beta_y)$  şeklinde bir  $LR$  bulanık sayıdır. Bu sayı  $l_y = c_y - \alpha_y$  ve  $u_y = c_y + \beta_y$  alınarak, sol uç, merkez ve sağ uç noktaları yardımıyla  $(l_y, c_y, u_y)$  şeklinde yazılabilir.  $l(x), c(x), u(x)$  fonksiyonlarının ise sırasıyla  $LR$  - fonksiyonun sol ucu, merkezi ve sağ ucu olarak,  $\mathcal{D}$  tanım kümesinde  $p$ . mertebeye kadar sürekli türeve sahip olmasıyla, bu fonksiyonlar verilen bir  $x_0 \in \mathcal{D}$  noktasının komşuluğunda yaklaşık olarak  $p$ . mertebeden Taylor polinomu olarak ele alınırlar ve aşağıdaki gibi yazılırlar;

$$l(x) \approx l(x_0) + l'(x_0)(x - x_0) + \dots + \frac{l^{(p)}(x_0)}{p!} (x - x_0)^p \quad (3)$$

$$c(x) \approx c(x_0) + c'(x_0)(x - x_0) + \dots + \frac{c^{(p)}(x_0)}{p!} (x - x_0)^p \quad (4)$$

$$u(x) \approx u(x_0) + u'(x_0)(x - x_0) + \dots + \frac{u^{(p)}(x_0)}{p!}(x - x_0)^p \quad (5)$$

Diamond uzaklığı (Diamond, 1988) kullanılarak, belirli bir  $K(\cdot)$  çekirdek fonksiyonu yardımıyla ağırlıklı yerel en küçük kareler yöntemi yardımıyla eşitlik 3-4-5'teki yerel tahminler elde edilir:

$$(\hat{l}(x_0), \hat{l}'(x_0), \dots, \hat{l}^{(p)}(x_0))^T = \left( X^T(x_0)W(x_0; h)X(x_0) \right)^{-1} X^T(x_0)W(x_0; h)l_y \quad (6)$$

$$(\hat{c}(x_0), \hat{c}'(x_0), \dots, \hat{c}^{(p)}(x_0))^T = \left( X^T(x_0)W(x_0; h)X(x_0) \right)^{-1} X^T(x_0)W(x_0; h)c_y \quad (7)$$

$$(\hat{u}(x_0), \hat{u}'(x_0), \dots, \hat{u}^{(p)}(x_0))^T = \left( X^T(x_0)W(x_0; h)X(x_0) \right)^{-1} X^T(x_0)W(x_0; h)u_y \quad (8)$$

Burada;

$$X(x_0) = \begin{pmatrix} \mathbf{1} & (x_1 - x_0) & (x_1 - x_0)^2 & \dots & (x_1 - x_0)^p \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{1} & (x_n - x_0) & (x_n - x_0)^2 & \dots & (x_n - x_0)^p \end{pmatrix}, \quad (9)$$

$$l_y = \begin{pmatrix} l_{y_1} \\ \vdots \\ l_{y_n} \end{pmatrix}, c_y = \begin{pmatrix} c_{y_1} \\ \vdots \\ c_{y_n} \end{pmatrix}, u_y = \begin{pmatrix} u_{y_1} \\ \vdots \\ u_{y_n} \end{pmatrix} \quad (10)$$

$$W(x_0; h) = \text{Diag}(K_h(|x_1 - x_0|), \dots, K_h(|x_n - x_0|)) \quad (11)$$

$$K_h(z) = \frac{1}{h} K\left(\frac{z}{h}\right) \quad (12)$$

olmak üzere  $n \times n$  boyutunda köşegen ağırlık matrisidir. Gauss çekirdek fonksiyonu  $K(\cdot)$  aşağıdaki gibi yazılır;

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (13)$$

Epanechnikov çekirdek fonksiyonu kullanılması durumunda,

$$K(x) = \begin{cases} 0.75(1 - x^2), & |x| < 1 \\ \mathbf{0} & , \quad d.d \end{cases} \quad (14)$$

şeklinde ifade edilir.

Çalışmada eşitlik 1'deki model için, polinomun mertebeleri  $p = 1$  ve  $p = 3$  alınarak, eşitlik 3-4-5'teki uygun formüller yardımıyla bulanık yerel doğrusal ve bulanık yerel kübik regresyon modelleri incelenmiştir. Bu amaçla bir sonraki bölümde yer alan simülasyon çalışması yapılmıştır.

### 3. SİMÜLASYON ÇALIŞMASI

Simülasyon çalışması; bulanık yerel doğrusal ve bulanık yerel kübik modelleri için uygun düzeltme parametresinin seçiminde çapraz geçerlilik (CV) kriterini kullanarak, bu modellerin performanslarını karşılaştırma amacıyla gerçekleştirilmiştir. Cheng ve Lee (1999) çalışmasındaki fonksiyonlar kullanılarak, boyutları 100 olan veri setleri türetilmiştir ve simülasyon denemelerinde 100 tekrarlıma gerçekleştirilmiştir. Bulanık yanıt çıktıları, merkez ve sağ-sol saçılımları Örnek-1 ve Örnek-2'deki gibi türetilen, simetrik üçgen bulanık sayılardır. Simülasyon çalışmasında gürültü, merkez için  $[-0.5, 0.5]$ , simetrik saçılımlar için ise  $[-0.25, 0.25]$  aralığında düzgün dağılımdan elde edilmektedir.

**Örnek-1.**  $g_1(x)$ ,  $[0,10]$  aralığında aşağıdaki gibi tanımlanmış bir fonksiyon olsun:

$$g_1(x) = \frac{1}{5}x^2 + 2\exp\left(\frac{x}{10}\right) \quad (15)$$

$[0,10]$  aralığındaki  $x_i = 0.1i$  ( $i = 1,2, \dots, 100$ ) noktaları için,

$$y_i = g_1(x_i) + \mathit{rand}[-0.5, 0.5] \quad (16)$$

$$\sigma_i = \frac{1}{4}g_1(x_i) + \mathit{rand}[-0.25, 0.25] \quad (17)$$

olarak hesaplanırlar. Burada  $\mathit{rand}[a_1, a_2]$ , her bir  $i$  için  $[a_1, a_2]$  aralığı üzerinde düzgün dağılımdan bağımsız olarak türetilen bir rassal sayıyı göstermektedir. Gözlenen bulanık çıktıların simetrik üçgen bulanık sayılar olduğu varsayılarak, bu çıktılara ait gösterim aşağıdaki gibi tanımlanır.

$$Y_i = (l_{y_i}, c_{y_i}, u_{y_i})_T = (y_i - \sigma_i, y_i, y_i + \sigma_i)_T \quad i = 1, 2, \dots, 100 \quad (18)$$

**Örnek-2.**  $[0,10]$  aralığında ikinci fonksiyon aşağıdaki gibi tanımlanmıştır:

$$g_2(x_i) = 10 + 5\sin(0.025\pi(1 - x)^2) \quad (19)$$

$[0,10]$  aralığındaki aynı  $x_i = 0.1i$  ( $i = 1,2, \dots, 100$ ) noktaları için

$$y_i = g_2(x_i) + \mathit{rand}[-0.5, 0.5] \quad (20)$$

$$\sigma_i = \frac{1}{3}g_2(x_i) + \mathit{rand}[-0.25, 0.25] \quad (21)$$

olarak hesaplanırlar. Bulanık çıktılar;

$$Y_i = (l_{y_i}, c_{y_i}, u_{y_i})_T = (y_i - \sigma_i, y_i, y_i + \sigma_i)_T \quad i = 1, 2, \dots, 100 \quad (22)$$

olarak belirlenirler. Çalışmada bu fonksiyonların yardımıyla türetilen iki veri seti için bulanık yerel doğrusal ve bulanık yerel kübik regresyon modelleri incelenmiştir. Her iki modele uygun düzeltme parametresinin (h) optimum değerinin seçimi için çapraz geçerlilik (CV) kriteri kullanılmıştır.

Ele alınan bulanık regresyon problemleri için CV kriteri;

$$\begin{aligned}
 CV(\mathbf{h}) &= \frac{1}{n} \sum_{i=1}^n d^2(y_i, \hat{F}(x_i, \mathbf{h})) \\
 &= \frac{1}{n} \sum_{i=1}^n [(l_{y_i} - \hat{l}_{(i)}(x_i, \mathbf{h}))^2 + (c_{y_i} - \hat{c}_{(i)}(x_i, \mathbf{h}))^2 + (u_{y_i} - \hat{u}_{(i)}(x_i, \mathbf{h}))^2] \quad (23)
 \end{aligned}$$

formülüyle hesaplanır. Burada  $(i)$  indisi tahminleme yapılırken  $i$ . gözlem değerinin çıkarıldığını göstermektedir. Optimum  $h_0$  düzeltme parametresi,  $CV(h_0) = \min_h CV(h)$  minimum probleminin çözümü olarak elde edilir.

Simülasyon hesaplamalarında rassallığın etkisini azaltmak için 100 tekrarlamaya yapılmıştır. Her iki veri seti ile yapılan uyumlamalarda ağırlık dizilerinin oluşturulmasında Gauss ve Epanechnikov çekirdekleri (kernel) kullanılmıştır. Modellerin performansları onların ASE değerleri yardımıyla karşılaştırılmıştır:

$$\begin{aligned}
 ASE(\mathbf{h}) &= \frac{1}{n} \sum_{i=1}^n d^2(F(x_i), \hat{F}(x_i)) \\
 &= \frac{1}{n} \sum_{i=1}^n [(l(x_i) - \hat{l}(x_i))^2 + (c(x_i) - \hat{c}(x_i))^2 + (u(x_i) - \hat{u}(x_i))^2] \quad (24)
 \end{aligned}$$

Tablo 1’de Örnek-1 için bulanık yerel doğrusal (LLS) ve kübik regresyon modellerinin (LCS), CV kriteri yardımıyla optimum düzeltme parametreleri seçilerek, Gauss ve Epanechnikov çekirdeklerine uygun ASE değerleri verilmiştir.

