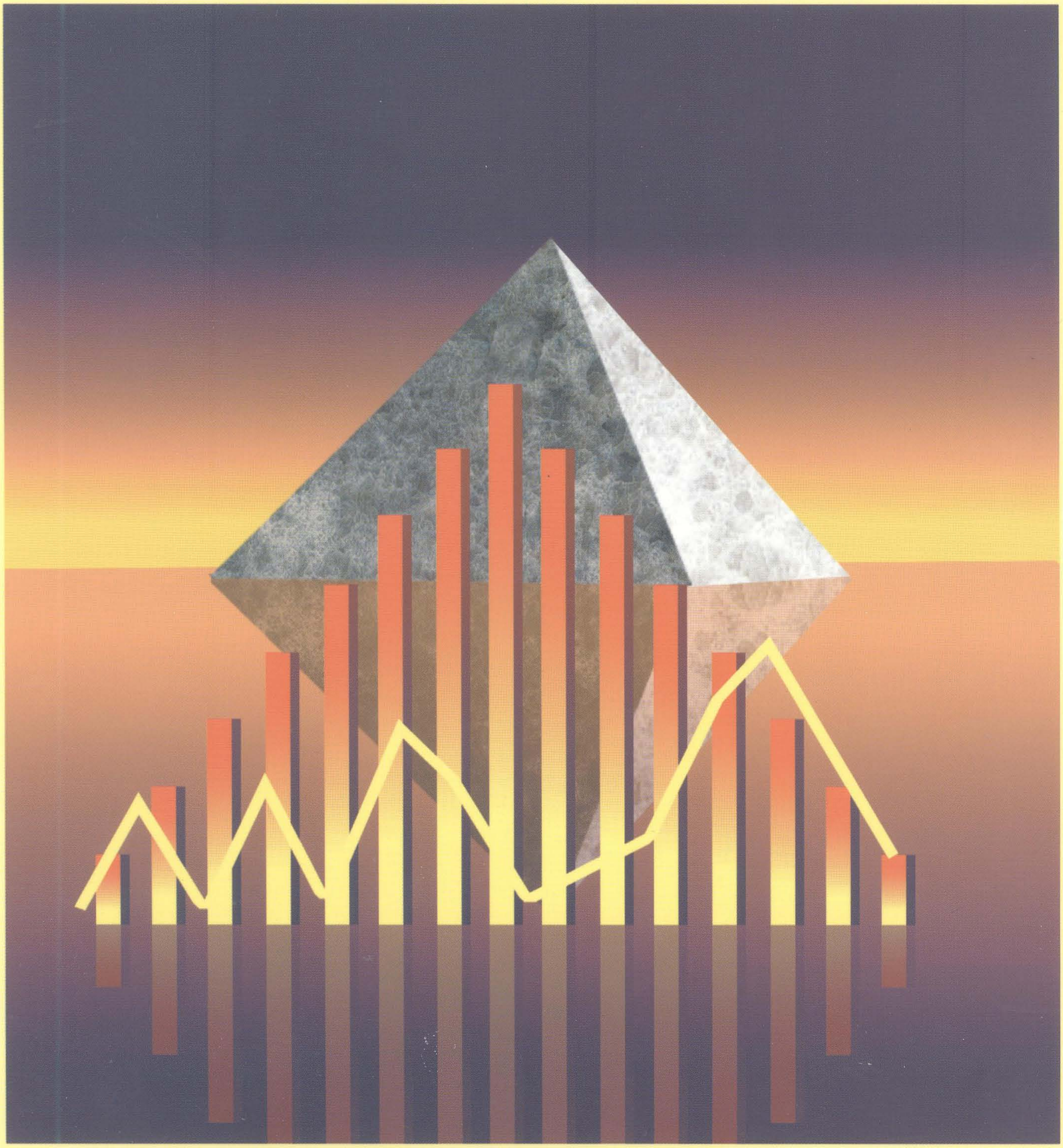
TÜİK

İstatistik Araştırma Dergisi

Journal of Statistical Research

Cilt 04 *Volume* No 03 *Number* Aralık 2005 *December*



İSTATİSTİK ARAŞTIRMA DERGİSİ

Amaç ve Kapsam

İstatistik Araştırma Dergisi, istatistiki araştırmaların kalitesinin arttırılmasını, istatistik metodolojisi ve uygulamasının geliştirilmesini, literatürde yer alan çalışmaların tartışılmasını, istatistik uygulamalarıyla ilgili araştırmaların ele alınmasını, teorik ve uygulamalı alanlardaki araştırmacılar arasındaki iletişimin ortak çalışma ve yayınlarla güçlendirilmesini amaçlayan bir yayındır.
İstatistik alanında aşağıdaki özellikleri taşıyan çalışmalar, Dergi kapsamında değerlendirilir:

1. İstatistik Teorisi, Olasılık Teorisi ve Stokastik Süreçler, Örnekleme ve Alan Araştırmaları, Uygulamalı İstatistik, İstatistiksel Kalite Kontrol, Biyoistatistik, Risk Aktüerya Analizi ve Sigortacılık, Ekonometri, Yöneylem Araştırması Uygulamaları, Demografi, Bilgisayar Uygulamaları ve Bilgi Sistemleri, gibi istatistiğin her dalında yeni bilgi üretimine yönelik olan tüm araştırmalar.
2. Sosyal Bilimler, Fen Bilimleri, Sağlık Bilimleri ve benzeri alanlara ilişkin veri derleme, veri çözümlemesi ve veri sunumu ile ilgili metodolojilerin geliştirilmesine yönelik araştırmalar.
3. Türkiye'de ve Dünya'da Resmi İstatistiklerin geliştirilmesine yönelik araştırmalar.
4. Yayımlanan istatistiki verileri, yeni bilimsel gelişmelerle analiz edip yorumlayan araştırmalar.

Yayın İlkeleri

1. Bu dergiye alınacak araştırmaların, özgün, yaratıcı, bilimsel kuram ve metodolojiye uygun olmaları, mevcut uygulama ve kurama katkıda bulunmaları esastır. Yayın dili Türkçe'dir.
2. Dergi, istatistiğin alanına giren tüm konuları kapsayan araştırmalara açıktır.
3. Dergi'de, 3 (üç) hakem tarafından incelenip, "**Yayımlanabilir**" olurunu almış araştırmalar yayımlanır. Yayımlanmayan makaleler sahibine geri verilmez.
4. Makaleler basılı dört kopya ve manyetik ortamda (3.5'lik diskette) dergi sekreteryasına gönderilir. Yayımlanmak üzere kabul edilmiş makale son düzeltmesi için yazar(lar)a gönderilir. Bu aşamada metnin değiştirilmesi değil, metne son şeklinin verilmesi beklenir.
5. 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.
6. Metin hazırlama kalıbına, telif haklarına uymayan veya herhangi bir yerde yayınlanmış ve yayınlanmak üzere kabul edilmiş çalışmalar, genel editör tarafından yazarına iade edilir.
7. Tüm yazışmalar dergi sekreteryası ile yapılır. Abonelik, eski nüshalar, makale ayrı basımları, reklamlar ve ödemelerle ilgili talepler isteme ve abone adresine yapılır

ISSN: 1303 - 6319

Genel Editör
Tel: +90 312 284 45 00/ 171
e-posta: fyildirim@cankaya.edu.tr
Editör Yardımcısı
Tel: +90 312 410 03 75
e-posta: seviluygur@tuik.gov.tr
İstatistik Araştırma Dergisi Sekreteryası
Basın ve Halkla İlişkiler Müşavirliği
Tel: +90 312 410 02 62 - 0312 410 01 10
Faks : +90 312 425 20 53
e-posta:dergi@tuik.gov.tr

T.C. Başbakanlık
Türkiye İstatistik Kurumu
Necatibey Cad. No: 114 06100 Yücetepe / ANKARA
Tel :+90 312 410 04 10 (Santral)
URL : http://www.tuik.gov.tr
Döner Sermaye İşletmesi Müdürlüğü
Tel : +90 312 410 03 23 -19
Faks : +90 312 417 58 86

Türkiye İstatistik Kurumu Matbaası – Ankara, Şubat 2007
MTB : 2007-0158-320 Adet

JOURNAL OF STATISTICAL RESEARCH

Owner
On Behalf of Turkish Statistical Institute
Ömer DEMİR

President, Turkish Statistical Institute

Editor in Chief
Fetih YILDIRIM

Assistant Editor
Sevil UYGUR

Objective and Scope

Journal of Statistical Research is a publication that aims to improve the quality of statistical researches, to develop the statistical methodology and application, to discuss the researches which take place in the literature, to assess the researches on statistical applications, to strengthen the communication between the researchers in theoretical and applied fields by associated studies and publications.

Researches having the following qualities in the field of statistics, are taken into consideration in the scope of the Journal:

1. Researches dealing with the production of new information on statistical matters such as Statistics Theory, Probability Theory and Stochastic Processes, Sampling and Survey, Applied Statistics, Statistical Quality Control, Biostatistics, Risk Actuary Analysis and Insurance, Econometrics, Operational Research, Demography, Computer Applications and Information Systems.
2. Researches dealing with the development of methodologies on data collection, evaluation and presentation in the fields of Social Sciences, Applied Sciences, Medical Sciences, etc.
3. Researches dealing with the development of Official Statistics in Turkey and in the world.
4. Researches, dealing with the interpretation and analyses of the statistical data published with new scientific developments.

Principles of Publication

1. Researches are to be original, creative, fit in methodology and science and contribute to the existing application and theory. Publication language is Turkish.
2. The Journal is open to researches covering all the subjects in the field of statistics.
3. Researches approved by a three referee's mission are published. Unpublished articles are not given back to the author.
4. Articles are sent to the Secretariat of Journal in the forms of print out (4 copies) and magnetic (3,5' diskette). Researches that are accepted to be published are re-sent to the author(s) for correction. It is expected that at this stage the article is to be given the final form and not to be changed any more.
5. According to the Law No. 5846, TURKSTAT holds the copyrights of this publication. The Journal is not duplicated or distributed without authorisation.
6. Researches, which are not in conformity with the form of text preparation, copyrights and previously published or accepted to be published are given back to the author by Editor in Chief.
7. All of the correspondence is to be done with the Secretariat. Requests regarding to the subscription, preceding issues, offprint, advertisements and payments are submitted to the address of Request and Subscription.

ISSN: 1303 – 6319

Editor in Chief
Tel: +90 312 284 45 00/ 171
e-mail: fyildirim@cankaya.edu.tr
Assistant Editor
Tel: +90 312 410 03 75
e-mail: seviluygur@tuik.gov.tr
Journal of Statistics Research Secretary
Tel: +90 312 410 02 62 - 0312 410 01 10
Faks: +90 312 425 20 53
e.mail: dergi@tuik.gov.tr

Prime Ministry Republic of Turkey
Turkish Statistical Institute
Necatibey Cad. No: 114 06100 Yücetepe / ANKARA
Tel : +90 312 410 04 10 (Telephone exchange)
URL: http://www.tuik.gov.tr
Revolving Fund Management
Tel : +90 312 410 03 23-19
Faks: +90 312 417 58 86

Turkish Statistical Institute, Printing Division, Ankara, February- 2007
MTB: 2007-0158-320 Copies

Değerli Okuyucularımız,

Nihayet Dergi'mizin elinizdeki sayısını da tamamlayarak, sizlere ulaştırmış bulunuyoruz. Dergi'mizin yayın hayatına girdiği tarihten bu yana, edinmiş olduğumuz olumlu olumsuz deneyimlerin de ışığında, Dergi'mizde içinde bulunduğumuz yıldan itibaren içerik ve biçim yönünden değişikliklerin yapılması yönünde Başkanlık koordinatörlüğünde, çalışmalara başlanmış olduğu konusunda sizleri bilgilendirmek isterim. Dergi'de yapılacak değişikliklerden bazı başlıklar şunlardır: *Dergi'nin yılda üç sayı yerine iki sayı olarak yayınlanması; Dergi'de yayınlanacak makalelerin dilinin Türkçe ve/veya İngilizce olması ve özetlerin de ana makalenin diline bağlı olarak diğer dilde olması; Dergi Editörü'nün belirli bir süre için seçilmesi ve bir sonraki dönem Editörü'nün de aynı zamanda seçilmesi; biri TÜİK'ten diğeri akademik çevreden olmak üzere iki Editör Yardımcısı'nın görev alması; Editörler Kurulu üyelerinin de belirli bir süre için görev yapmaları; Dergi'de yayınlanmış tüm makalelere Dergi web sitesinden erişimin sağlanması ve makale tam metinlerinin indirilebilmesi; Dergi'ye yayınlanmak üzere gönderilmiş çalışmaların hangi değerlendirme aşamasında olduğunun dinamik ve güncel bir yapıda makale yazarları tarafından takip edilebilmesi; makale taslaklarının hakemler tarafından değerlendirilmesi sürecinde kullanılmak üzere, çok daha ayrıntılı bilgi içeren formların hazırlanmış olması; Dergi'de yayınlanmak üzere gönderilen makalelere ilişkin makale formatının yeniden düzenlenmesi ve değerlendirme işlemleri takviminin verilmesi; Dergi'nin özgün araştırma, gözden geçirme, teknik notlar, eleştirel derleme, tartışma ve güncel çevirilerden olmak üzere altı gruptaki makale çalışmalarına açık olması ... gibi* yenilikler. Yukarıda önemli olanlardan bazılarını belirtmiş olduğum tüm konuların açıklandığı, **Dergi Kılavuzu** ve **Uygulama Yönergesi** hazırlanmıştır. Yakın bir gelecekte sizlere bu konu hakkında daha geniş bilgilerin ulaştırılması amaçlanmaktadır.