**Tablo 1. Örnek-1 için simülasyon sonuçları**

Yöntem	Çekirdek	Düzeltilme Parametresi (h)	ASE
LLS	Gauss	0,51	0,2642107
	Epanechnikov	1,20	0,2739353
LCS	Gauss	1,10	0,2616963
	Epanechnikov	1,25	0,2496074

Benzer şekilde Tablo 2’de Örnek-2 için bulanık yerel doğrusal ve kübik regresyon modellerinin, CV kriteri yardımıyla optimum düzeltme parametreleri seçilerek, Gauss ve Epanechnikov çekirdeklerine uygun ASE değerleri verilmiştir.

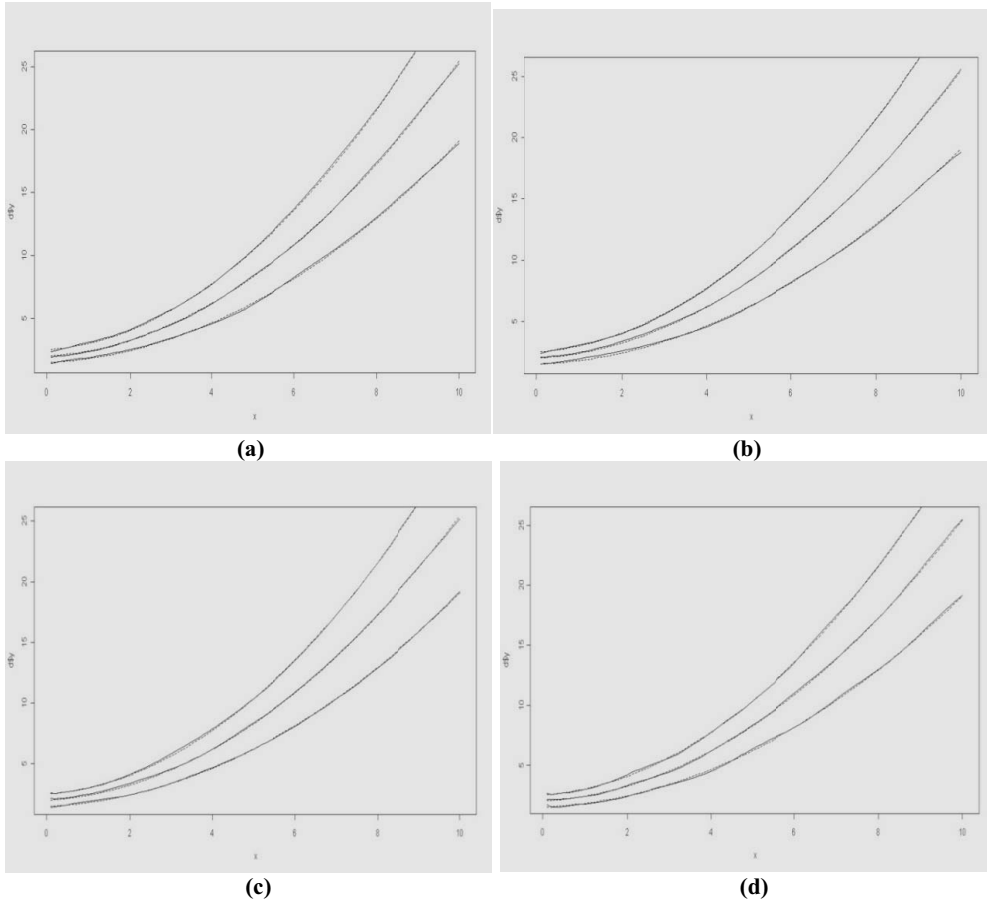
**Tablo 2. Örnek-2 için simülasyon sonuçları**

Yöntem	Çekirdek	Düzeltilme Parametresi (h)	ASE
LLS	Gauss	0,21	0,2301000
	Epanechnikov	0,52	0,2573107
LCS	Gauss	0,40	0,2281975
	Epanechnikov	1,30	0,2483770

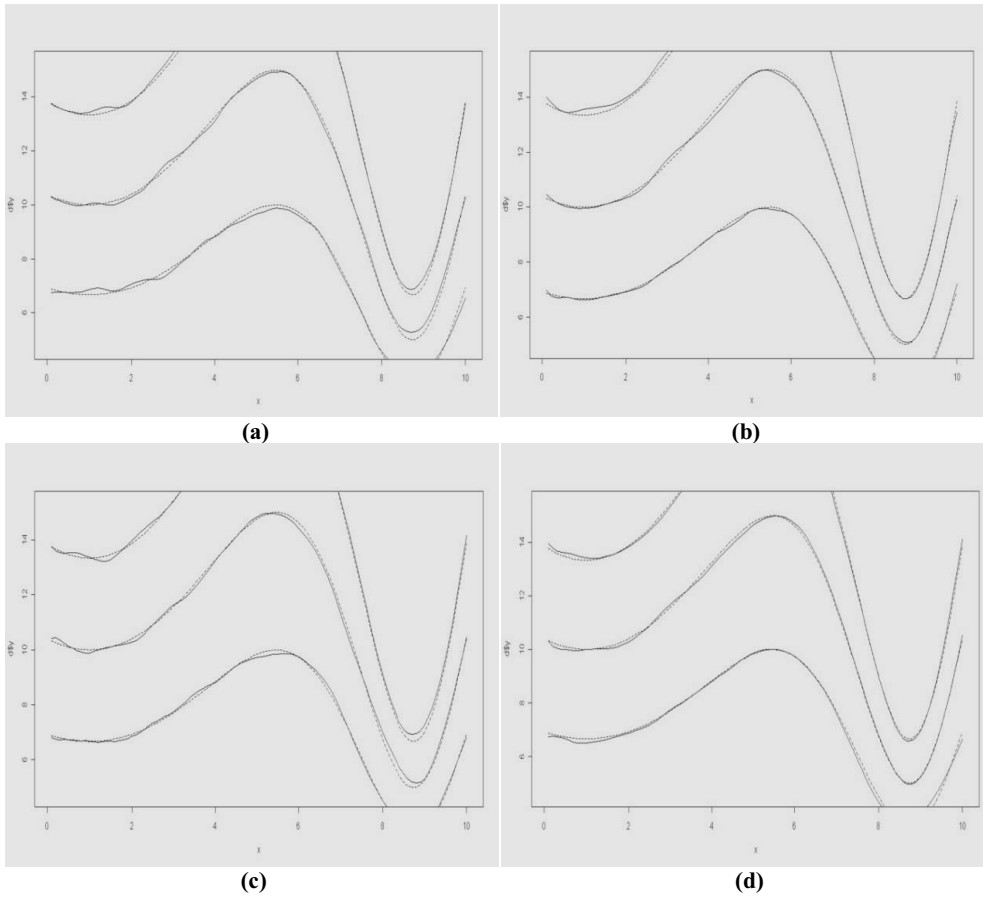
Tablo 1 ve Tablo 2’de LLS ve LCS modelleri için uygun çekirdek fonksiyonu ve düzeltme parametresine karşılık gelen performans kriteri olan ASE değerleri görülmektedir. Tablolar incelendiğinde bulanık yerel kübik modelin performansı

çekirdek fonksiyonları açısından daha iyi sonuçlar vermektedir. Tablolarda yer alan bilgilerin grafiksel gösterimleri Şekil 1 ve Şekil 2’de verilmiştir.

Şekil 1 ve Şekil 2’de yer alan grafikler incelendiğinde, bulanık yerel kübik regresyon modeline ait grafiklerde, merkez, sağ – sol saçılım eğrilerinin gerçek bulanık regresyon fonksiyonu eğrisi ve onun tahminlerini gösteren eğri arasındaki sapmanın, bulanık yerel doğrusal regresyon modeline göre daha az olduğu görülmektedir. Şekil 1 ve Şekil 2’de, Örnek-1 ve Örnek-2 için ilgili grafikler verilmiştir.



Şekil 1. Örnek-1 için Gauss çekirdeği kullanıldığında: (a)-  $h=0.51$  ve  $p=1$ , (b)-  $h=1.1$  ve  $p=3$   
Epanechnikov çekirdeği kullanıldığında: (c)-  $h=1.20$  ve  $p=1$ , (d)-  $h=1.25$  ve  $p=3$



Şekil 2. Örnek-2 için Gauss çekirdeği kullanıldığında: (a)-  $h=0.21$  ve  $p=1$ , (b)-  $h=0.4$  ve  $p=3$   
Epanechnikov çekirdeği kullanıldığında: (c)-  $h=0.52$  ve  $p=1$ , (d)-  $h=1.3$  ve  $p=3$

#### 4. TARTIŞMA VE SONUÇ

Parametrik olmayan yerel polinomial regresyon problemlerinde yerel polinomun mertebesi ile ilişkili olarak düzeltme parametresinin ( $h$ ) seçimi oldukça önemli bir rol oynar. Bu modellerde polinomun mertebesi arttıkça yerel düzeltme parametresinin genişliği de artar.

Bulanık yerel polinomial regresyon modelleri incelenirken, genel teknik olarak ağırlıklı en küçük kareler yöntemi, bulanık uzaklık olarak ise Diamond uzaklığı kullanılmıştır. Yapılan simülasyon çalışması ile öncelikle bulanık yerel doğrusal ve bulanık yerel kübik regresyon modelleri için çapraz geçerlilik kriteri yardımıyla uygun düzeltme parametresi seçilmiştir. Gauss ve Epanechnikov çekirdek fonksiyonları modellerin uyumlanmasında kullanılmıştır. Tüm bu kriterler göz önüne alınarak bulanık yerel doğrusal ve bulanık yerel kübik regresyon modelleri karşılaştırılmıştır. Bulanık yerel kübik regresyonun, bulanık yerel doğrusal regresyon modeline göre her iki çekirdek fonksiyonu içinde  $ASE$  değerleri bakımından daha iyi performans gösterdiği gözlenmiştir.

Sonuçta, polinomun mertebesi artarken düzeltme parametresinin genişliği de artmaktadır. Bu durum, işlemsel olarak yerel uyumlama noktalarının azalmasından dolayı işlem basamaklarını azaltmakta ve zamandan kazanç sağlanmaktadır. Özellikle daha eğrisel olan veri yapılarında yerel doğrusal modellerin kullanılması çok uygun olmamaktadır. Çünkü bu tür durumlarda düzeltme parametresinin değeri oldukça küçülmekte ve işlem karmaşasına yol açmaktadır.

Sonraki çalışmalarda düzeltme parametresinin seçimi için farklı kriterlerin bulanık versiyonlarının belirlenmesi, modelin bu parametreye bağlı olarak esnekliğinin incelenmesi merak uyandırmaktadır. Ayrıca, diğer parametrik olmayan regresyon modellerinin bulanık versiyonları üzerine çalışmaların yapılması oldukça önemlidir.

## 5. KAYNAKLAR

Cheng, C. B., Lee E. S., 1999. Nonparametric Fuzzy Regression – k-NN and Kernel Smoothing Techniques, *Computers and Mathematical with Applications*, 38, 239-251.

Cheng, C. B., Lee E. S., 2001. Fuzzy Regression with Radial Basis Function Networks, *Fuzzy Sets and Systems*, 119, 291-301.

Diamond, P., 1988. Fuzzy Least Squares, *Information Sciences* 46, 141-157.

Fan, J., Gijbels, I., 1996. *Local Polynomial Modelling and its Applications*, Chapman & Hall/CRC.

Hong, H., Song, J. K., Do, H., 2001. Fuzzy Least Squares Linear Regression Analysis Using Shape Preserving Operations, *Information Sciences* 138, 185-193.

Hong, Y. Y., Chao, Z. T., Yang, M. S. (2004). A Fuzzy Multiple Linear Regression Based Loss Formula in Electric Distribution Systems, *Fuzzy Sets and Systems* 142, 293-306.

Ishibuchi, H., Tanaka, H. (1992). Fuzzy Regression Analysis Using Neural Networks, *Fuzzy Sets and Systems* 50, 257-265.

Kao, C., Chyu, C. L., 2002. A Fuzzy Linear Regression Model with Better Explanatory Power, *Fuzzy Sets and Systems* 126, 401-409.

Loader, C., 1999. *Local Regression and Likelihood*, Springer.

Nasrabadi, M., Nasrabadi, E., 2004. A Mathematical-Programming Approach to Fuzzy Linear Regression Analysis, *Applied Mathem. and Computation* 155, 673-688.

Sanchez, J. D., Gomez, A. T., 2003. Applications of Fuzzy Regression in Actuarial Analysis, *Journal of Risk Insurance* 70, 665-699.

Tanaka, H., Uejima, S., Asai, K., 1982. Linear Regression Analysis with Fuzzy Model, *IEEE Transactions on Systems, Man. and Cybernetics* 12, 903-907.

Wang, N., Zhang, W. X., Mei, C. L., 2007. Fuzzy Nonparametric Regression Based on Local Linear Smoothing Technique, An International Journal Information Sciences, 177, 3882-3900.

Wood, S. N., 2006. Generalized Additive Models (An Introduction with R), Chapman & Hall/CRC.

## FUZZY LOCAL POLYNOMIAL NONPARAMETRIC REGRESSION MODELS

### ABSTRACT

*In this study, nonparametric fuzzy regression models are considered. Existing fuzzy regression models are compared with improved fuzzy version of local nonparametric cubic regression model. In this new approach, it is found that performance of this model has better results according to others.*

**Keywords:** Fuzzy nonparametric regression, LR Fuzzy number, Local polynomial smoothing.



## SAYISAL GÖRÜNTÜLERDEKİ NESNELERİN BASKIN NOKTALARININ ALGILANMASINDA İSTATİSTİKSEL YAKLAŞIMLAR

Orhan KESEMEN\*

Çiğdem GÜNGÖR\*\*

### ÖZET

*Bu çalışmada, sayısal görüntülerdeki nesnelerin baskın noktalarının araştırılmasında dört ayrı yöntem önerilmiş ve bu yöntemlerin başarımları incelenmiştir. Bu önerilen yöntemleri gerçekleştirmek için bazı istatistiksel ve geometrik kuralları kullanılmıştır.*

**Anahtar Kelimeler:** Baskın nokta algılama, Çokgensel yaklaşım, Köşe algılama, Sayısal doğrular, Şekil çözümlenme.

### 1. GİRİŞ

Geleceğin bilimleri arasında gösterilen yapay zekânın, dünyayı gören gözü olan görüntü işlemenin kullanımı günümüzde birçok alanda giderek yaygınlaşmaktadır. Görüntü çözümlenmede en önemli konulardan birisi sayısal ortama aktarılan görüntülerin içeriklerinin karakteristik özelliklerini ortaya çıkarmaktır. Aynı şekilde bir görüntü içerisindeki nesnelerin de karakteristik özellikleri çıkarılabilir.

Görüntü içerisinde birbirine bitişik olan ve zemin renginden farklı, aynı renkteki gözeklerin (pixel) oluşturduğu kümelere nesne veya tanecik denir. Görüntü içerisinde bulunan değişik biçimlerde ve boyutlardaki nesnelerin özelliklerinin araştırılması uygulama alanlarında oldukça önemli bir yer tutmaktadır. Bu uygulama alanları: tıpta birçok hastalığın teşhisi; askeri alanlarda uydu görüntüleriyle istihbaratların derlenmesi, akıllı füzelerin hedeflerini bulması; bitki çeşitliliğinin araştırılması, hayvan kitlelerinin göçlerinde ve nüfuslarının incelenmesi; kayaçların mineral içeriklerinin araştırılması, vb. konular sayılabilir (Kesemen vd., 2010).

İstatistiksel özellikler yardımıyla basit şekilli nesnelerin birçok tanımı yapılabilir. Bu istatistiksel özellikler karmaşık nesnelerin tanımlanması için de kullanılabilir (Bribiesca vd., 1980). Bir nesnenin en belirgin özelliği, nesnenin kenarlarındaki ani değişimin olduğu bölgelerin uçlarında bulunan merkezi noktalar. Bu noktalar nesnenin baskın noktaları olarak tanımlanır (O’Gorman vd., 2009). Sayısal görüntülerdeki nesneler, değişik geometrik şekillere sahip olabilmektedirler. Bu nesnelerin baskın noktaları yardımıyla çokgenlere benzetilip kenar sayılarının ve dolayısıyla köşe noktalarının belirlenmesi ve nesnelerin daha az veriyle temsil edilmesi uygulamada önemli problemlere çözüm olabilmektedir. Karmaşık biçimlerdeki nesnelerin daha basit bir çokgenle temsil edilmesi görüntü tanıma yöntemlerinde kolaylık sağlayacaktır.

Literatürde, sayısal görüntülerdeki nesnelerin baskın noktalarının araştırılmasında, sınır gözekleri boyunca tanımlanan iki yöney arasındaki açının 180 dereceden küçük bir oranda ( $180^\circ \pm s^\circ$ ) farklı olmasına göre çözüm bulunmaktadır (The vd., 1989), (Marji vd., 2003), (Costa vd., 2001). Ancak bu çalışmalarda kenarları düzgün olmayan

\*Yrd. Doç. Dr., Karadeniz Teknik Üniversitesi, İstatistik ve Bilgisayar Bilimleri Bölümü, Trabzon, e-posta: [okescemen@email.com](mailto:okescemen@email.com)

\*\*Arş. Gör., Karadeniz Teknik Üniversitesi, İstatistik ve Bilgisayar Bilimleri Bölümü, Trabzon, e-posta: [ciğdemgungor@windowslive.com](mailto:ciğdemgungor@windowslive.com)

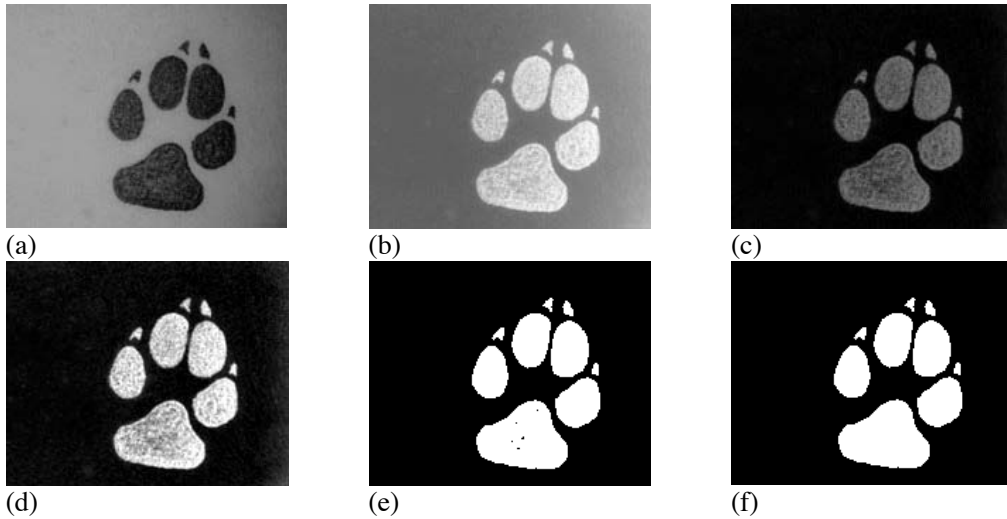
nesnelerin baskın noktalarının araştırılmasında aykırı değerler (doğrudan sapmış) varlığı yanlış sonuçların doğmasına neden olmaktadır.