Sevgili okuyucularımız, büyük emeklerle sizlere ulaştırılmış Dergi'mizin bu sayısında, yine başta desteğini hiç esirgemeyen TÜİK Başkanı Sayın Doç. Dr. Ömer DEMİR'e, makale çalışmaları ile Dergi'ye destek veren makale sahiplerine, özverili katkılarından dolayı hakemlerimize, Dergi Editör Yardımcısı Sayın TÜİK Uzmanı Sevil UYGUR'a, Editörler Kurulu üyelerine ve Dergi Sekreteryasında görev yapan arkadaşlarımıza ayrı ayrı teşekkür etmeyi bir borç bilirim.

Daha sonraki sayılarda buluşmak ümidiyle, saygılarımı sunarım.

Prof. Dr. Fetih YILDIRIM
Dergi Genel Editörü

İSTATİSTİK ARAŞTIRMA DERGİSİ'NİN 2005 YILI NİSAN, AĞUSTOS ve ARALIK SAYILARINA BİLİMSEL KATKI SAĞLAYAN HAKEMLER

Prof. Dr. Aydın ÖZTÜRK
Ege Üniversitesi

Dr. Barış SÜRÜCÜ
Orta Doğu Teknik Üniversitesi

Prof. Dr. Berna DENGİZ
Başkent Üniversitesi

Yrd. Doç. Dr. Canan DEMİRÜSTÜ
Osmangazi Üniversitesi

Prof. Dr. Cenap ERDEMİR
Hacettepe Üniversitesi

Prof. Dr. Dolun ÖKSOY
Ankara Üniversitesi

Prof. Dr. Ergun KARAAĞAOĞLU
Hacettepe Üniversitesi

Prof. Dr. Fatin SEZGİN
Bilkent Üniversitesi

Prof. Dr. Fetih YILDIRIM
Çankaya Üniversitesi

Prof. Dr. Fevzi KUTAY
Gazi Üniversitesi

Doç. Dr. Gül ERGÜN
Hacettepe Üniversitesi

Prof. Dr. Gülay BAŞARIR KIROĞLU
Mimar Sinan Güzel Sanatlar Üniversitesi

Prof. Dr. Gülser KÖKSAL
Orta Doğu Teknik Üniversitesi

Prof. Dr. Hamza EROL
Çukurova Üniversitesi

Prof. Dr. Hasan BAL
Gazi Üniversitesi

Prof. Dr. Hülya BAYRAK
Gazi Üniversitesi

Prof. Dr. Hülya ÇINGI
Hacettepe Üniversitesi

Prof. Dr. Hüseyin TATLIDİL
Hacettepe Üniversitesi

Doç. Prof. Dr. İhsan ALP
Gazi Üniversitesi

Yrd. Doç. Dr. İhsan KARABULUT
Ankara Üniversitesi

Yrd. Doç. Dr. İlknur ÖZMEN
Başkent Üniversitesi

Doç. Dr. İrini DİMİTRİYADİS
Bahçeşehir Üniversitesi

Prof. Dr. İsmail ERDEM
Başkent Üniversitesi

Doç. Dr. İsmet DOĞAN
Afyon Kocatepe Üniversitesi

Prof. Dr. Kazım ÖZDAMAR
Osmangazi Üniversitesi

Doç. Dr. Kıvılcım METİN ÖZCAN
Bilkent Üniversitesi

Prof. Dr. M. Aydın ERAR
Mimar Sinan Güzel Sanatlar Üniversitesi

Yrd. Doç. Dr. Mehmet UYSAL
Hacettepe Üniversitesi

İSTATİSTİK ARAŞTIRMA DERGİSİ'NİN 2005 YILI NİSAN, AĞUSTOS ve ARALIK SAYILARINA BİLİMSEL KATKI SAĞLAYAN HAKEMLER (DEVAM)

Prof. Dr. Olcay ARSLAN

Çukurova Üniversitesi

Doç. Dr. Osman SARAÇBAŞI

Hacettepe Üniversitesi

Dr. Özlem TÜRKER

Çankaya Üniversitesi

Dr. Özlem İLK

Orta Doğu Teknik Üniversitesi

Prof. Dr. Refik BURGUT

Çukurova Üniversitesi

Prof. Dr. Refik SOYER

George Washington University

Prof. Dr. Sadullah SAKALLIOĞLU

Çukurova Üniversitesi

Prof. Dr. Sahamet BÜLBÜL

Marmara Üniversitesi

Doç. Dr. Sevil BACANLI

Hacettepe Üniversitesi

Prof. Dr. Soner GÖNEN

Gazi Üniversitesi

Yrd. Doç. Dr. Suat KASAP

Hacettepe Üniversitesi

Prof. Dr. Süleyman GÜNAY

Hacettepe Üniversitesi

Prof. Dr. T. Erkan TÜRE

International Saray Bosna University

Doç. Dr. Tülay SARAÇBAŞI

Hacettepe Üniversitesi

Doç. Dr. Utku UTKULU

Dokuz Eylül Üniversitesi

Prof. Dr. Ümit ŞENESEN

İstanbul Teknik Üniversitesi

Doç. Dr. Yılmaz AKDİ

Ankara Üniversitesi

Yrd. Doç. Dr. Yüksel TERZİ

Afyon Kocatepe Üniversitesi

Prof. Dr. Zehra MULUK

Başkent Üniversitesi

MARJİNAL ANALİZ YOLUYLA GÜVENİLİRLİK OPTİMİZASYONU

Ümit YÜCEER*

ÖZET

Marjinal analizin güvenilirlik redandansı optimizasyonuna bir uyarlaması anlatılmaktadır. Amaç fonksiyonu ayrışık hale dönüştürülüp, konkav olduğu gösterilecek ve marjinal analiz metodunun çok seri ve hızlı bir şekilde, optimal olmasa bile, optimala çok yakın çözümler türettiği bir örnekle gösterilecektir.

Anahtar Kelimeler *: Ayrık Konkav, Kaynak Tahsisi Problemleri, Marjinal Tahsis Algoritması.*

1. GİRİŞ

Çözülmek istenen problem, matematiksel olarak aşağıda görüldüğü gibi ifade edilebilir.

$$\max \; R(x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} (1 - q_j^{x_j+1}) \tag{1}$$

kısıtlar

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i \quad i=1,2,\ldots,m \tag{2}$$

$$x_j \geq 0, \text{ tamsayı, her bir } j=1,2,\ldots,n$$

Amaç fonksiyonu (1), sistemin güvenilirliğini gösterir ve $p_j = 1 - q_j$ teriminde her bir $j=1,2,\ldots,n$ elemanının güvenilirlik olasılığıdır. Özellikle Pages ve Gondran (1986) bu çeşit güvenilirlik optimizasyonu problemlerini tartışmaktadırlar. Daha genel anlamda, kaynak tahsisi problemleri Ibaraki ve Katoh (1988) tarafından tanıtılmıştır. Kaynak tahsisi problemlerinin daha genel bir sınıflandırması ve uygun çözüm teknikleri, Katoh ve Ibaraki (1998) tarafından yapılmıştır. Değişkenlerin sayısının az olduğu bir problem, bütün geçerli kombinasyonları deneyerek çözülebilir. Fakat bu metot, orta veya daha büyük sayıda değişkenleri olan problemlere uygulanamaz. Eğer bütün p_j'ler yeterince büyük (1'e yakın)

* Doç. Dr. Çankaya Üniversitesi, Mimarlık Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Ankara, TÜRKİYE

olmaları durumunda, $R(x_1,x_2,\dots,x_n)$ fonksiyonu (1) yaklaşık olarak (Pages ve Gondran (1986)) şöyle ifade edilebilir: $R \approx 1-\sum_{j=1}^{n} q_j^{x_j+1}$. Daha iyi bir sonuç, amaç fonksiyonu (1)'in logaritmasını alarak elde edilebilir.

$$\max f(x_1,x_2,\dots,x_n)=\sum_{j=1}^{n}\log(1-q_j^{x_j+1}) \tag{3}$$

Kolaylık olması için, $f_j(x_j)=\log(1-q_j^{x_j+1})$ tanımlayalım. Böylece amaç fonksiyonu $f(x_1,x_2,\dots,x_n)=\sum_{j=1}^{n} f_j(x_j)$ ayrışık bir şekilde ifade edilmiş olur. Bir fonksiyonun, ayrışık olma özelliği, çok güçlü bir özelliktir (Yüceer (2002)).

2. AYRIK KONVEKSLİK KAVRAMI

D herhangi ayrık bir küme olsun. Verilen bir noktanın (Z) komşuluğu $\{W \in D : \|W - Z\| < 1\}$ olarak tanımlanır. Bu tanımda Z noktasının D içinde olması aranmaz. Eğer $Z \in D$ ise $N(Z)=\{Z\}$ olur.

Tanım 1 Herhangi bir ayrık küme D üzerine tanımlanmış bir fonksiyon $\Phi(X)$, eğer bütün X,Y $\in$ D ve $\alpha \in (0,1)$ için aşağıdaki şartı (4) sağlıyorsa, $\Phi(X)$ ayrık konveks bir fonksiyondur.

$$\alpha\Phi(X)+(1-\alpha)\Phi(Y) \geq \min_{W\in N(Z)} \Phi(W) \tag{4}$$

Bu tanım, önemli bir özelliği simgeler ve yerel minimum, ayrık konvekslik kavramı altında global minimumdur (Miller (1971)). Daha genel kavramlar ve kuramsal gelişmeler Murota (2003) tarafından sunulmuştur. Ayrık D kümesini, genelliği bozmadan, n-boyutlu tam sayılar kümesi olarak alabiliriz.

Tanım 2 Eğer $-\Phi(X)$ bütün X $\in$ D için ayrık konveks bir fonksiyonsa, $\Phi(X)$ ayrık konkav bir fonksiyondur.

Tek değişkenli bir ayrık fonksiyonun ($\Phi(X)$), birinci ileri farkları şöyle gösterilebilir: $\Delta\Phi(x)=\Phi(x+1)-\Phi(x)$ ve ikinci ileri farklarıda $\Delta^2\Phi(x)=\Delta\Phi(x+1)-\Delta\Phi(x)$ olarak gösterilebilir.

Ayrık bir küme üstüne tanımlanmış bir fonksiyon için, klasik konvekslik tanımı birinci ileri farkların azalmayan (artan) bir dizi olması anlamına gelir. Başka bir deyişle, bu ikinci ileri farkların sıfır veya pozitif olması demektir. Bu sonuç, sürekli ve tek değişkenli bir fonksiyonun ikinci türevinin pozitif olmasıyla analog bir benzeyiştir. Klasik tanımının, ayrık konvekslik tanımıyla eşdeğerliliği Yüceer (2006) tarafından ispatlanmıştır.

Teorem 1 Amaç fonksiyonu (3) her bir *j=1,2,...,n* için terimi $f_j(x_j)$ ayrık konkavdır veya $-f_j(x_j)$ ayrık konvekstir.