## 2. YÖNTEMLER

Ele alınan sayısal görüntüler, öncelikle boz ölçekli bir görüntüye dönüştürülür (Şekil 1(a)). Görüntüde artalan (background) içerisinde ilgilenilen nesnelere artalan renginden daha koyu ise ters renk görüntüsü alınarak (Şekil 1(b)) nesnelerin açık renge dönüştürülmesi sonraki işlemlerde kolaylık sağlayacaktır. Kayan ortalama süzgeci yardımıyla ters renkli görüntünün artalan görüntüsü elde edilir ve bu görüntü ters renkli görüntüden çıkarılarak artalan sıfırlanmış görüntü elde edilir (Şekil 1(c)). Elde edilen son görüntü değerleri artalanın çıkarılmasından dolayı dar bir sıklık aralığı sıkıştığı için görüntü ayrırlılığında (contrast) düşme meydana gelmektedir. Aşağıdaki özayrımılık (autocontrast) eşitliği kullanılarak,

$$\hat{g}(x, y) = \frac{g(x, y) - \text{düşük}}{\text{yüksek} - \text{düşük}} \quad (1)$$

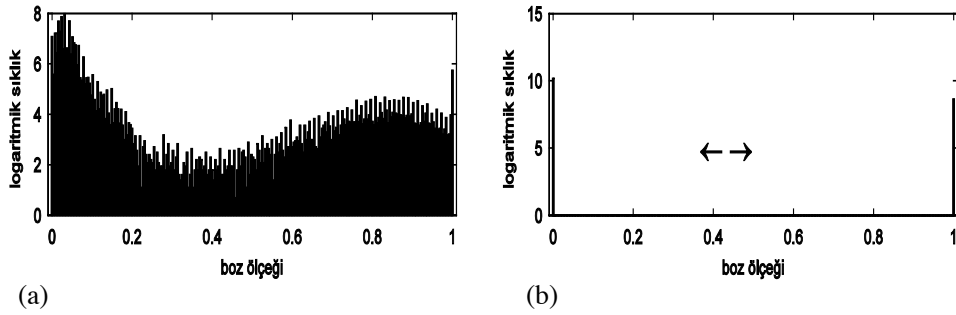
görüntü ayrırlılığ artırılır (Şekil 1(d)). Burada  $g(x, y)$ ,  $x, y$  noktasındaki gözet değeri ( $0 \leq g(x, y) \leq 255; g(x, y) \in \mathbb{N}^2$ ), düşük ve yüksek değişkenleri sayısal görüntüdeki sırasıyla en düşük ve en yüksek gözet değerleridir. Burada  $\hat{g}(x, y)$ , işlenmiş görüntü gözetlerini göstermektedir. Bu yöntem bir anlamda bir birimleştirme (normalization) yöntemi olarak da kullanılır (González vd., 2007).



Şekil 1. Sayısal görüntüler

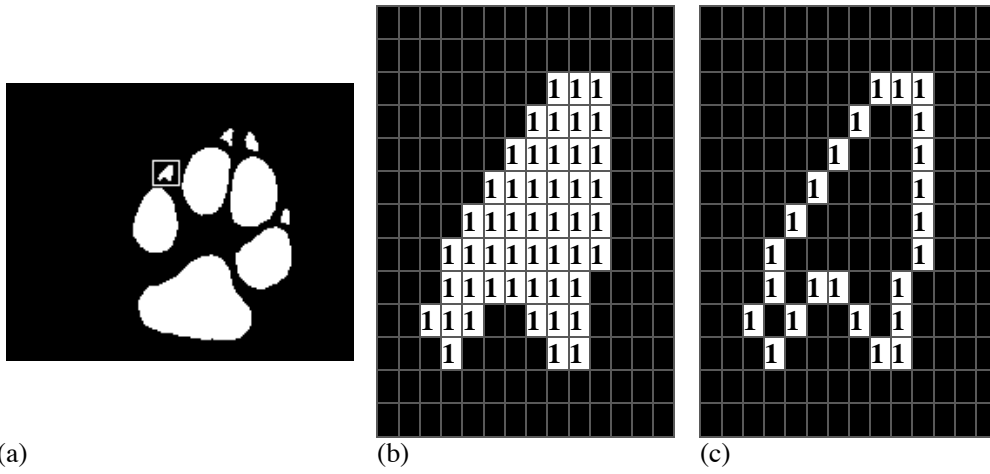
(a) boz ölçekli özgün görüntü; (b) ters renkli görüntü; (c) artalan çıkarılmış görüntü; (d) ayrırlılığ artırılmış görüntü; (e) eşiklenmiş görüntü; (f) eşiklenme sonucunda nesnelerin içinde kalan siyah lekeler büyüklükte (maximization) süzgeci yardımıyla temizlenmiş görüntü

Şekil 1(d)'deki görüntünün iki ayrı veri kümesinin (artalan-nesnelere) birleşimi olduğu Şekil 2'deki sıklık göstergesinde açıkça görülmektedir. Bu iki kümenin kesişim noktası ( $\sim 0.4$ ) bulunarak, bu noktadan küçük olanlar siyah  $\{0\}$ , büyük olanlar beyaza  $\{1\}$  eşiklenirse görüntü Şekil 1(e)'ye dönüşür. Bu görüntü iki farklı  $\{0,1\}$  değer kümesinden oluşmaktadır.



(a) (b)  
Şekil 2. Şekil 1(d) görüntüsünün, (a) logaritmik sıklık göstergesi (b) eşiklenmiş sıklık göstergesi

Değer kümelerinin daha iyi anlaşılması için görüntünün küçük bir kesiti (Şekil 3(a) dikdörtgen alan içine alınan) alınarak sayısal olarak incelenirse Şekil 3(b)'deki dizey elde edilir. Bu dizeyde siyah gözetler 0 değerini alırken, beyaz gözetler 1 değerine karşılık gelmektedir.

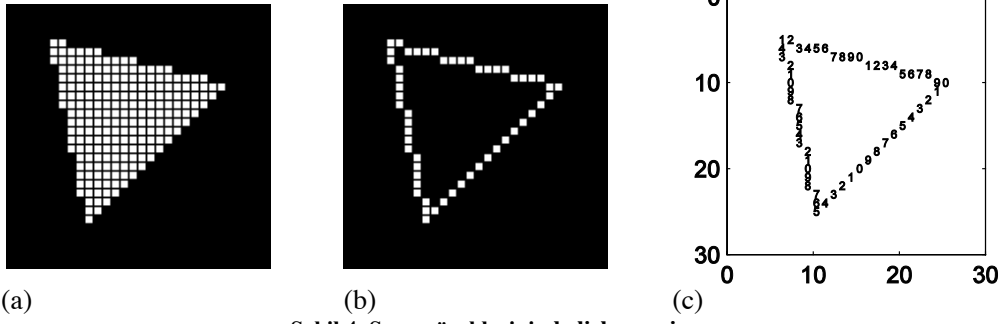


(a) (b) (c)  
Şekil 3. Sayısal görüntü ve dikdörtgen içinde seçilmiş nesne  
(a) büyütülmüş ve ayrıklaştırılmış görüntü; (b) sayısal görüntü dizeyi; (c) sınırları belirlenmiş sayısal görüntü dizeyi

Şekil 1'de ön işlemlerden geçirilerek kullanıma hazır duruma getirilmiş görüntü içerisindeki nesnelere tek tek etiketlenir. Etiketleme işlemi her gözetin 4 komşusuna (kuzey, güney, doğu ve batı) bakılarak, her bir nesneye farklı gözet değerleri atanması yoluyla gerçekleştirilir (Kesemen vd., 2010). Etiketlenen her nesne ayrı ayrı ele alınarak, iki boyutlu ikinci dereceden sayısal türev yöntemiyle nesnenin sınır gözetleri bulunur. Bu sınır gözetleri nesnenin sol-üst köşesinden itibaren saat yönünde gezilerek her gözetin köşe olup olmayacağı araştırılır.

## 2.1 Baskın Nokta Algılama

Bu çalışmada köşe gözetlerinin bulunması için dört yöntem önerilmiştir. Bu yöntemlerin başarımını ölçmek için köşe konumları net (tartışmasız) olarak bilinen şekillerin kullanılması gerekmektedir. Gerçek yaşamdan örnekler seçildiğinde bunların baskın noktaları tam olarak bilinmemektedir. Dolayısıyla, Şekil 4(a)'daki sayısal görüntü ele alınmış ve bu görüntünün sınır gözetleri (Şekil 4(b)) Şekil 4(c)'deki gibi rakamlarla işaretlenmiştir. ( $C = \{p_n = (x_n, y_n), n = 1, 2, \dots, N\}$ ).