İspat: Amaç fonksiyonu (3), ayrışık bir fonksiyon olduğu için, her bir j=1, 2,... , n için $f_j(x_j)$ fonksiyonun ayrık konkav veya $-f_j(x_j)$ ayrık konveks olduğunu göstermek yeterlidir. Bu sebeble, $f_j(x_j)$ fonksiyonunun birinci ileri farklarının monotonik azalan (artmayan) bir dizi oluşturduğunu veya ikinci ileri farklarının negatif veya sıfır olduğunu göstermemiz gerekir.

$$
\begin{aligned}
\Delta f_j(x_j) &= f_j(x_j+1) - f_j(x_j) \\
&= \log(1-q_j^{x_j+2}) - \log(1-q_j^{x_j+1}) \\
&= \log\left(\frac{1-q_j^{x_j+2}}{1-q_j^{x_j+1}}\right) \\
\Delta^2 f_j(x_j) &= \Delta f\,(x_j+1) - \Delta f_j(x_j) \\
&= \left(\frac{(1-q_j^{x_j+1})(1-q_j^{x_j+3})}{(1-q_j^{x_j+2})^2}\right)
\end{aligned}
\qquad (5)
$$

Öbür taraftan, bütün $0 \le q \le 1$ değerleri için $(1-q)^2 = 1-2q+q^2 \ge 0$ ve $1+q^2 \ge 2q$ sağlanır. Bu ilişkiyi kullanarak, yukarıda (5)'in paydasının $1-q_j^{x_j+1}-q_j^{x_j+3}+q_j^{2x_j+4}$, daima $1-2q_j^{x_j+2}+q_j^{2x_j+4}$ teriminden küçük veya eşit olduğu söylenebilir. Argümanin birden küçük veya eşit olması ise logaritmanın negatif veya sıfır olmasını gerektirir. Sonuç olarak, amaç fonksiyonunun konkav olduğunu söyleyebiliriz. Argümanımızı, bir adım daha ileri götürerek, aşağıda tanımlanan Lagrangian fonksiyonunun da ayrık konkav olduğunu gösterebiliriz.

Tanım 3 Her bir *i=1,2, ...,m* için, $\lambda_i \ge 0$ olsun.

$$
\Im(x_1,x_2,\ldots,x_n\lambda_1,\lambda_2,\ldots,\lambda_m) = \sum_{j=1}^{n} \log(1-q_j^{x_j+1}) + \sum_{i=1}^{m} \lambda_i \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right) \qquad (6)
$$

Lagrangian fonksiyonu da ayrışık bir fonksiyondur. Aşağıdaki tanım bu konuda çok yararlı olacaktır.

Tanım 4

$$\Im_j(x_j,\lambda_1,\lambda_2,\ldots,\lambda_m)=\log(1-q_j^{x_j+1})-\sum_{i=1}^{m}\lambda_i a_{ij}x_j \tag{7}$$

Böylece $\Im(x_1,x_2,\ldots x_n,\lambda_1,\lambda_2,\ldots,\lambda_m)=\sum_{j=1}^{n}\Im_j(x_j,\lambda_1,\lambda_2,\ldots,\lambda_m)+\sum_{i=1}^{m}\lambda_i b_i$

Teorem 2 $\Im(x_1,x_2,\ldots,x_n,\lambda_1,\lambda_2,\ldots,\lambda_m)$ ayrık konkav bir fonksiyondur.

İspat: Her bir *j=1,2,...,n* için $\Delta^2\Im_j(x_j,\lambda_1,\lambda_2,\ldots,\lambda_m)=\Delta^2 f_j(x_j)\leq 0$ olduğu açıkça görülebilir. Sonuç olarak, Lagrangian fonksiyonuda ayrık konkavdır.

Lagrangian fonksiyonunun konkav olması, marginal analiz algoritmasının uygulanabilmesine müsaade eder. (Fox (1966), Kao (1976) veya Yüceer (1999)).

3. TEK KISITLI KAYNAK TAHSİSİ PROBLEMİ

Verilen bir bütçe seviyesini aşmadan, güvenilirliği maksimum kılacak şekilde bütün elemanların sayısını (x_j) bulmamız gerekiyor. Bu durumda kısıtlar yeniden şu şekilde ifade edilecektir.

$$\sum_{j=1}^{n}a_j x_j \leq b \tag{8}$$

Görüldüğü üzere bu küçük problemin bir tek kısıtı vardır ve bu problem için marjinal analiz algoritması aşağıdaki gibi tarif edilebilir.

Başlangıç: Her j=1,2,...,n için $x_j=0$.

Adım 1. Her j=1,2,...,n için aşağıdaki oranı hesapla.

$$r_j=\frac{\log\left(\dfrac{1-q_j^{x_j+2}}{1-q_j^{x_j+1}}\right)}{a_j}$$

Adım 2. $\max_{1\leq j\leq n}\{r_j\}=r_{j^*}$. *Eğer* $\sum a_j x_j + a_{j^*}\leq b$, $x_{j^*}=x_{j^*}+1$ *ve Adım 1'e geri dön.*

Aksi takdirde optimale yakın bir çözüm elde edilmiştir.

Ayrıca, $f(x_1,x_2,\ldots,x_{j^*}+1,\ldots,x_n)$ gerçek optimum için bir üst sınır değeri verir. Optimala daha yakın bir değer, Adım 2 de küçük bir değişiklik yaparak elde edilebilir. Maksimum oranı bulma kuralı şu şekilde tanımlanabilir; $r_{j^*}=\max\left\{r_j:\sum_{i=1}^{m}a_ix_i+a_j\le b\right\}$.

4. GENEL KAYNAK TAHSİSİ PROBLEMLERİ

Verilen bir bütçe seviyesini, belirlenen bir ağırlığı, belirlenen bir hacmi, vesaire aşmadan, güvenilirliği maksimum kılacak şekilde her $j=1,2,\ldots,n$ için bütün elemanların sayısının bulunması isteniyor. Şimdiki problemde bir grup kısıt (2) vardır ve tamsayılar üstüne kurulu böyle bir problemi çözmekte sonsuz güçlükler yaratır. Pratik bir yaklaşımla, yeni bileşik bir kısıt aşağıda görüldüğü gibi elde edilebilir.

$$\sum_{j=1}^{n}\left(\sum_{i=1}^{m}\mu_i a_{ij}\right)x_j\le\sum_{i=1}^{m}\mu_i b_i \tag{9}$$

Bu yeni kısıttaki μ_i değerleri, $\sum_{i=1}^{m}\mu_i=1$ ve her $i=1,2,\ldots,m$ için $\mu_i\ge 0$. Bu değerler, esasen eşlenik değişkenlerdir (dual variables) ve kısıt grubu (2) doğrusal kısıtlar oldukları için, eşlenik problemi çözerek elde edilebilirler. Bu değiskenlerin Kuhn-Tucker şartlarını kullanarak nasıl hesaplanabileceği, örneğin Luenberger (1973) de anlatılmıştır. Genel problem için, Adım 1'deki oran aşağıdaki gibi (10) hesaplanacaktır.

$$r_j=\frac{\log\left(\dfrac{1-q_j^{x_j+2}}{1-q_j^{x_j+1}}\right)}{\sum_{i=1}^{m}\mu_i a_{ij}} \tag{10}$$

5. BİR ÖRNEK

Tek kısıtlı problemimiz, örnek uygulama olarak kullanılacaktır. Bu örnek problemimizde beş tane eleman bulunmaktadır. Bütçe Seviyeside diyelim ki 25000 YTL olsun. Gerekli bilgiler (a_j YTL cinsindendir) tabloda verilmiştir. Bu değerler tesadüfi seçilmiştir.

j	1	2	3	4	5
p_j	.75	.63	.81	.77	.72
a_j	2725	1535	2500	2015	1805

Marjinal analiz metodu X=(2,3,2,2,3) çözümünü türetir. Güvenilirlik ise R(2,3,2,2,3)=.94180 veya f(2,3,2,2,3)= -0.05996, ayrıca toplam maliyet de 24485 YTL'dir. Her adımda türetilen ara çözümler Tablo1'de verilmiştir. Bütün geçerli çözümleri deneyerek de aynı sonuç elde edilmiştir. Bu sadece bir tesadüftür. Ayrıca, optimal değer için bir üst sınırda elde edilmiştir; X=(2,4,2,2,3), ve R(2,4,2,2,3)=.95137, maliyeti de 26020 YTL'dir.

Tablo 1. Marjinal Analiz Çözümü

İterasyon	Oran Vektörü ($\times 10^{-5}$)	Çözüm (X)	R (X)
0	(8.19,20.51,6.96,10.27,13.71)	(0,0,0,0,0)	.21218
1	(8.19,6.20,6.96,10.27,13.71)	(0,1,0,0,0)	.29069
2	(8.19,6.20,6.96,10.27,3.30)	(0,1,0,0,1)	.37208
3	(8.19,6.20,6.96,2.09,3.30)	(0,1,0,1,1)	.45766
4	(1.79,6.20,6.96,2.09,3.30)	(1,1,0,1,1)	.57208
5	(1.79,6.20,1.20,2.09,3.30)	(1,1,1,1,1)	.68077
6	(1.79,2.15,1.20,2.09,3.30)	(1,2,1,1,1)	.74880
7	(1.79,2.15,1.20,2.09,0.89)	(1,2,1,1,2)	.79467
8	(1.79,0.78,1.20,2.09,0.89)	(1,3,1,1,2)	.82138
9	(1.79,0.78,1.20,0.47,0.89)	(1,3,1,2,2)	.85670
10	(0.43,0.78,1.20,0.47,0.89)	(2,3,1,2,2)	.89958
11	(0.43,0.78,0.22,0.47,0.89)	(2,3,2,2,2)	.92682
12	(0.43,0.78,0.22,0.47,0.25)	(2,3,2,2,3)	.94180

6. SONUÇLAR

Gerçek hayatta sık karşılanılan bir problem, hızlı ve seri bir şekilde marjinal analiz kullanarak çozülmüştür. Esasen bazı aksiyomlar yaparak veya yaklaşım teknikleri kullanarak (Pages and Gondran (1986)), optimal çözümü bulmak mümkündür. Burada anlatılan metot herhangi bir aksiyom veya yaklaşım tekniği kullanmamıştır.

Küçük örneğimizin incelenmesi, marjinal analiz metodunun çalışması, hızı, türettiği çözümlerin kalitesi hakkında bir bilgi vermektedir. Daha geniş kapsamlı bilgi Yüceer (1999) tarafından verilmiştir. Pratisyenler açısından, her $j=1,2,...,n$ için r_j oranı, birim başına güvenilirlik oranının katkısı olarak düşünülebilir. Bilhassa, pratikte bu oranın anlamı ve kullanılması oldukça değerlidir.

KAYNAKLAR

FOX, B. (1966), *Discrete Optimization via Marginal Analysis*, Management Science, 13, 210-216.

IBARAKI, T., KATOH, N. (1988), *Resource Allocation Problems*, The MIT Press.

KAO, E. (1976), *On Incremental Analysis in Resource Allocations, OR Quarterly*, 27, 759-763.

KATOH, N., IBARAKI, T. (1998), *Resource Allocation Problems*, D.--Z. Du, P.M. Pardalos, eds., Handbook of Combinatorial Optimization, Kluwer Academic, Boston, 159-260.

LUENBERGER, D.G. (1973), *Introduction to Linear and Nonlinear Programming*, Addison-Wesley.

MILLER, B.L. (1971), *On Minimizing Nonseparable Functions Defined on Integers with an Inventory Application, SIAM Journal of Applied Mathematics*, 21/1, 166-185.

MUROTA, K. (2003), *Discrete Convex Analysis*, SIAM.

PAGES, A., GONDRAN, M. (1986), *System Reliability*, Springer-Verlag.

YÜCEER, Ü. (1999), *Marginal Allocation Algorithm for Nonseparable Functions, INFOR*, 37, 97-113.

YÜCEER, Ü. (2002), *Discrete Convexity: Convexity for Functions Defined on Discrete Spaces,* Discrete Applied Mathematics, 119/3, 299-306.

YÜCEER, Ü. (2006), *The Equivalence of Discrete Convexity and the Classical Definition of Convexity,* International Mathematical Forum, 1/5, 299-308.

REDUNDANCY OPTIMIZATION VIA MARGINAL ANALYSIS

ABSTRACT

An application of marginal analysis to redundancy optimization is presented. The objective function is transformed into a separable function and its discrete concavity is proved. Consequently, the marginal allocation algorithm obtains near optimal solutions very rapidly. An example illustrates the method.

Key Words: *Discrete Concavity, Marginal Allocation Algorithm, Resource Allocation Problems.*

İstatistik Araştırma Dergisi Cilt 04 No 03 Sayfa 9-29 Aralık 2005

DÖRT FARKLI FAKTÖR ANALİZİ YÖNTEMİNİN BİR ÖRNEK ÜZERİNDE KARŞILAŞTIRILMASI

Seval SÜZÜLMÜŞ* Sadullah SAKALLIOĞLU**

ÖZET

Faktör analizi modelindeki faktör sayısını belirlemek, Λ ve Ψ parametrelerini tahmin etmek ve faktör skorlarını oluşturmak için çeşitli yöntemler vardır. Bunlar arasında yaygın olarak kullanılanları Temel Bileşenler Analizi; En Çok Olabilirlik; Ağırlıklandırılmamış En Küçük Kareler ve Genelleştirilmiş En Küçük Kareler yöntemleridir. Bu çalışmada bu yöntemler hakkında kısaca bilgi verilerek, Türkiye İstatisitk Kurumu(TÜİK)'ten alınan, Aralık 2003 yılına ait Türkiye'nin çeşitli bölgelerinde tüketilen bazı gıda maddelerinin fiyatlarına ait verilerin SPSS 12.0 paket programı kullanılarak, analizi yapıldı. Amaç, gıda fiyatlarıyla ilgili yapılacak olan bir araştırmada fazla bilgi kaybı olmadan daha az sayıda olan hangi gıda maddelerinin seçilmesi gerektiğini belirlemektir. Yapılan analizde 4 farklı faktör çıkarma yöntemi ve her bir yöntem için 2 farklı döndürme yöntemi kullanıldı. Elde edilen sonuçlar karşılaştırılarak yöntemler arasındaki farklar incelendi.

Anahtar Kelimeler: *Ağırlıklandırılmamış En Küçük Kareler, En Çok Olabilirlik, Faktör Analizi, Genelleştirilmiş En Küçük Kareler, Quartimax, Temel Bileşenler Analizi, Varimax.*

1. GİRİŞ

$x = \Lambda f + e$ şeklinde gösterilen faktör analizi modelinde x'in varyans-kovaryans matrisi $\Sigma = \Lambda\Lambda' + \Psi$ şeklindedir. Burada; $x : px1$ tipinde gözlenebilir rastgele değişkenlerin vektörü; $f : kx1$ tipinde ortak faktörler olarak adlandırılan, gözlenemeyen değişkenlerin vektörü; $e : px1$ tipinde gözlenemeyen değişkenlerin vektörü (hataların vektörü); Λ : $[\lambda_{ir}]_{pxk}$ faktör yükleri olarak adlandırılan bilinmeyen sabitlerin matrisi; Ψ ise pxp tipinde, hatalara ait varyans-kovaryans matrisi olup, köşegenindeki elemanları ψ_1 , ψ_2,....., ψ_p olan köşegen matristir.

* Çukurova Üniversitesi, Osmaniye Meslek Yüksekokulu, Adana, TÜRKİYE
suzulmus@mail.cu.edu.tr
** Çukurova Üniversitesi, Fen-Edebiyat Fakültesi İstatistik Bölümü Adana, TÜRKİYE
sadullah@mail.cu.edu.tr

Faktör analizi, aralarında yüksek korelasyon bulunan değişkenleri bir araya getirerek daha az sayıda temel bileşenler ya da faktörler olarak adlandırılan yeni değişkenler bulmayı amaçlar. Faktör analizi, psikoloji ile başlamış olup, günümüzde başta Sosyal Bilimler olmak üzere ekonomi, botanik, biyoloji, ziraat, tıp gibi uygulamalı bilim dallarında yaygın olarak kullanılan çok değişkenli istatistik analiz yöntemlerinden biridir.

Faktör analizi ile ilgili ilk çalışmalar 20. yüzyılın başında Spearman, Karl Pearson, Thomson, Thurstone ve Burt tarafından yapılmıştır. Kovaryans ya da korelasyon matrislerinin yapısı analiz edilirken, iki yaklaşım söz konusudur. Bunlardan en iyi bilineni temel eksenler yöntemini geliştiren Pearson (1901)'ı takiben, Hotelling (1933) bu yöntemi Temel Bileşenler Analizine genişletmiş ve Spearman (1904, 1926) faktör analizi kavramını geliştirmiştir. Çoklu faktör kavramına Garnett (1919) ile girilmiş fakat bu kavramın gelişimi 1930 ve 1940'lı yıllarda Thurstone tarafından gerçekleştirilmiştir. Çoklu faktör analizi terimini ortaya atan Thurstone (1931) daha sonra basit yapı olarak da bilinen Merkezi Faktör Rotasyonu kavramını geliştirmiştir (Darton, 1980).

Bu çalışmanın ikinci bölümünde faktör çıkarma yöntemleri ve faktör döndürmesi yöntemleri hakkında bilgi verilmiştir. Üçüncü bölümde ise TÜİK'ten alınan, Aralık 2003 yılında Türkiye'nin çeşitli bölgelerinde tüketilen bazı gıda maddelerinin fiyatlarına ait verilerin SPSS 12.0 paket programı kullanılarak, 4 yöntem ile faktör analizi yapılmış ve elde edilen sonuçlar karşılaştırılmıştır.

2. YÖNTEMLER

2.1 Temel Bileşenler Analizi Yöntemi

p tane değişkeni

$$x' = (x_1, x_2, \ldots, x_p)$$

vektörü ile gösterelim. Değişkenlerin sadece varyans ve kovaryansları ile ilgilendiğimizden $x_1, x_2, \ldots, x_p$'lerin her birinin ortalaması sıfır kabul edilebilir. Kovaryans matrisi Σ'nın $\delta_1, \delta_2, \ldots, \delta_p$ özdeğerlerinin farklı ve azalan sırada düzenlendiğini kabul edelim. Kovaryans matrisi;

$$\Sigma = \Gamma \Delta \Gamma'$$

olarak yazılabilir. Δ, köşegeninde $\delta_1, \delta_2, \ldots, \delta_p$ olan köşegen matris ve Γ pxp tipinde ortogonal bir matristir. Γ'nın i. sütunu δ_i özdeğerlerine karşılık gelen özvektör olup, Σ'nın özdeğerleri farklı olduğundan, Γ tek olarak tanımlıdır.

Temel Bileşenler cebirsel olarak x_i değişkenlerinin doğrusal bileşeni olarak ifade edilir. Yeni $y_1, y_2, \ldots, y_p$ değişkenlerini $y = \Gamma' x$ eşitliği ile tanımlayalım. Bu durumda y'ler ilişkisiz ve y_i'nin varyansı δ_i'dir. Yani; X_{nxp} veri matrisine uygun dönüşüm yapılarak, birbirleri ile ilişkisiz kolonlardan oluşan bir veri kümesi elde edilmiş olur. Temel bileşenler analizinde x_i değişkenlerine ait toplam varyans p-tane y_i değişkenleri (temel bileşenler) bulunduğunda açıklanabilir. Yeni $y_1, y_2, ..., y_p$ değişkenlerini $x_1, x_2, ..., x_p$ değişkenlerinin lineer kombinasyonları olarak yazıp, her bir y_i'nin varyansı sırasıyla maksimum yapılmaktadır (Chatfield ve Collins, 1980). Eğer verideki toplam varyansın büyük bir miktarı daha az sayıda temel bileşen tarafından açıklanıyorsa, o zaman yorum yaparken araştırmacı, bu az sayıdaki temel bileşenleri p orijinal değişkenler yerine kullanır. Böylece temel bileşenler analizi veri indirgeme tekniği olarak da kullanılır (Sharma, 1996).

2.2 Ağırlıklandırılmamış En Küçük Kareler Yöntemi

Bu yöntemde

$$U = tr\left[(S - \Sigma)^2\right]$$

fonksiyonu minimum yapılır. Burada S örneklem varyans-kovaryans matrisidir (Jöreskog ve Goldberger, 1972).

$$x = \mu + \Lambda f + e$$

faktör modeli için μ ortalama vektörü, Λ. faktör yükleri ve Ψ varyanslarının bilindiğini kabul edelim. Hatalar olarak

$$e' = \lfloor e_1, e_2, ..., e_p \rfloor$$

faktörlerini kabul edelim.

$i = 1, 2, ..., p$ için $var(e_i) = \Psi_i$'lerin eşit olmadığı durumlarda ortak faktör değerlerini tahmin etmek için ağırlıklı en küçük kareler yöntemi kullanılmaktadır. Özel faktörlerin kendi varyanslarının tersleriyle ağırlıklandırılmış olan kareler toplamı

$$\sum_{i=1}^{p} e_i^2 / \psi_i = (x - \mu - \Lambda f)' \Psi^{-1} (x - \mu - \Lambda f)$$

şeklindedir. Bu kareler toplamını, en küçük kareler yöntemine göre minimum yapan f değeri

$$\hat{f} = \left(\hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda}\right)^{-1} \hat{\Lambda}' \hat{\Psi}^{-1} (x - \mu)$$

dır. i. inci durum için faktör skorları,

$$\hat{f}_i = \left(\hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda}\right)^{-1} \hat{\Lambda}' \hat{\Psi}^{-1} (x_i - \bar{x}) \qquad i = 1, 2, ..., p$$

olarak ele edilir (Johnson ve Wichern, 2002).

2.3 Genelleştirilmiş En Küçük Kareler Yöntemi

Bu yöntemde S ve Σ arasındaki farklar S^{-1}'in elemanlarıyla ağırlıklandırılarak

$$G = tr\left[(I - S^{-1}\Sigma)^2\right]$$

fonksiyonu minimum yapılır (Jöreskog ve Goldberger, 1972).

2.4 En Çok Olabilirlik Yöntemi

Bu yöntemde f, e ve x vektörlerinin elemanlarının birbirinden bağımsız çok değişkenli normal dağılımlı olduğu ve ortak faktör sayısı k'nın bilindiği varsayılır.

x değişkenlerinin n $(n > p)$ birimlik rastgele örneklemi elde edilsin. x gözlemlerini x_α $(\alpha = 1, 2, ..., n)$ sütun vektörleriyle gösterelim. O zaman örneklem kovaryans matrisi $S = \lfloor s_{ij} \rfloor$;

$$S = \frac{1}{n-1} \sum_{\alpha} (x_\alpha - \bar{x})(x_\alpha - \bar{x})'$$

olarak tanımlanır. Örneklem ortalama vektörü $\bar{x} = (1/n) \sum_{\alpha} x_\alpha$'dir. Burada S, Σ'nın bir yansız tahmin edicisidir. L örneklemin olabilirlik fonksiyonu olmak üzere, bu fonksiyonun logaritması,

$$ln\, L = -\frac{1}{2} n\, ln|\Sigma| - \frac{1}{2} n \sum_{i,j} s_{ij} \sigma^{ij}$$

dir. Burada σ^{ij}, Σ^{-1}'in i. satır ve j. inci sütunundaki elemanıdır. Ayrıca logaritmik olabilirlik fonksiyonu $\ln L$,

$$\ln L = -(1/2)n\left[\ln|\Sigma| + tr(S\Sigma^{-1})\right]$$

şeklinde de yazılabilir. Olabilirlik fonksiyonu Σ' nın dolayısıyla Λ ve Ψ'nin bir fonksiyonudur. Buradan; $\ln L$'yi maksimum yapan $\hat{\Lambda}$ ve $\hat{\Psi}$ değerleri bulunur (Lawley and Maxwell, 1971). En çok olabilirlik yöntemi

$$M = tr(\Sigma^{-1}S) - ln|\Sigma^{-1}S| - p$$

fonksiyonunun minimum yapılması olarak da bilinir.

2.5 Faktör Döndürmesi

Faktörlerin elde edilmesinden bir sonraki adım; faktörleri daha iyi yorumlayabilmek için "döndürme" yapılmasıdır. Faktör döndürmesinde iki yöntem kullanılmaktadır. Bunlardan birincisi eksenler dik olacak şekildeki döndürmedir. Buna "dik döndürme" adı verilir. İkinci yöntemde ise her faktör birbirinden bağımsız olarak döndürülür. "Eğik döndürme" adı verilen bu yöntemde eksenlerin birbirlerine dik olması gerekli değildir, eğik döndürme farklı açılarla yapılmaktadır. Sonuç olarak, iki döndürme yöntemi arasındaki en önemli istatistiksel farklılık; dik döndürmede faktörler ilişkisiz (dik bağımsız) iken, eğik döndürmede ilişkilidir. Bu çalışmadaki yapılan uygulamada dik döndürme yöntemlerinden Varimax ve Quartimax yöntemleri kullanıldığından, bu yöntemler hakkında bilgi verilecektir.