Şekil 4. Sınır gözeklerinin belirlenmesi  
(a) sayısal görüntü; (b) sınır gözekleri; (c) sınır numaralandırılması

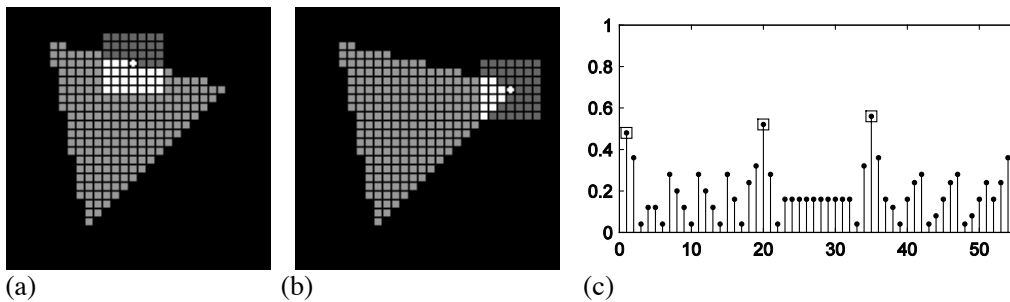
İşaretlenmiş sınır gözekleri boyunca her gözek tek tek ele alınarak (etkin gözek) baskın nokta olup olmadığı aşağıda önerilen yöntemler yardımıyla incelenmiştir. Her yöntemde elde edilen sonuçlar  $[0,1]$  aralığına indirgenerek (anlamlılık işlevi,  $e_n$ ) görsel olarak karşılaştırılması sağlanmıştır.

### 2.1.1 Kütle-zemin oranı yöntemi

Etkin gözek komşuluğundaki gözeklerden nesne olarak etiketlenmiş gözeklerin sayısı ile zemin olarak etiketlenmiş gözeklerin sayısının oranlanması yöntemidir. Bu oran kenar gözekleri boyunca 1'e yakın olup küçük sapmalar göstermekte, ancak köşeye yaklaşıldıkça 1'den uzaklaşmakta ve büyük sapmalar göstermektedir. Bu ise ilgili etkin gözeğin muhtemel bir köşe noktası olduğu anlamına gelmektedir. Şekil 5(a) ve (b)'de en koyu noktalar zemin gözeklerini gösterirken en açık noktalar kütle (nesne) gözeklerini göstermektedir. Yuvarlak olarak işaretlenen gözek ise o andaki etkin gözeği göstermektedir. Etkin gözek etrafında  $m = 3$  komşuluğunda toplam  $M = (2m + 1)^2 = 49$  gözek bulunmaktadır. Bu gözeklerden nesneye ait olanların  $s_n$  ile gösterilmektedir. Denklem (2) verilen eşitlikte anlamlılık işlevi

$$e_n = 1 - \frac{2}{M} \begin{cases} M - s_n, & s_n > M/2 \\ s_n, & s_n \leq M/2 \end{cases}, \quad n = 1, 2, \dots, N \quad (2)$$

biçiminde hesaplanır. Şekil 5(c)'de 1, 20 ve 34. noktalarında anlamlılık işlevi komşu değerlere göre en yüksek olan üç noktayı verilmektedir. Bu noktalar baskın gözek konumlarına karşılık gelmektedir.



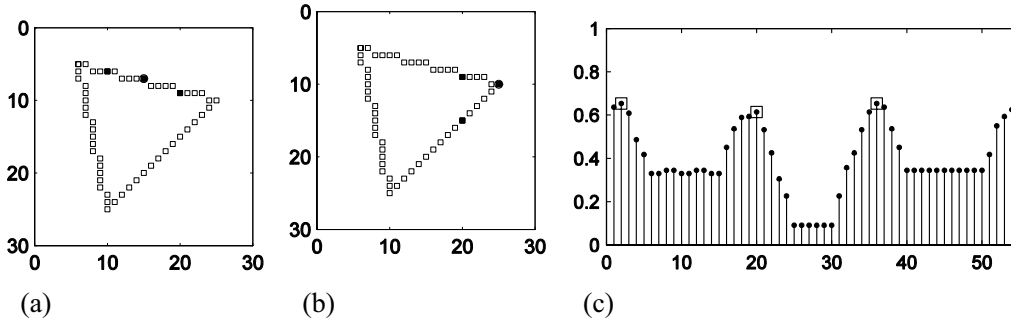
Şekil 5. Sınır gözeklerinin kütle-zemin oranlaması  
(a) kenardan seçilmiş bir etkin gözek 3 komşuluğu; (b) köşeden seçilmiş bir etkin gözek 3 komşuluğu; (c) tüm sınır gözekleri için anlamlılık işlev değerlerinin grafiği ve dörtgen kutucuk içerisinde alınmış baskın noktalar

### 2.1.2 Göreceli uzaklıklar yöntemi

Sınır gözekleri boyunca kaydırılarak seçilen her etkin gözekten birbirine zıt yönde  $m$  gözek ötedeki iki gözek ele alınır. Bu iki gözek arasındaki uzaklık ( $d_n$ ), her üç gözek de aynı kenar üzerindeyse (Şekil 6(a)) en fazla  $\sqrt{2M}$  uzunluğunda olacaktır. Ancak etkin gözek köşe noktasına yaklaştıkça bu uzaklık köşenin oluşturduğu açının etkisiyle azalacaktır (Şekil 6(b)). Uzaklıklar, anlamlılık işlevi yardımıyla,

$$e_n = 1 - \frac{d_n}{\sqrt{2M}} \quad (3)$$

biçiminde birleştirilir. Uzaklığın en az olduğu, yani belirlenen iki gözeğin birbirine en yakın olduğu nokta ise olası köşe noktası olarak kabul edilecektir (Şekil 6(c)). Bu noktalarda anlamlılık işlevi komşularına göre en yüksek değeri vermektedir.



Şekil 6. Sınır gözeklerinin uzaklık ölçütü

(a) kenardan seçilmiş bir etkin gözek ve zıt yönde a gözek uzaklığında iki gözek; (b) köşeden seçilmiş bir etkin gözek ve zıt yönde a gözek uzaklığında iki gözek; (c) tüm sınır gözekleri için anlamlılık işlev değerlerinin grafiği ve dörtgen kutucuk içerisine alınmış baskın noktalar

### 2.1.3 Kesişim açısı yöntemi

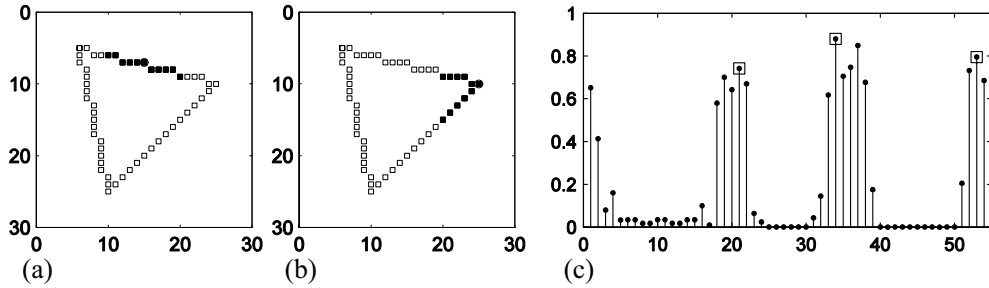
Etkin gözeğin zıt yönlerindeki  $m$  kadar gözeğin konumlarının oluşturduğu iki veri kümesinden hesaplanan iki bağlantım (regression) doğrusunun kesişim açısı baskın noktanın belirlenmesinde bir ölçüt olarak alınabilir. Bu iki bağlantım doğrusunun kesişim açısı,

$$\theta = \text{atan} \left( \frac{m_1 - m_2}{1 + m_1 m_2} \right) \quad (4)$$

eşitliğiyle bulunur. Burada  $m_1$  ve  $m_2$  sırasıyla birinci ve ikinci bağlantım doğrusunun eğimleridir. Kesişim açısının anlamlılık işlevi,

$$e_n = \frac{2}{\pi} |\theta| \quad (5)$$

biçiminde tanımlanabilir. Eğer etkin gözek köşe gözeği ise  $e_n$  değeri 1'e yaklaşacaktır. Her etkin nokta için kesişim açısı hesaplanır, istenen sayıda köşe için en yüksek anlamlılık işlev değerine sahip etkin gözekler işaretlenerek belirlenir (Şekil 8(c)).