Quartimax yönteminin amacı, orijinal Λ faktör yüklerinin karelerinin varyansı maksimum olan yeni bir D faktör matrisinin oluşması için dik (ortogonal) dönüşüme karar vermektir. p değişkenli k faktörlü faktör modelinde, faktör yüklerinin karelerinin varyansı

$$Q = Var(DD') = (1/pk)\sum_{i=1}^{p}\sum_{j=1}^{k}\left(d_{ij}^{2} - \bar{d}^{2}\right)^{2} \qquad (1)$$

dır. Burada faktör yüklerinin karelerinin ortalaması

$$\bar{d}^{2} = (1/pk)\sum_{i=1}^{p}\sum_{j=1}^{k} d_{ij}^{2}$$

dır. Böylece (1) eşitliği

$$Q = (1/pk)\sum_{i=1}^{p}\sum_{j=1}^{k} d_{ij}^{4} - \left(\bar{d}^{2}\right)^{2}$$

şekline gelir (Harman, 1968).

Varimax yöntemi, quartimax yöntemine benzerdir. Varimax yönteminde de faktör varyansları en büyük olacak şekilde döndürme yapılır (Kaiser, 1958).

$$D = (d_{ij}) = \Lambda\Gamma$$

eşitliğinde, Γ_{kxk} ortogonal bir matris ($\Gamma\Gamma' = I_k$), D dik dönüştürülmüş yüklerin matrisi,

ve $d_j = \sum_{i=1}^{p} d_{ij}^{2}$ $j = 1, 2, \ldots\ldots, k$ olmak üzere

$$\sum_{j=1}^{k}\sum_{i=1}^{p}(d_{ij}^{2} - p^{-1}d_j)^{2} \qquad (2)$$

eşitliği maksimum yapılır. Böyle bir yöntem, D'nin sütunlarını ya büyük (mutlak değerce) değerli ya da 0 değerli yapmaya çalışır. Böylece, yöntem değişkenlerle güçlü ilişkileri veren faktörleri ya da hiç ilişkili olmayan faktörleri verir. Varimax yöntemi; faktör yük matrisinin her bir kolonunun normalize edilmesiyle elde edilen yüklerin, varyanslarının toplamının maksimum yapılması olarak da tanımlanır ve (2) eşitliğini maksimum yapma yerine, $d_{ij}^{*} = d_{ij}/h_i$ olmak üzere,

$$V = (1/p^{2})\sum_{j=1}^{k}\left[p\sum_{i=1}^{p} d_{ij}^{*4} - \left(\sum_{i=1}^{p} d_{ij}^{*2}\right)^{2}\right]$$

fonksiyonu maksimum yapılır. Burada

$$h_i^2 = \sum_{j=1}^{k} d_{ij}^2 = \sum_{j=1}^{k} \lambda_{ij}^2$$

şeklinde olup, *i*. değişkenle tüm *k* faktörleri arasındaki ortak varyansı göstermektedir. Bu ortak varyans dik dönüşüm altında değişmez. Burada, λ_{ij} dönüştürülmemiş faktör yüklerini göstermektedir (Srivastava, 2002).

3. UYGULAMA

Çalışmanın bu bölümünde Türkiye genelinde 20 farklı ilde (Bursa, İstanbul, Kocaeli, Denizli, İzmir, Adana, Antalya, İçel, Ankara, Eskişehir, Kayseri, Konya, Samsun, Trabzon, Zonguldak, Erzurum, Malatya, Diyarbakır, Gaziantep, Türkiye) tüketilen bazı gıda çeşitlerinin fiyatları kullanılmıştır. Analizler SPSS 12.0 paket programı kullanılarak yapılmıştır. Faktör analizi; faktör çıkarma yöntemlerinden "Temel Bileşenler Analizi (TBA)"; "En Çok Olabilirlik (EÇO)"; "Genelleştirilmiş En Küçük Kareler Yöntemi (GEKK)" ve "Ağırlıklandırılmamış En Küçük Kareler Yöntemi (AEKK)"; faktör döndürme yöntemlerinden ise "Varimax" ve "Quartimax" yöntemi seçilerek yapılmıştır.

Tablo 1. Analizde Kullanılan Değişkenlerin Listesi

Değişken Adı	**Notasyon**
Normal Ekmek	nrmekmek
Buğday Unu	bugdayun
Pirinç Unu	pirinunu
Pirinç 1.Bersani	pbersani
Pirinç 2.Baldo	prcbaldo
Makarna	makarna
Bulgur	bulgur
Yufka	yufka
Şehriye	sehriye
Bisküvi 1.Sade	bis1sade
Bisküvi 2.Bebe	bis2bebe
Yaş Pasta	yaspasta
Kuru Pasta	kurupast
Hamur Tatlısı 1. Baklava	baklava
Hamur Tatlısı 2. Tulumba	tulumba

Kaiser-Meyer-Olkin (KMO); örneklem uygunluğu ölçütünü incelediğimizde nrmekmek çıkarıldıktan sonra KMO=0,613 olarak elde edilmiş, bu değer yaklaşık % 60 değerine karşılık gelip, faktör analizini uygulamak için orta derece olarak yorumlanır.

Korelasyon matrisinin birim matris olup olmadığını test etmek için kullanılan Bartlett Küresellik Testi sonucunu incelediğimizde, ki-kare değerinin 225,497 olması ve p-değerinin 0,000 olmasından dolayı korelasyon matrisinin birim matris olmadığı hipotezi kabul edilmiştir. Buradan korelasyon matrisinin faktörleştirilebilir olduğu sonucuna varılmıştır.

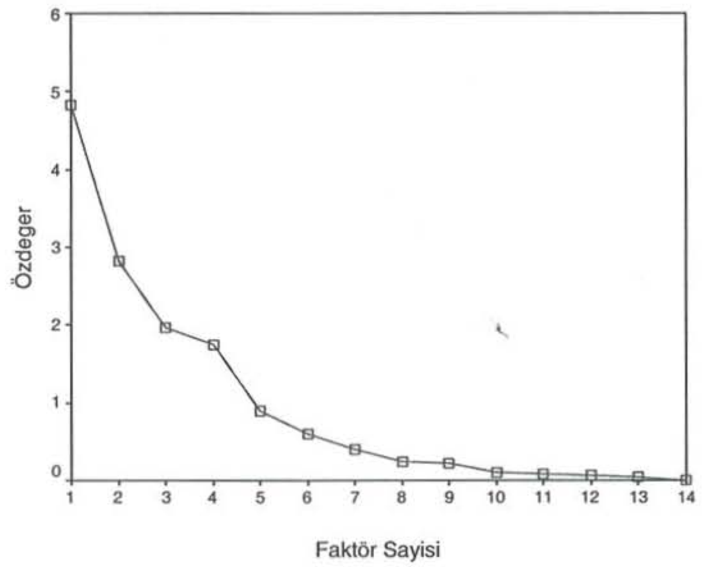
Tablo 2. Türkiye'de 20 Farklı Bölgede Tüketilen Gıda Fiyatlarının Ortalama ve Standart Sapma Değerleri

	Ortalama	Standart Sapma	N
BUGDAYUN	1,085289	6,7523E-02	20
PIRINUNU	3,359843	,43388874	20
PBERSANI	1,731266	,19317339	20
PRCBALDO	2,421677	,21125453	20
MAKARNA	1,088025	6,3607E-02	20
BULGUR	1,145427	,12143882	20
YUFKA	1,825863	,28902566	20
SEHRIYE	1,085755	6,2493E-02	20
BIS1SADE	3,512285	,12126564	20
BIS2BEBE	4,928129	,41376717	20
YASPASTA	9,454155	2,96461701	20
KURUPAST	6,612499	1,46198414	20
BAKLAVA	8,458560	2,22362367	20
TULUMBA	5,231382	1,20376497	20

Tablo 3. KMO ve Bartlett Ölçümü

Kaiser-Meyer-Olkin Ölçümü	,613
Bartlett Küresellik Yaklaşık ki-kare Değeri Testi	225,497
serbestlik derecesi	91
p-değeri	,000

Anti-imaj korelasyon matrisinin köşegen elemanları aynı zamanda örneklem yeterliliğinin bir göstergesidir. Örneklem yeterlik değerleri küçük olan değişkenler analizden çıkarılabilir. Eğer korelasyon matrisi faktörleştirilebilir ise anti imaj matrisinin köşegen elemanları dışındaki elemanlar oldukça küçük değerli olacaktır (Hair vd, 1998). Anti-İmage Korelasyon matrisinin ana köşegeni üzerinde yer alan MSA_i değerleri büyük olduğundan, veri kümesinin analiz için uygun olduğu sonucuna varılmıştır. Değişkenlerin ortak faktör içermeleri için, diğer değişkenlerin doğrusal etkileri göz ardı edildiğinde, değişken çiftleri arasındaki kısmi korelasyon katsayılarının küçük olmaları gerekmektedir (George ve Mallery, 2003). NRMEKMEK değişkeninin diğer değişkenlerle kısmi korelasyonlarının küçük olduğu gözlendiğinden, bu değişkenin analizden çıkarılmasının pek bir fark yaratmayacağı sonucuna varılmıştır. Bu nedenle NRMEKMEK çıkarılıp 14 değişken kullanılarak, analiz yapılmıştır. Normal ekmeğin analizden çıkarılmasıyla elde edilen ortalama ve standart sapma değerleri Tablo 2'de verilmiştir.



Şekil 1. Çizgi Grafiği

Şekil 1'deki Çizgi Grafiğinden ilk 4 özdeğerden sonra, eğimimiz düzleştiğinden analize ilk 4 faktörle devam edilmiştir.

3.1 Temel Bileşenler Analizi (TBA) Yöntemi

Tablo 4. TBA Yöntemi Uygulandığında Ortak Faktör Varyansları

	Başlangıç Değeri	Ortak Faktör Varyansı
BUGDAYUN	1,000	,858
PIRINUNU	1,000	,709
PBERSANI	1,000	,839
PRCBALDO	1,000	,914
MAKARNA	1,000	,869
BULGUR	1,000	,716
YUFKA	1,000	,844
SEHRIYE	1,000	,866
BIS1SADE	1,000	,784
BIS2BEBE	1,000	,710
YASPASTA	1,000	,887
KURUPAST	1,000	,860
BAKLAVA	1,000	,654
TULUMBA	1,000	,863

Ortak Faktör Varyans (OFV)'ları, analize dahil edilen her bir değişkene ait varyansın ortak faktörler tarafından açıklanma miktarını gösterir. Faktör çıkarma yöntemlerinden temel bileşenler analizi kullanılarak elde edilen, ortak faktör varyansları Tablo 4'te verilmiştir. NRMEKMEK değişkeni analizdeyken OFV'ları %25.2 ile %91.8 arasında; NRMEKEMEK çıkarıldıktan sonra ise %65.4 ile %91.4 arasında değiştiğinden, buradan da NRMEKMEK değişkeninin analizden çıkarılmasının uygun olduğu yorumu yapılmıştır.

TBA kullanılarak elde edilen dört faktöre göre yapılan çözümlemede özdeğerler, varyans açıklama yüzdeleri ve birikimli varyans Tablo5'te verilmiştir.