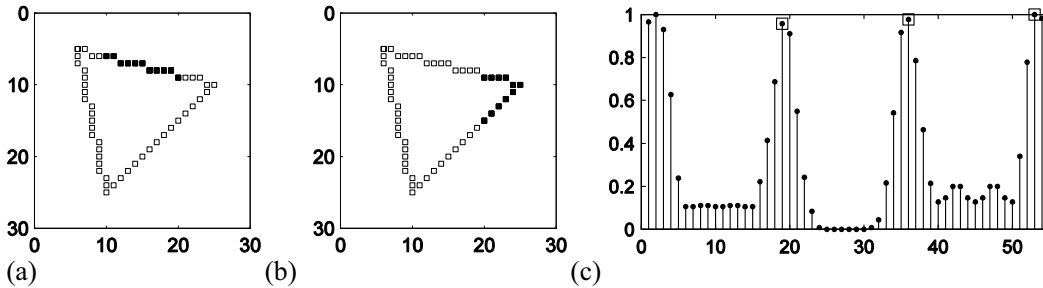
Şekil 8. Etkin gözetken geçen iki bağlanım doğrusu arasındaki açının belirlenmesi  
 (a) kenardan seçilmiş etkin gözeğin zıt yönlerinde a tane gözeğin oluşturduğu iki değerler kümesi; (b) köşeden seçilmiş etkin gözeğin zıt yönlerinde m tane gözeğin oluşturduğu iki değer kümesi; (c) tüm sınır gözetkenleri için anlamlılık işlev değerlerinin grafiği ve dörtgen kutucuk içerisine alınmış baskın noktalar

#### 2.1.4 Belirtme katsayısı yöntemi

Kesişim açısı yöntemindeki iki bağlanım doğrusu yerine tek bağlanım doğrusu seçilir. Seçilen bu noktalardan belirtme katsayısı ( $r^2$ ) hesaplanır. Belirtme katsayısının anlamlılık işlevi,

$$e_n = 1 - r^2 \quad (6)$$

biçiminde hesaplanır.  $e_n$  değerinin 1'e yaklaşması (belirtme katsayısının düşmesi) baskın noktaya yaklaşıldığını göstermektedir. İlgili gözetkenin doğru etrafında sapmaları düşükse düz bir kenar olduğu (Şekil 9(a)), sapmaları yüksekse, yani köşelere yaklaşırsa (Şekil 9(b)) baskın noktaya yaklaşıldığı söylenebilir (Şekil 9(c)'de dörtgen kutucuklar).



Şekil 9. Etkin gözetken merkezli  $2m + 1$  sayıda gözeğin oluşturduğu kümenin belirtme katsayıları  
 (a) kenardan seçilmiş bir dizi gözetken; (b) köşeden seçilmiş bir dizi gözetken; (c) tüm sınır gözetkenleri için anlamlılık işlev değerlerinin grafiği ve dörtgen kutucuk içerisine alınmış baskın noktalar

### 3. BULGULAR ve SONUÇLAR

Önerilen tüm yöntemler sonucunda elde edilen çözüm dizisinde en büyük değerlerin öncelikle bulunması gerekir. Ancak köşe (baskın) değerlere yaklaştıkça çözüm yoğunlaşması bu konumda birçok gözetkenin baskın nokta olabilme olasılığını artırmaktadır. Bu durumda çözüm dizisi düşey yönde iki ayrı küme gibi değerlendirilip parçalara ayrılır. Daha sonra baskın nokta olma olasılığı yüksek olan dizi kümelerinin ortanca değeri kullanılarak yaklaşık baskın nokta bulunur (Güngör, 2012).

Bu inceleme sonucunda, Kütle zemin oranı yöntemi, seçilen komşuluk dörtgeninin büyük olması durumunda iyi sonuçlar vermektedir. Göreceli uzaklıklar yöntemi köşelerin dar açılı, geniş açılı, çokgenin içbükey veya dışbükey olması durumunda bile

yüksek başarı göstermektedir. Kesişim açısı yöntemi yöneysel yaklaşımlara (The vd., 1989), (Marji vd., 2003), (Costa vd., 2001) göre düzgün olmayan kenarlarda da başarılı sonuçlar vermektedir. Belirtme katsayısı yöntemi dar açılarda başarı gösterirken geniş açılarda baskın nokta ölçütünün belirlenmesinde diğer yöntemlerden daha başarılı değildir.

Önerilen yöntemlerin başarılarını sınamak ve karşılaştırmak için birçok farklı şekle sahip çokgenler üzerinde denenmesi için kenar sayısı rastgele seçilen 1000 tane yapay çokgen şekil içeren sayısal görüntü oluşturuldu. Oluşturulan çokgenlerin köşe noktaları önceden bilindiğinden önerilen yöntemler sonucunda elde edilen baskın noktalar ile karşılaştırılır yöntemlerin başarımının incelenmesi mümkün olmuştur (Tablo-1).

**Tablo 1. Yöntemlerin benzetimsel karşılaştırılması**

	$\mu_B$	$\sigma_B$	$\mu_t$	$\sigma_t$
Kütle/Zemin Oranı	95.6000	2.9136	0.7452	0.0222
Göreceli Uzaklık	91.3000	2.5841	0.7060	0.0152
Kesişim Açısı	79.0000	4.2947	5.5617	0.0533
Belirtme Katsayısı	81.4000	4.6236	1.4246	0.0276

Ortalama başarı oranı ( $\mu_B$ ), seçilen  $i$ . çokgende kestirilen tüm köşe noktalarının ( $q_{in}$ ) hepsinin bilinen köşe noktalarının ( $p_{in}$ ) 3 gözetim yakınında olması durumunda başarılı, aksi durumda başarısız sayılmaktadır (González vd., 2007).

$$\mu_B = \frac{1}{10} \sum_{i=1}^{1000} \prod_{n=1}^{N_i} \begin{cases} 1, & |p_{in} - q_{in}| < 3 \\ 0, & \text{diğer durumlarda} \end{cases} \quad (7)$$

$\mu_B$ : ortalama başarı oranı (%);  $\sigma_B$ : başarı oranı standart sapması (%);  
 $\mu_t$ : ortalama hesaplama zamanı (sn);  $\sigma_t$ : hesaplama zamanı standart sapması (sn).

Tablo 1'e göre her yöntemin köşe konumları bilinen nesnelere ne kadar yaklaştığı veya tam olarak doğru bulup bulmadığı ortalama başarı oranları ile ortaya konmuştur.

#### 4. KAYNAKLAR

Kesemen, O., Aktaş, N., Solak, Y.N., 2010. Sayısal Görüntülerdeki Taneciklerin Topaklık Ölçüsünün Belirlenmesi, 9. Matematik Sempozyumu, 20-22 Ekim, Trabzon.

Bribiesca, E., Guzman, A., 1980. How to Describe Pure Form and How to Measure Differences in Shapes Using Shape Numbers, Pattern Recognition, 12 (2) 101-112.

O'Gorman, L., Kasturi, R., 2009. Document Image Analysis, IEEE Computer Society Executive Briefings.

The, C. H., Chin, R. T., 1989. On the Detection of Dominant Points on Digital Curves, IEEE Transaction on Pattern Analysis and Machine Intelligence, 11 (8) 859-872.

Marji, M., Siy, P., 2003. A New Algorithm for Dominant Points Detection and Polygonization of Digital Curves, Pattern Recognition, 36 (3) 2239-2251.

Costa, L. da F., Cesar Jr., R. M., 2001. Shape Analysis and Classification, CRC Press.

González, R. C., Woods, R. E. (2007), Digital Image Processing, Prentice Hall. New Jersey.

Kesemen, O., Karakaya, G., Gürbüz, F., 2010. Sayısal Görüntülerde Taneciklerin Bulanık Yönelimlerinin Belirlenmesi, İstatistik Günleri Sempozyumu, 28-30 Haziran, Ankara.

Güngör, Ç., 2012. İkili Görüntülerde Nesnelerin Baskın Noktalarının Belirlenmesinde Optimizasyon Algoritmaları, Yüksek Lisans Tezi, KTÜ, Trabzon.

## STATISTICAL APPROACHES FOR THE DETECTION OF DOMINANT POINTS OF OBJECTS IN DIGITAL IMAGES

### ABSTRACT

*In this paper, four different methods are proposed for the detection of dominant points of objects in digital images, and these methods are examined for performance. Some statistical methods and geometry rules are used in the proposed algorithms.*

**Keywords:** Dominant point detection, Polygonal approximation, Corner detection, Digital curves, Shape analysis.



# ÜNİVERSİTE ÖĞRENCİLERİNİN KAN BAĞIŞI HAKKINDAKİ TUTUMLARININ LOJİSTİK REGRESYON ANALİZİ İLE İNCELENMESİ

Barış ERGÜL\*

Zeynep FİLİZ\*\*

## ÖZET

*Günümüz araştırmalarında, verilerin çok değişkenli bir yapıda olması, ileri istatistik yöntemlerin kullanılmasını gerektirmektedir. Çok değişkenli çözümlerinin yaygınlaştığı günümüzde, bu yöntemlerin en çok uygulandığı alanlardan birisi de sağlık konusudur. Bu çalışmada Lojistik Regresyon Analizi'nin tanıtılmasının yanı sıra, sağlık alanındaki bir uygulama ile sunumu amaçlanmaktadır. Lojistik Regresyon Analizini uygulamak için, Eskişehir Osmangazi Üniversitesinde öğrenim gören toplam 929 öğrencinin, kan bağıışı hakkındaki bilgi, tutum ve davranışlarını öğrenmek amacıyla 17 sorudan oluşan, bir anket formu uygulandı.*

**Anahtar Kelimeler:** Kan bağıışı, Lojit, Lojistik regresyon, Üniversite öğrencileri.

## 1. GİRİŞ

Doğada olaylar incelendiğinde olayların detayları anlaşıldıkça, arka planda yatan birçok nedenin bulunduğu da fark edilmektedir. Sağlık alanındaki oluşumlar; teknolojik gelişmelere ve karşılaşılan problemlere göre daha çok ayrıntılı biçimde incelemeyi gerektiren boyutlara ulaşmış bulunmaktadır. Gün geçtikçe basit, tek değişkenli ve ikili ilişkilerle açıklanmaya çalışılan sorunlar artık daha ayrıntılı biçimde çoklu ya da çok değişkenli yöntemlerle açıklanma gereğini gündeme getirmektedir. Eskiden bilimsel çalışmalarda yer verilmeyen değişkenler, değişkenler arası ilişkiler de günümüzdeki araştırmalarda ele alınarak çözümlenmesi hedeflenmektedir.