Tablo 5. TBA Uygulandığında Özdeğerler ve Toplam Açıklanan Varyans Miktarları

	Faktörleştirme Sonrası Yüklerin Kareleri Toplamı			Faktör Döndürmesi Sonrası Yüklerin Kareleri Toplamı		
Faktörler	Toplam	Varyans %	Birikimli %	Toplam	Varyans %	Birikimli %
1	4,832	34,513	34,513	3,118	22,274	22,274
2	2,824	20,171	54,684	2,897	20,690	42,964
3	1,973	14,093	68,777	2,870	20,499	63,464
4	1,745	12,464	**81,240**	2,489	17,776	**81,240**

Tablo 5'te "Başlangıç Değerler" sütununda, başlangıç özdeğerlerine bağlı olarak çıkarılan faktörler hakkında bilgi verilmektedir. Başlangıç çözümü için değişken sayısı kadar faktör vardır. İlk temel bileşen tüm değişkenlerdeki maksimum varyansı açıklar. İlk özdeğer varyansın %34.513'lik kısmını, ikinci özdeğer %20.171, üçüncü özdeğer %14.093 ve dördüncü özdeğer ise %12.464'lük kısmını açıklamaktadır. Böylece dört temel bileşen orijinal değişkenlerdeki varyansın %81.240'lık kısmını açıklamaktadır. Benzer şekilde devam edildiğinde temel bileşenlerin varyans açıklama yüzdeleri giderek azalacaktır. Faktör sayısı belirleme kriteri olarak 1'den büyük özdeğer sayısının 4 olduğu da göz önüne alınarak, ilk 4 faktöre ait bilgiler verilmiştir. "Faktör Çıkarma Sonrası Yüklerin Kareleri Toplamı" sütununda, çıkarılan faktörler hakkında bilgi verilmektedir. "Faktör Döndürmesi Sonrası Yüklerin Kareleri Toplamı" sütununda yer alan özdeğerler ve varyans açıklama yüzdeleri ise önceki sütunlarda hesaplanan değerlerden farklıdır. Fakat faktör kümesi için birikimli varyans yüzdesi aynı kalmıştır. Döndürme yapıldıktan sonra faktörlerin sayısı ve toplam varyansı açıklama oranı değişmeyip, her bir faktörün bireysel olarak açıkladığı oranda farklılık olmuştur. Faktörleştirme işleminden sonra, analize kaç faktör ile devam edileceğine karar verebilmek için, açıklanan toplam varyans miktarı oldukça önemli bir göstergedir. Varyans açıklama yüzdesi Tablo 5'te 81.240'tır. Bir başka ifadeyle, %19 gibi bir bilgi kaybıyla değişkenler açıklanmıştır.

Tablo 6. TBA Yöntemi Uygulandığında Rotasyonsuz Faktör Matrisi

	Faktörler			
	1. Faktör	**2. Faktör**	**3. Faktör**	**4. Faktör**
BULGUR	**0,745**	-0,237	0,259	0,197
PIRINUNU	**0,705**	0,274	-0,272	-0,251
PRCBALDO	**0,704**	-0,419	-0,367	0,328
PBERSANI	**0,675**	-0,32	-0,161	**0,505**
BAKLAVA	**0,67**	-0,126	**-0,433**	-0,039
KURUPAST	**0,616**	**-0,602**	0,229	-0,256
MAKARNA	**0,596**	**0,55**	0,224	0,402
SEHRIYE	**0,588**	**0,537**	0,288	0,385
BIS1SADE	**0,575**	**0,536**	0,385	-0,133
TULUMBA	**0,574**	-0,227	**0,573**	-0,393
YASPASTA	**0,538**	**-0,66**	-0,065	-0,397
YUFKA	-0,243	-0,341	**0,615**	**0,539**
BUGDAYUN	**0,443**	**0,574**	**-0,575**	-0,029
BIS2BEBE	0,308	**0,474**	0,329	**-0,532**

Tablo 6'da TBA yöntemi için verilen Rotasyonsuz Faktör Matrisi, faktörlerin her bir değişken üzerindeki faktör yüklerini göstermektedir. Bu katsayılar, her bir faktöre ne kadar ağırlık düştüğünü gösterdiklerinden "Faktör Yükleri" adını alırlar. Tahmin edilen faktörler birbirleri ile ilişkisiz, yani ortogonal olduklarında, faktör yükleri ayrıca faktörler ile değişkenler arasındaki korelasyonu ifade edecektir.

Faktörler dik olsun ya da olmasın faktör yükleri, orijinal değişkenlerin bağımlı ve faktörlerin bağımsız değişkenleri ifade ettiği çoklu regresyon denklemindeki standartlaştırılmış regresyon katsayılarını ifade eder. Faktörler ilişkisiz ise katsayı değerleri de birbirleri ile ilişkisiz olacak ve bu katsayı değerleri, değişkenlerin her bir faktöre sağladığı katkı oranını ifade eder (Norusis, 1993).

Faktör matrisi, faktörleştirme aşamasında elde edilen faktörler ile değişkenler arasındaki ilişkiyi gösterdiği halde, bu matriste bir değişken birden fazla faktörle aynı anda yüksek ilişkili olduğundan bu matrise dayanarak anlamlı faktörleri tanımlamak oldukça güç olacağından rotasyonlu faktör matrisi incelenir. Örneğin; Tablo 6'da Makarna değişkeni 1 ve 2 nolu faktörle hemen hemen aynı oranda ilişkiye sahip olduğundan bu değişkenin hangi faktör tarafından açıklanabildiğini söylemek zordur. Ancak Tablo 7'de döndürme sonrası Makarna değişkeninin sadece 1. faktörle yüksek ilişkiye sahip olduğu görülmektedir.

Tablo 6'da en yüksek faktör yüküne sahip olan değişken BULGUR olup, bu değişken için, birinci faktör $(0,745)^2$'lik, ikinci faktör $(-0,237)^2$'lik, üçüncü faktör $(0,259)^2$'lik, dördüncü faktör ise $(0,197)^2$'lik bir varyansı açıklayacaktır. Böylelikle bu değişkene ait toplam varyans açıklama oranı: $(0{,}745)^2 + (-0{,}237)^2 + (0{,}259)^2 + (0{,}197)^2 = 0{,}717084$ olarak bulunmuştur.

Tablo 7. TBA ve Varimax Yöntemi Uygulandığında Rotasyonlu Faktör Matrisi

	Faktörler			
	1. Faktör	**2. Faktör**	**3. Faktör**	**4. Faktör**
SEHRIYE	**0,906**	-0,042	0,199	0,054
MAKARNA	**0,894**	-0,081	0,231	0,097
BIS1SADE	**0,791**	0,277	-0,161	0,238
TULUMBA	0,308	**0,873**	-0,052	-0,054
KURUPAST	-0,029	**0,857**	0,353	-0,015
YASPASTA	-0,285	**0,796**	0,364	0,199
PRCBALDO	0,032	0,238	**0,902**	0,204
PBERSANI	0,228	0,165	**0,871**	0,004
BAKLAVA	0,064	0,25	**0,556**	0,527
BULGUR	0,422	0,51	**0,523**	-0,068
BIS2BEBE	0,464	0,362	**-0,496**	0,343
YUFKA	0,094	0,007	0,114	**-0,907**
BUGDAYUN	0,331	-0,279	0,184	**0,798**
PIRINUNU	0,352	0,257	0,206	**0,69**

Tablo 8. TBA ve Quartimax Yöntemi Uygulandığında Rotasyonlu Faktör Matrisi

	Faktörler			
	1. Faktör	**2. Faktör**	**3. Faktör**	**4. Faktör**
SEHRIYE	**0,906**	0,201	-0,053	0,033
MAKARNA	**0,895**	0,234	-0,092	0,076
BIS1SADE	**0,799**	-0,153	0,271	0,223
PRCBALDO	0,036	**0,907**	0,228	0,192
PBERSANI	0,228	**0,874**	0,154	-0,011
BAKLAVA	0,075	**0,565**	0,245	0,519
BULGUR	0,424	**0,529**	0,5	-0,084
BIS2BEBE	0,476	**-0,487**	0,364	0,338
TULUMBA	0,316	-0,042	**0,87**	-0,063
KURUPAST	-0,022	0,362	**0,854**	-0,022
YASPASTA	-0,274	0,374	**0,795**	0,197
YUFKA	0,075	0,104	0,002	**-0,91**
BUGDAYUN	0,344	0,191	-0,281	**0,79**
PIRINUNU	0,368	0,218	0,253	**0,68**

3.2 En Çok Olabilirlik (EÇO) Yöntemi

Tablo 9. EÇO Yöntemi Uygulandığında Ortak Faktör Varyansları

	Başlangıç Değeri	Ortak Faktör Varyansı
BUGDAYUN	,797	,788
PIRINUNU	,733	,564
PBERSANI	,926	,853
PRCBALDO	,943	,999
MAKARNA	,985	,981
BULGUR	,862	,591
YUFKA	,758	,844
SEHRIYE	,982	,993
BIS1SADE	,805	,377
BIS2BEBE	,813	,189
YASPASTA	,891	,892
KURUPAST	,892	,941
BAKLAVA	,824	,533
TULUMBA	,805	,661

Tablo 9'da OFV'ları % 37,7 ile % 99,9 arasında değişmektedir. Tablo 11'de görüldüğü gibi burada da döndürmeden önce bazı değişkenler birden fazla faktörle ilişkilidir.

Tablo 10. EÇO Yöntemi Uygulandığında Özdeğerler ve Toplam Açıklanan Varyans Miktarları

	Faktörleştirme Sonrası Yüklerin Kareleri Toplamı			Faktör Döndürmesi Sonrası Yüklerin Kareleri Toplamı		
Faktörler	Toplam	Varyans %	Birikimli %	Toplam	Varyans%	Birikimli %
1	3,479	24,851	24,851	2,751	19,653	19,653
2	2,730	19,501	44,352	2,749	19,638	39,291
3	2,088	14,917	59,269	2,546	18,188	57,479
4	1,909	13,639	72,908	2,160	15,429	72,908

Tablo 11. EÇO Yöntemi Uygulandığında Rotasyonsuz Faktör Matrisi

	Faktörler			
	1. Faktör	**2. Faktör**	**3. Faktör**	**4. Faktör**
PRCBALDO	**0,999**	-0,037	-0,010	-0,001
PBERSANI	**0,882**	0,109	-0,085	-0,236
BULGUR	**0,664**	0,247	0,266	-0,134
BAKLAVA	**0,643**	0,091	0,105	0,317
SEHRIYE	0,178	**0,980**	0,017	-0,014
MAKARNA	0,204	**0,969**	-0,030	0,021
BIS1SADE	0,108	**0,575**	0,083	0,168
KURUPAST	**0,485**	-0,001	**0,840**	0,001
TULUMBA	0,132	0,236	**0,767**	0,002
YASPASTA	**0,518**	-0,239	**0,724**	0,206
YUFKA	-0,081	0,019	0,093	**-0,910**
BUGDAYUN	0,289	0,363	-0,363	**0,664**
PIRINUNU	0,352	0,354	0,106	**0,551**
BIS2BEBE	-0,171	0,218	0,146	0,302

Tablo 12. EÇO ve Varimax Yöntemi Uygulandığında Rotasyonlu Faktör Matrisi

	Faktörler			
	1. Faktör	**2. Faktör**	**3. Faktör**	**4. Faktör**
SEHRIYE	**0,987**	0,119	0,027	0,063
MAKARNA	**0,974**	0,146	-0,009	0,109
BIS1SADE	**0,571**	0,021	0,103	0,201
PRCBALDO	0,009	**0,960**	0,207	0,188
PBERSANI	0,162	**0,904**	0,085	-0,045
BULGUR	0,291	**0,598**	0,385	-0,025
BAKLAVA	0,106	**0,523**	0,265	0,422
BIS2BEBE	0,196	**-0,267**	0,124	0,254
KURUPAST	0,039	0,299	**0,922**	-0,010
YASPASTA	-0,210	0,323	**0,839**	0,198
TULUMBA	0,258	-0,037	**0,768**	-0,055
YUFKA	0,067	0,093	-0,002	**-0,912**
BUGDAYUN	0,331	0,189	-0,246	**0,763**
PIRINUNU	0,342	0,182	0,215	**0,606**

Tablo 13. EÇO ve Quartimax Yöntemi Uygulandığında Rotasyonlu Faktör Matrisi

	Faktörler			
	1.Faktör	2.Faktör	3.Faktör	4.Faktör
PRCBALDO	**0,967**	0,010	0,182	0,175
PBERSANI	**0,906**	0,158	0,061	-0,061
BULGUR	**0,608**	0,290	0,368	-0,038
BAKLAVA	**0,536**	0,114	0,250	0,413
BIS2BEBE	**-0,259**	0,202	0,129	0,254
SEHRIYE	0,124	**0,988**	0,019	0,044
MAKARNA	0,150	**0,975**	-0,018	0,089
BIS1SADE	0,028	**0,575**	0,099	0,190
KURUPAST	0,323	0,043	**0,914**	-0,013
YASPASTA	0,346	-0,203	**0,832**	0,199
TULUMBA	-0,017	0,261	**0,768**	-0,058
YUFKA	0,080	0,050	-0,003	**-0,914**
BUGDAYUN	0,195	0,343	-0,254	**0,754**
PIRINUNU	0,197	0,353	0,207	**0,598**

3.3 Ağırlıklandırılmamış En Küçük Kareler Yöntemi

Tablo 14. AEKK Yöntemi Uygulandığında Ortak Faktör Varyansları

	Başlangıç Değeri	Ortak Faktör Varyansı
BUGDAYUN	,797	,822
PIRINUNU	,733	,612
PBERSANI	,926	,802
PRCBALDO	,943	,995
MAKARNA	,985	,832
BULGUR	,862	,612
YUFKA	,758	,774
SEHRIYE	,982	,831
BIS1SADE	,805	,667
BIS2BEBE	,813	,462
YASPASTA	,891	,861
KURUPAST	,892	,819
BAKLAVA	,824	,516
TULUMBA	,805	,845

Tablo 14'te OFV'ları % 46,2 ile % 99,5 arasında değişmektedir. Faktörleştirme sonrası faktörlerin değişkenlere ait varyansı açıklama oranı değişmektedir.