Olaylar arasındaki neden-sonuç ilişkilerinin ortaya konulması, risk faktörlerinin belirlenmesi ve bu risk faktörlerinin hangilerinin majör hangilerinin minör yapıda, önemsenmeyecek boyutlarda olduğunun belirlenmesi ileri istatistiksel yöntemlerle ortaya konmaya çalışılmaktadır.

Günümüzde yapılan bilimsel araştırmalarda güvenilir sonuçlara varılabilmesi için kararların, istatistik yöntemlere dayandırılması gerekmektedir. İncelenen olayların karmaşık ve bu olayların çözümü için önerilen yolların da fazla olması, olayı açıklamada kullanılacak değişken sayısını da arttırmaktadır. Bu amaçla olayların çözümünde, birden fazla değişkeni konu alıp, bunların analizleriyle uğraşan istatistiksel modellerin kullanılması gerekmektedir.

Kategoriler şeklinde sonuçlanan olaylar ile bu olayları etkileyen faktörler arasındaki neden-sonuç ilişkisini belirlemede Lojistik Regresyon Analizi kullanılmaktadır (Çolak,2004).

\*Arş. Gör., Eskişehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, e-posta: [bergul@ogu.edu.tr](mailto:bergul@ogu.edu.tr)

\*\*Doç. Dr., Eskişehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, e-posta: [zfiliz@ogu.edu.tr](mailto:zfiliz@ogu.edu.tr)

Lojistik Regresyon Analizi, klinik uygulamalarda elde edilen değişkenlerin her zaman süreklilik göstermemesi ve Lojistik Regresyon Analizinin bu değişkenlerle çözümlemede başarılı olması nedeniyle son yıllarda önemini arttırmıştır (Akkuş vd., 2005).

Lojistik Regresyon Analizinde hata terimlerinin normal dağılım yerine binom dağılım göstermesi kritik bir noktadır. Bu durum ele alınan konuda doğrusal regresyon tekniğinin kullanılmasını engellemektedir. Lojistik Regresyon Analizi, gözlem değerlerinin herhangi bir gruba atanması işlemini yapmaya yardımcı olan bir regresyon yöntemidir. Normal dağılım varsayımı, süreklilik varsayımı gibi ön koşullara ihtiyacı yoktur. Böylece bağımlı değişken üzerinde açıklayıcı değişkenlerin etkilerinin olasılık olarak elde edilmesi sağlanır (Albert vd.,1986), (Lemeshow vd., 2000), (Tatlıdil,1996).

Bu çalışmada çok değişkenli analiz tekniklerinden Lojistik Regresyon Analizi'nin tanıtımı yanında sağlık alanında bir uygulamanın sunumu amaçlanmaktadır.

## 2. LOJİSTİK REGRESYON ANALİZİ

Lojistik Regresyon Fonksiyonu aşağıdaki gibi tanımlanır (Daştan, 2003).

$$P_i = E(Y = 1|X_i) = \frac{1}{1+e^{-(b_0+b_1x_i)}} = \frac{1}{1+e^{-z_i}} \quad (1)$$

Burada doğal logaritma tabanı;  $e = 2,718$  ve  $z_i = b_0 + b_1x_i$  şeklindedir.

$Y_i = 1$  bir olayın olması,  
 $Y_i = 0$  bir olayın olmaması,  
 $X_i = k$  sayıda gruba sahip bağımsız değişkeni

belirtir.

Bir olayın olma olasılığı  $P_i$  iken,  $1 - P_i$  ise bir olayın olmama olasılığını gösterir. O halde  $P_i$ , aşağıdaki şekilde edilir;

$$1 - P_i = \frac{e^{-z_i}}{1+e^{-z_i}} \text{ ve } \frac{P_i}{1-P_i} = \frac{\frac{1}{1+e^{-z_i}}}{\frac{e^{-z_i}}{1+e^{-z_i}}} = \frac{1}{e^{-z_i}} \quad (2)$$

Bu orana, bir olayın olması lehine fark oranı adı verilir.

İşte lojistik model, yukarıdaki modelin her iki tarafının doğal logaritması alınarak şöyle elde edilmektedir.

$$L_i = \ln \left[ \frac{P_i}{1-P_i} \right] = z_i = b_0 + b_1x_i \quad (3)$$

Bu modelde,  $L_i$ , fark oranı logaritması olup, hem  $X_i$  hem de  $b_i$  parametrelerine göre doğrusaldır. Lojistik modelde,  $Z_i$  değişkeni  $-\infty$ 'dan  $+\infty$ 'a kadar olan aralıkta değişirken,  $P_i$  de 0 ve 1 arasında değer alır ve  $Z_i$  ile ilişkisi doğrusal değildir.

Koşullu en çok olabilirlik fonksiyonunun kısmi türevleri alınarak bilinmeyen model katsayılarının ve standart hatalarının kestirimi yapılır (Albert vd., 1986), (Lemeshow vd., 2000).

Gözlemlerin ele alınan gruplardan birine yerleştirilmesinde, son aşamada elde edilen değerler yardımıyla  $P_i$  olasılıkları bulunarak, bu tahmini değer  $P_i < 0,50$  olduğunda gözlem değeri,  $Y_i = 0$  grubuna;  $P_i > 0,50$  olduğunda ise diğer gruba yerleştirilir (Tatlıdil, 1996).

## 2.1 Modelin ve Katsayıların İstatistiksel Anlamlılığının Sınanması

### 2.1.1 Modelin genel anlamlılığının sınanması

Çoklu regresyondaki hata kareleri toplamına benzer olarak -2 Log Likelihood ile ifade edilen bu değer olabilirlik değerinin -2 kez logaritmasıdır.

Regresyon analizinde,  $k - 1$  ve  $n - k$  serbestlik derecesinde  $F$  istatistiğinde, kesim noktası hariç tüm katsayılarının birleştirildiği  $H_0$  hipotezi test edilir.  $H_0$  hipotezi kabul edildiğinde  $\chi^2$  dağılımı kullanılarak aşağıdaki istatistik yardımı ile hesaplanır.

$$C = -2 \log \left[ \frac{L_0}{L_1} \right] = (-2 \log L_0) - (-2 \log L_1) \quad (4)$$

$L_1$ , modelin tamamı için olabilirlik fonksiyonunun değerini verir.  $L_0$  ise kesim noktası hariç tüm katsayılar 0 ise olabilirlik fonksiyonunun maksimum değerini verir. Bu kesim noktası hariç tüm katsayılar 0 olduğu hipotezi için hesaplanan  $\chi^2$  değeri ile test edilir.  $\chi^2$  için serbestlik derecesi ( $k - 1$ ) başka bir deyişle hipotezdeki katsayı sayısıdır. Testin uygulanmasında ( $k - 1$ ) serbestlik derecesiyle ve  $\alpha$  anlamlılık düzeyinde  $\chi^2$  dağılımı tablosundan alınan kritik değer ile hesaplanan  $C$  değeri karşılaştırılarak hipotez test edilir (Aldrich vd., 1984), (Buse, 1982).

### 2.1.2 Modelin katsayılarının anlamlılık sınaması

Model katsayıları ve katsayıların standart hataları, en çok olabilirlik yöntemi kullanılarak elde edilir (Albert vd., 1986), (Lemeshow vd., 2000).

Katsayıların anlamlılık testi,

$$WALD^{1/2} = \left[ \frac{b_i - 0}{s_{b_i}} \right] \quad (5)$$

şeklinde WALD istatistiği ile yapılır ve bu değer standart normal dağılıma ait tablo ile karşılaştırılarak katsayıların anlamlılık sınamaları yapılır (Ürük, 2007).

## 3. UYGULAMA

ESOGÜ'de öğrenim gören 929 öğrenciden 17 sorudan oluşan ankete cevap vermeleri istenmiştir. Çalışmada cinsiyet, yaş, öğrenim görülen fakülte, kan bağışı yapma sıklığı, başkasının kanı ile yaşama, kan bağışının nereye yapıldığı, kan grubu, bağış yapmadan kan grubunun bilinip bilinmemesi, kan bağışının faydalı olup olmadığı, kan bankası, iğne ya da doktor korkusu, öğrencileri bilgilendirmek için kan bağışı ile ilgili bir dersin

konulması, kan bağıışı tanıtım kampanyası, kan sorunu yaşayan akraba ve kan bağıışında bulunup bulunmama deęişkenleri incelemeye alınmıştır.

Uygulanan analiz sonucunda, uyum iyilięi testleri sonuçları tabloda gösterildięi gibidir.

**Tablo 1. Modelin genel anlamlılık testi sonucu**

Uyum iyilięi testi	Deęeri
-2 Log Likelihood	56,127

Modelin genel olarak anlamlılıęının test edilmesinde ise  $\chi^2$  testi yöntemi uygulanır. Testin uygulanmasında;  $(\chi^2(k-1, \alpha))$   $\alpha$  anlamlılık düzeyi olmak üzere,  $(k-1)$  serbestlik derecesi ile  $\chi^2$  daęılımı tablosundan alınan kritik deęeri ile  $C$  hesaplanan istatistięi ile karşılaştırılarak hipotez test edilir.

$H_0: b_0 = b_1 = \dots = b_k = 0$  (Model genel olarak anlamsızdır.)