Tablo 16'da görüldüğü gibi burada da döndürmeden önce bazı değişkenler birden fazla faktörle ilişkilidir.

Tablo 15. AEKK Yöntemi Uygulandığında Özdeğerler ve Toplam Açıklanan Varyans Miktarları

	Faktörleştirme Sonrası Yüklerin Kareleri Toplamı			Faktör Döndürmesi Sonrası Yüklerin Kareleri Toplamı		
Faktörler	**Toplam**	**Varyans %**	**Birikimli %**	**Toplam**	**Varyans%**	**Birikimli%**
1	4,587	32,767	32,767	2,896	20,686	20,686
2	2,613	18,663	51,429	2,746	19,612	40,297
3	1,740	12,426	63,855	2,647	18,905	59,202
4	1,511	10,793	**74,649**	2,163	15,446	**74,649**

Tablo 16. AEKK Yöntemi Uygulandığında Rotasyonsuz Faktör Matrisi

	Faktörler			
	1. Faktör	**2. Faktör**	**3. Faktör**	**4. Faktör**
PRCBALDO	**0,730**	-0,403	-0,412	0,360
BULGUR	**0,709**	-0,195	0,202	0,175
PBERSANI	**0,670**	-0,278	-0,176	**0,494**
PIRINUNU	**0,661**	0,257	-0,220	-0,248
BAKLAVA	**0,662**	-0,092	-0,345	-0,043
KURUPAST	**0,616**	**-0,578**	0,228	-0,231
MAKARNA	**0,586**	**0,568**	0,238	0,331
SEHRIYE	**0,579**	**0,553**	0,298	0,319
BIS1SADE	**0,540**	**0,493**	0,340	-0,132
YASPASTA	**0,544**	**-0,649**	-0,058	-0,375
BUGDAYUN	**0,431**	**0,574**	**-0,548**	-0,085
TULUMBA	**0,569**	-0,227	**0,574**	-0,374
YUFKA	-0,231	-0,321	**0,563**	**0,549**
BIS2BEBE	0,271	0,386	0,261	**-0,414**

Tablo 17. AEKK ve Varimax Yöntemi Uygulandığında Rotasyonlu Faktör Matrisi

	Faktörler			
	1. Faktör	**2. Faktör**	**3. Faktör**	**4. Faktör**
SEHRIYE	**0,887**	0,202	-0,039	0,047
MAKARNA	**0,874**	0,232	-0,076	0,088
BIS1SADE	**0,744**	-0,085	0,233	0,228
BIS2BEBE	**0,430**	-0,340	0,270	0,297
PRCBALDO	-0,002	**0,957**	0,213	0,184
PBERSANI	0,194	**0,862**	0,145	-0,018
BULGUR	0,378	**0,511**	0,453	-0,048
BAKLAVA	0,088	**0,508**	0,242	0,438
TULUMBA	0,338	-0,036	**0,852**	-0,055
KURUPAST	-0,005	0,347	**0,836**	-0,022
YASPASTA	-0,261	0,359	**0,792**	0,192
YUFKA	0,060	0,093	-0,007	**-0,873**
BUGDAYUN	0,325	0,201	-0,266	**0,778**
PIRINUNU	0,355	0,226	0,242	**0,613**

Tablo 18. AEKK ve Quartimax Yöntemi Uygulandığında Rotasyonlu Faktör Matrisi

	Faktörler			
	1. Faktör	**2. Faktör**	**3. Faktör**	**4. Faktör**
SEHRIYE	**0,886**	0,204	-0,059	0,013
MAKARNA	**0,875**	0,234	-0,096	0,055
BIS1SADE	**0,756**	-0,075	0,222	0,203
BIS2BEBE	**0,445**	-0,329	0,270	0,287
PRCBALDO	0,004	**0,964**	0,194	0,167
PBERSANI	0,193	**0,865**	0,124	-0,039
BULGUR	0,382	**0,520**	0,436	-0,071
BAKLAVA	0,106	**0,520**	0,232	0,425
TULUMBA	0,351	-0,019	**0,847**	-0,070
KURUPAST	0,007	0,363	**0,828**	-0,031
YASPASTA	-0,241	0,377	**0,790**	0,191
YUFKA	0,030	0,078	-0,013	**-0,876**
BUGDAYUN	0,346	0,209	-0,273	**0,764**
PIRINUNU	0,379	0,242	0,233	**0,596**

3.4 Genelleştirilmiş En Küçük Kareler Yöntemi

Tablo 19. GEKK Yöntemi Uygulandığında NRMEKMEK Değişkeni Analizdeyken Ortak Faktör Varyansları

	Başlangıç Değeri	Ortak Faktör Varyansı
NRMEKMEK	,321	,458
BUGDAYUN	,798	,868
PIRINUNU	,753	,860
PBERSANI	,929	,990
PRCBALDO	,944	,972
MAKARNA	,985	,999
BULGUR	,866	,939
YUFKA	,759	,856
SEHRIYE	,983	,988
BIS1SADE	,816	,861
BIS2BEBE	,816	,987
YASPASTA	,892	,936
KURUPAST	,903	,966
BAKLAVA	,831	,945
TULUMBA	,805	,868

Bu yöntemde normal ekmek değişkeni çıkarılarak yapılan analizde 25. iterasyonda minimum değer bulunamadığından sonuçlar normal ekmek analize dahil iken verilmiştir.

Tablo 19'da OFV'ları % 45,8 ile % 99,9 arasında değişmektedir

Tablo 20. GEKK; Yöntemi Uygulandığında NRMEKMEK Değişkeni Analizdeyken Özdeğerler ve Toplam Açıklanan Varyans Miktarları

	Faktörleştirme Sonrası Yüklerin Kareleri Toplamı			**Faktör Döndürmesi Sonrası Yüklerin Kareleri Toplamı**		
Faktörler	**Toplam**	**Varyans %**	**Birikimli %**	**Toplam**	**Varyans%**	**Birikimli%**
1	3,181	21,205	21,205	2,754	18,360	18,360
2	3,176	21,170	42,376	2,724	18,159	36,519
3	2,048	13,651	56,027	2,640	17,599	54,118
4	1,658	11,056	**67,082**	1,945	12,964	**67,082**

Tablo 20'de yer alan "Faktör Döndürmesi Sonrası Yüklerin Kareleri Toplamı" sütunu incelendiğinde, toplam açıklanan varyans yüzdesi ve faktörlerin her biri ile açıklanan varyans oranında değişmenin olduğu göze çarpar.

Tablo 21. GEKK Yöntemi Uygulandığında Rotasyonsuz Faktör Matrisi

	Faktörler			
	1. Faktör	**2. Faktör**	**3. Faktör**	**4. Faktör**
MAKARNA	**0,999**	-0,033	-0,024	0,005
SEHRIYE	**0,986**	-0,044	0,012	0,022
BUGDAYUN	**0,441**	0,024	0,188	-0,265
PIRINUNU	**0,437**	0,233	0,384	0,043
PRCBALDO	0,199	**0,927**	0,049	-0,040
PBERSANI	0,308	**0,899**	0,072	-0,256
BULGUR	0,388	**0,614**	0,304	0,116
BAKLAVA	0,264	**0,564**	0,209	0,093
BIS2BEBE	0,161	-0,281	**0,930**	-0,126
BIS1SADE	**0,587**	-0,101	**0,629**	-0,076
YUFKA	-0,018	0,095	**-0,241**	-0,055
KURUPAST	0,094	**0,530**	0,265	**0,765**
YASPASTA	-0,124	**0,574**	0,250	**0,689**
TULUMBA	0,246	0,171	**0,502**	**0,588**
NRMEKMEK	0,125	0,125	0,131	**0,258**

Tablo 22. GEKK ve Varimax Yöntemi Uygulandığında Rotasyonlu Faktör Matrisi

	Faktörler			
	1. Faktör	**2. Faktör**	**3. Faktör**	**4. Faktör**
KURUPAST	**0,942**	0,237	0,046	-0,005
YASPASTA	**0,882**	0,268	-0,176	-0,039
TULUMBA	**0,726**	0,022	0,197	0,347
NRMEKMEK	**0,311**	0,050	0,108	0,059
PBERSANI	0,097	**0,970**	0,152	-0,018
PRCBALDO	0,294	**0,895**	0,061	-0,105
BULGUR	0,400	**0,606**	0,266	0,188
BAKLAVA	0,335	**0,540**	0,163	0,098
MAKARNA	-0,002	0,131	**0,985**	0,104
SEHRIYE	0,019	0,115	**0,971**	0,135
BUGDAYUN	-0,174	0,210	**0,381**	0,283
BIS2BEBE	0,045	-0,102	0,059	**0,985**
BIS1SADE	0,074	0,088	0,497	**0,705**
PIRINUNU	0,227	0,301	0,348	**0,362**
YUFKA	-0,082	0,083	-0,002	**-0,238**

Tablo 23. GEKK ve Quartimax Yöntemi Uygulandığında Rotasyonlu Faktör Matrisi

	Faktörler			
	1. Faktör	**2. Faktör**	**3. Faktör**	**4. Faktör**
KURUPAST	**0,944**	0,045	0,226	-0,020
YASPASTA	**0,884**	-0,178	0,258	-0,047
TULUMBA	**0,732**	0,205	0,013	0,330
NRMEKMEK	**0,313**	0,109	0,046	0,051
MAKARNA	0,001	**0,988**	0,129	0,076
SEHRIYE	0,023	**0,975**	0,113	0,108
BUGDAYUN	-0,167	**0,390**	0,211	0,275
PIRINUNU	0,236	**0,358**	0,298	0,348
PBERSANI	0,109	0,153	**0,969**	-0,025
PRCBALDO	0,303	0,060	**0,892**	-0,112
BULGUR	0,410	0,272	**0,601**	0,174
BAKLAVA	0,343	0,166	**0,536**	0,087
BIS2BEBE	0,059	0,086	-0,102	**0,982**
BIS1SADE	0,085	0,516	0,087	**0,690**
YUFKA	-0,085	-0,008	0,083	**-0,236**

Tablo 24. Dört Farklı Faktörleştirme Yöntemleri ve Varimax Döndürme Yöntemi Uygulandığında Genel Faktör Sıralamaları

Sıra No	TBA Yöntemi	EÇO Yöntemi	AEKK Yöntemi	GEKK Yöntemi
1	İstanbul	İstanbul	Diyarbakır	Diyarbakır
2	Gaziantep	Gaziantep	İstanbul	İstanbul
3	Diyarbakır	Diyarbakır	Gaziantep	Ankara
4	Ankara	Ankara	Ankara	Zonguldak
5	İçel	İçel	İçel	İzmir
6	Antalya	Samsun	Eskişehir	Gaziantep
7	İzmir	Eskişehir	İzmir	Türkiye
8	Eskişehir	Antalya	Samsun	Antalya
9	Denizli	Denizli	Antalya	Bursa
10	Bursa	Zonguldak	Bursa	Denizli
11	Samsun	Türkiye	Denizli	İçel
12	Zonguldak	Bursa	Türkiye	Erzurum
13	Türkiye	İzmir	Erzurum	Samsun
14	Erzurum	Trabzon	Zonguldak	Eskişehir
15	Adana	Adana	Trabzon	Trabzon
16	Trabzon	Kocaeli	Adana	Konya
17	Kocaeli	Erzurum	Kocaeli	Kocaeli
18	Konya	Konya	Konya	Adana
19	Kayseri	Kayseri	Kayseri	Kayseri
20	Malatya	Malatya	Malatya	Malatya