$H_1: En az bir b_k anlamlı$  (Model genel olarak anlamlıdır.)

$$C = -2LL = 56,127 > \chi^2(k-1, \alpha) = 7,815$$

olduęu için  $H_0$  hipotezi reddedilir. Buna göre kurulan regresyon modeli genel olarak anlamlıdır denilir.

Modeldeki deęişkenlerin anlamlılık sınamaları sonucunda, modelin sabit katsayısının, bağımsız deęişkenlerden; kan bağıışı yapma sıklıęı, yaşı, kan bağıışı ile ilgili tanıtım kampanyası deęişkenlerinin katsayılarının istatistiksel olarak anlamlı olduęuna karar verilir (WALD >2).

**Tablo 2. Modelin katsayı kestirimleri sonuçları**

Deęişken adı	b	S.E.	Wald
AKRABA	0,596	0,863	0,477
BANKA	-0,104	0,356	0,085
CİNSİYET	1,194	1,056	1,278
DERS	0,047	1,059	0,002
FAKÜLTE	-0,105	0,354	0,088
GÖNÜL	-9,848	29,781	0,109
GRUPBİL	-0,858	0,860	0,994
KAÇBAĞIŞ	-6,940	1,056	43,202*
KANGRB	-0,257	0,199	1,666
KORKU	-0,201	0,956	0,044
NEREYE	-0,149	0,857	0,030
SAĞLIK	-0,017	0,373	0,002
TANITIM	-0,730	0,473	2,375*
YAŞ	0,866	0,588	2,168*
YAŞAM	1,363	2,030	0,451
Constant (Sabit)	4,369	6,242	2,206*

#### 4. SONUÇ VE ÖNERİLER

Toplumunu oluşturan bireyler yaşamı boyunca değişik zamanlarda çeşitli nedenlerden dolayı kendisi, ailesi ya da yakın çevresiyle ilgili olarak zaman zaman kan ve kan ürünlerine gerek duymaktadır. Kan ve kandan elde edilen ürünler laboratuvar koşullarında elde edilemediğinden, hayati önem taşıyan kanın sağlıklı bireylerden sağlanması gerekmektedir.

Yapılan analizde modelin genel olarak anlamlı olduğu görülmüştür. Modeldeki değişkenlerin anlamlılık sınımaları sonucunda, modelin sabit katsayısının, bağımsız değişkenlerden; kan bağışı yapma sıklığı, yaş, kan bağışı ile ilgili tanıtım kampanyası değişkenlerinin katsayılarının istatistiksel olarak anlamlı olduğuna karar verilmiştir. Toplumumuzun önemli bir kısmının kan bağışlamadığı gerçeğinden yola çıkılarak bu çalışmanın sonuçları değerlendirilip bu veriler ışığı altında kan bağışı konusunda toplumun eğitilmesi ve gönüllü kan bağışında bulunmanın özendirilmesi gerekmektedir.

En yakın zamanda kan bağışı hakkında toplumumuzu bilinçlendirme çalışmalarına başlamamız gerekmektedir. Kan bağışı yapmayı en fazla etkileyen değişkenler kan bağışı yapma sıklığı, yaş, kan bağışı ile ilgili tanıtım çalışmalarının yapılmasıdır. Bu da bize gösteriyor ki bir kez kan bağışının toplum ve kendisi için önemini öğrenen kişiler tekrar kan bağışında bulunmaktadır. Kan bağışı hakkında tüm toplumu bilinçlendirerek onları teşvik etmenin yolları aranmalıdır.

Toplumumuzda bir damla acil kana ihtiyacı olan, yarına umutla bakmak ve gülmek belki de tekrar eskisi gibi mutlu olmak isteyen birçok insanı düşünerek kan bağışı yapmaya tüm toplumumuzu teşvik etmeliyiz.

Ayrıca ülkemizde insanları gönüllü kan bağışında bulunmaya teşvik etmek için çeşitli kampanyaların düzenlenmesi gerektiği de yapılan analiz sonucunda ortaya çıkmıştır.

#### 5. KAYNAKLAR

Aldrich, H. J., Nelson D. F., 1984. Linear Probability, Logit and Probit Models, Sage University Paper, Series: Quantitative Applications in the Social Sciences No: 45.

Akkuş, Z., Çelik, Y., Satıcı, Ö., Daşdağ, N. M., Sanisoğlu, Y., 2005. Hastane Personelinin Kan Bağışı Hakkındaki Bilgi, Tutum ve Davranışlarının Çok Değişkenli Lojistik Regresyon Yöntemiyle İncelenmesi, İnönü Üniversitesi Tıp Fakültesi Dergisi, 12(1). 25-29.

Albert, A., Lesaffre, E., 1986. Multiple Group Logistic Discrimination, Computational Mathematics With Applications, 12 A, 2. 209-224.

Buse, A., 1982. The Likelihood Ratio, Wald and Lagrange Multiplier Tests: An Expository Note, The American Statistician, August, Vol 8, No.3, Part 1. 153-157.

Çolak, E., 2004. Koşullu ve Sınırlandırılmış Lojistik Regresyon Yöntemlerinin Karşılaştırılması ve Bir Uygulama, ESOĞÜ Sağlık Bilimleri, Eskişehir.

Daştan, S., 2003. Lojistik Regresyon Tekniđi İle Hisse Senedi Getirilerinin Analizi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Ana Bilim Dalı, İstanbul.

Lemeshow, S., Hosmer, D., 2000. Applied Logistic Regression, Wiley Series in Probability and Statistic, Wiley Interscience, 2 Sub Edition, New York.

Tatlıdil, H., 1996. Uygulamalı Çok Deđişkenli İstatistiksel Analiz, Ankara.

Ürük, E., 2007. İstatistiksel Uygulamalarda Lojistik Regresyon Analizi, Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

## **EVALUATION OF THE ATTITUDE OF UNIVERSITY STUDENTS ON BLOOD DONATION BY LOGISTIC REGRESSION**

### *ABSTRACT*

*The multivariate structure of data in today's researches require the use of advanced statistical methods. Now that the multivariate analyzes are so prevalent, health branch is one of the most commonly applied fields of these methods. In this study, the introduction of Logistic Regression Analysis is aimed together with a presentation of an application in the health field. To apply the Logistic Regression Analysis, a questionnaire consisting of 17 statements is conducted to 929 students at the University of Eskisehir Osmangazi for determining their knowledge, attitudes and behaviors of the students on blood donation.*

**Keywords: Blood donating, Logit, Logistic regression, University students.**

## DANIŐMA KURULU ÜYELERİ - ADVISORY BOARD MEMBERS

Ali YAZICI	Atılım Üniversitesi
Alper GÜVEL	Çukurova Üniversitesi
Asaf Savaş AKAT	Bilgi Üniversitesi
Aşır GENÇ	Selçuk Üniversitesi
Aydın ÖZTÜRK	Ege Üniversitesi
Ayşe GÜNDÜZ HOŐGÖR	Orta Doęu Teknik Üniversitesi
Bedriye SARAÇOęLU	Gazi Üniversitesi
Coşkun Can AKTAN	Dokuz Eylül Üniversitesi
Deniz GÖKÇE	Bahçeşehir Üniversitesi
Ekrem ERDEM	Erciyes Üniversitesi
Ercan UYGUR	Türkiye Ekonomi Kurumu
Erdem BAŐCI	T.C. Merkez Bankası
Erinç YELDAN	Bilkent Üniversitesi
Erol TAYMAZ	Orta Doęu Teknik Üniversitesi
Eser KARAKAŐ	Bahçeşehir Üniversitesi
Fatih ÖZATAY	TOBB Ekonomi ve Teknoloji Üniversitesi
Fatin SEZGİN	Bilkent Üniversitesi
Fikri AKDENİZ	Çukurova Üniversitesi
Fikri ÖZTÜRK	Ankara Üniversitesi
Gülay BAŐARIR KIROęLU	Mimar Sinan Güzel Sanatlar Üniversitesi
Güven SAK	TOBB Ekonomi ve Teknoloji Üniversitesi
Haluk LEVENT	Galatasaray Üniversitesi
Hamza EROL	Çukurova Üniversitesi
İlhan TEKELİ	Orta Doęu Teknik Üniversitesi
İmdat KARA	Başkent Üniversitesi
İnsan TUNALI	Koç Üniversitesi
Levent KANDİLLER	Yaşar Üniversitesi
Mehmet KAYTAZ	Işık Üniversitesi
Meltem DAYIOęLU TAYFUR	Orta Doęu Teknik Üniversitesi
Metin TOPRAK	İstanbul Üniversitesi
Mustafa ACAR	Aksaray Üniversitesi
Mustafa AYTAÇ	Uludağ Üniversitesi
Nihat BOZDAę	Gazi Üniversitesi
Onur BASKAN	Ege Üniversitesi
Orhan GÜVENEN	Bilkent Üniversitesi
Ömer Faruk ÇOLAK	Gazi Üniversitesi
Ömer L. GEBİZLİOęLU	Kadir Has Üniversitesi
Özkan ÜNVER	Ufuk Üniversitesi
Öztaş AYHAN	Orta Doęu Teknik Üniversitesi
Reşat KASAP	Gazi Üniversitesi
Savaş ALPAY	SESRTCIC
Seyfettin GÜRSOY	Bahçeşehir Üniversitesi
Süleyman GÜNAY	Hacettepe Üniversitesi
Turan EROL	Ankara Strateji Enstitüsü
Ümit OKTAY FIRAT	Marmara Üniversitesi
Yasin AKTAY	Selçuk Üniversitesi
Yılmaz AKDİ	Ankara Üniversitesi