Tablo 25. Dört Farklı Faktörleştirme Yöntemleri ve Quartimax Döndürme Yöntemi Uygulandığında Genel Faktör Sıralamaları

Sıra No	TBA Yöntemi	EÇO Yöntemi	AEKK Yöntemi	GEKK Yöntemi
1	İstanbul	İstanbul	Diyarbakır	Diyarbakır
2	Gaziantep	Diyarbakır	İstanbul	İstanbul
3	Diyarbakır	Gaziantep	Gaziantep	Ankara
4	Ankara	Ankara	Ankara	İzmir
5	İçel	İçel	İçel	Zonguldak
6	Antalya	Samsun	Eskişehir	Gaziantep
7	İzmir	Eskişehir	İzmir	Antalya
8	Bursa	Antalya	Bursa	Bursa
9	Eskişehir	Bursa	Samsun	Denizli
10	Denizli	Denizli	Antalya	Erzurum
11	Samsun	Zonguldak	Türkiye	İçel
12	Türkiye	Trabzon	Denizli	Türkiye
13	Zonguldak	Adana	Erzurum	Samsun
14	Erzurum	İzmir	Trabzon	Eskişehir
15	Adana	Kocaeli	Zonguldak	Konya
16	Trabzon	Türkiye	Adana	Trabzon
17	Kocaeli	Erzurum	Kocaeli	Kocaeli
18	Konya	Konya	Konya	Adana
19	Kayseri	Kayseri	Malatya	Kayseri
20	Malatya	Malatya	Kayseri	Malatya

"Temel Bileşenler Analizi", "En Çok Olabilirlik" ve "Ağırlıklandırılmamış En Küçük Kareler" yöntemleri ile birlikte "Varimax" ve "Quartimax" döndürme yöntemleri uygulandığında ilk beş ilin "İstanbul, Gaziantep, Diyarbakır, Ankara, İçel" şeklinde ve son üçünün ise "Konya, Kayseri, Malatya" şeklinde olduğu görülmektedir.

"Genelleştirilmiş En Küçük Kareler" yöntemi ile birlikte "Varimax" ve "Quartimax" yöntemleri uygulandığında ilk beş ilin "Diyarbakır, İstanbul, Ankara, Zonguldak, İzmir" şeklinde sıralandığı ve yine son üç ilin de "Adana, Kayseri, Malatya" şeklinde olduğu belirlenmiştir. Bu yöntemdeki farklılık normal ekmek değişkeninin de analize dahil edilerek yapılmasından kaynaklanmaktadır. Diğer illerde gözlenen sıralama farklılıkları kullanılan yöntemlerdeki ağırlıklandırmadan kaynaklanmaktadır.

4. TARTIŞMA VE SONUÇ

Analiz sonuçlarından görüldüğü gibi tüm yöntemlerde genel olarak Şehriye, Makarna, Bis1sade; Tulumba, Kurupasta, Yaşpasta; Prcbaldo, Pbersani, Baklava, Bulgur, Bis2bebe ve Yufka, Bugdayun, Prinunu değişkenlerinden oluşan 4 faktör elde edilmiştir. Oluşan faktörler "makarna ürünleri", "tatlı ürünleri", "tahıl ürünleri" ve "un ürünleri" kavramları ile adlandırılabilir. Yapılacak olan araştırmada örneğin 14 değişken yerine bu faktörlerden anlamlı olan 4 faktör kullanılarak ilgili analizin yapılmasının uygun olacağı sonucuna varılmıştır.

Elde edilen sonuçlara göre tüm yöntemlerde faktörleştirme ile ilgili benzer yapı elde edildiğinden, Temel Bileşenler Analizi yönteminde varyans açıklama oranı diğer yöntemlere göre daha fazla olduğundan, bu yöntem diğerlerine göre tercih edilebilir.

KAYNAKLAR

AKGÜL, A. (2003), *Tıbbi Araştırmalarda İstatistiksel Analiz Teknikleri,* Ankara, Emek Ofset Ltd.Sti.

CHATFIELD, C. ; COLLINS, A.J. (1980), *Introduction to Multivariate Analysis*, London, Chapman and Hall, Ltd.

DARTON, R.A. (1980), *Rotation in Factor Analysis,* the Statistician, 29(3): 167-194.

GARNETT, J.C.M. (1919), *On Certain Independent Factors in Mental Measurements*, Proceedings of the Royal Society, London, A, 96, 91-111.

GEORGE, D. ; MALLERY, P. (2003), *SPSS for Windows Step by Step a Simple Guide and Reference 11.0 Update,* United States of America, Pearson Education Inc.

HAIR; J.F. ; ANDERSON R.E. ; TATHAM, R.L., BLACK W.C. (1998). *Multivariate Data Analysis with Readings,* USA, Pentice-Hall International Inc.

HARMAN, H.H. (1968), *Modern Factor Analysis,* Chicago, the University of Chicago Press.

HOTELLING, H. (1933), *Analysis of a Complex of Statistical Variables into Principal Components,* Journal of Educational Psychology, 24, 417-441, 498-520.

JOHNSON, R.A. ; WICHERN, D.W. (2002), *Applied Multivariate Statistical Analysis,* United States of America, Prentice-Hall Inc.

JORESKOG, K.G. ; GOLDBERGER, A.S. (1972), *Factor Analysis by Generalized Least Squares,* Psychometrika, 37(3): 243-260.

KAISER, H.F. (1958), *The Varimax Criterion for Analytic Rotation in Factor Analysis,* Psychometrika, 23(3): 187-200.

LAWLEY, D. N. ; MAXWELL, A. E. (1971), *Factor Analysis as a Statistical Method,* New York , American Elsevier Publishing Company, Inc.

NORUSIS, M.J. (1993), *SPSS for Windows,* Chicago, Professional Statistics.

PEARSON, K. (1901), *On Lines and Planes af Closest Fit to a System of Points in Space,* Phil. Mag., 2(6): 557-572.

SHARMA, S. (1996), *Applied Multivariate Techniques,* USA.

SPEARMAN, C. (1904), *General Intelligence Objectively Determined and Measured,* Am. J. Psychol., 15, 201-293.

SPEARMAN, C. (1926), *The Abilities of Man,* MacMillan, London

SRIVASTAVA, M.S.(2002), *Methods of Multivariate Statistics,* Canada, John Wiley &Sons, Inc.

THURSTONE, L.L. (1931). *Multiple Factor Analysis.* Psychological Review, 38, 406-427.

COMPARING FOUR FACTOR EXTRACTION METHODS WITH AN EXAMPLE

ABSTRACT

There are several methods, for determining the number of factors, estimating of the Λ and Ψ parameters and forming factor scores in a factor analysis. Some of the popular methods are Principal Component Analysis, Maximum Likelihood, Unweighted Least Squares and Generalized Least Squares. In this study, we analyze the data provided by Turkish Statistical Institute (TURKSTAT), which is about the prices of miscellaneous food products belonging to 20 province of Turkey for December 2003 with SPSS statistical package. The variables are processed by using 4 different extraction methods and for each method we use 2 different rotations. The purpose of this study is to determine which food products are selected without losing a lot of information in the research of the prices of food products. We compare the results and attempt to determine the differences of these methods.

Key Words: *Factor Analysis, Generalized Least Squares, Maximum Likelihood, Principal Component Analysis, Unweighted Least Squares, Quartimax, Varimax.*

İstatistik Araştırma Dergisi Cilt 04 No 03 Sayfa 30-44 Aralık 2005

BAYES AĞLARDA BİRLEŞME AĞAÇLARINI KULLANARAK OLAYLARIN GENİŞLEMESİ

Hülya OLMUŞ* Semra ERBAŞ*

ÖZET

Bir Bayes ağ, yön verilmiş döngüsel olmayan bir grafiktir. Bu grafikte, düğümler rastgele değişkenleri gösterir ve kenarlar değişkenler arası doğrudan bağımlılıkları tanımlar. Kenarlar arası bağımlılıkların gücü, koşullu olasılıklar ile tanımlanmıştır. Bayes ağlarda, olayların genişlemesi, birleşme ağacı algoritması ile tanıtılmıştır. Uygulama da yer alan veriden hareketle, koşullu olasılıklar hesaplanmış ve bu olasılıklar incelenmiştir. Burada, olumsallık tabloların oluşturulması için SPSS 8.0 paket programı ve olasılık hesaplamaları için de HUGIN paket programı kullanılmıştır.

Anahtar Kelimeler: *Bayes Ağlar, Birleşme Ağacı, Koşullu Bağımsızlık, Koşullu Olasılık.*

1. GİRİŞ

Bir grafik, hem yön verilmiş hem de döngüsel olmama durumlarına sahip ise, grafiğe yön verilmiş döngüsel olmayan grafik denir. Bayes ağlar, ilgilenilen değişkenler arası olasılık ilişkilerini gösteren grafik modelleridir. Bayes ağlar, değişkenler arasındaki ilişkileri grafik olarak göstermek ve uzman sistemlerde kesin olmayışlık ile ilgilenmek için güçlü bir yöntemdir. Yön verilmiş döngüsel olmayan grafik modelleri geniş uygulama alanlarında önemli bir yer almıştır. Bayes ağlar; özellikle, hastalık tanı sistemleri, karar destek sistemleri, planlama ve kontrol, dinamik sistemler, zaman serileri, genel veri analizi ve istatistikteki bilgi alanlarında kullanılmaktadır. Bu alanlarda kullanımları, Lauritzen ve Spiegelhalter (1988), Pearl (1986), Spiegelhalter ve Lauritzen (1990) ve Spiegelhalter vd., (1993), Friedman ve Goldszmidt (1996) tarafından detaylı bir şekilde incelenmiştir.

Bir Bayes ağ, koşullu bağımsızlık özelliklerin yapısını da tanımlayan yön verilmiş döngüsel olmayan grafiktir. Bayes ağ, değişkenler ve değişkenler arası yön verilmiş kenarların kümesinden oluşur. Her bir değişken karşılıklı bağımsız durumların sonlu bir kümesine sahiptir. Her düğüm rastgele değişkenle gösterilir. Kenarlar, değişkenler arası olasılık bağımlılıklarını gösterir. Bu bağımlılıklar koşullu olasılıkların kümesinden oluşur. Her bir değişkenin ebeveynleri verildiğinde, değişkenin koşullu olasılığı belirlenir. Bir düğümün ebeveynleri olmadığı zaman, bir değişken marjinal bir olasılığa sahiptir. Düğümün ebeveynleri üzerinde koşullandırma yapılarak, her düğüm

* Gazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Beşevler, Ankara, TÜRKİYE

için koşullu olasılıklar $P(X \backslash eb(X))$ belirlenir. Burada, $eb(X)$, X düğümünün ebeveynidir. Tüm değişkenler kümesi üzerinde ortak olasılık,

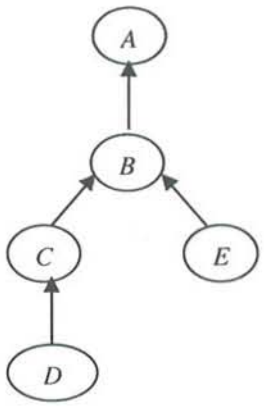
$$P(U) = \prod_{X} P(X / eb(X))$$

ifadelerin çarpımıyla verilir (Jensen, 1996; Liarokapis, 1999).

Örneğin, beş tane değişkenden (düğüm) oluşan A,B,C,D,E bir ağa sahip olunduğu düşünülsün. Zincir kuralı ile bu beş değişkene ilişkin ortak olasılık dağılımı $P(A,B,C,D,E)$ hesaplanabilir. Beş düğüme ilişkin ortak olasılık dağılımı, aşağıdaki şekilde yazılır.

$$P(A,B,C,D,E)=P(A \backslash B,C,D,E)\ P(B \backslash C,D,E)\ P(C \backslash D,E)\ P(D \backslash E)\ P(E)$$

Ancak; bir Bayes ağda bağımlılıklar açık bir şekilde modellenmiştir. Bunun için aşağıda bir ağ verilmiştir (Şekil 1).



Şekil 1. Beş Düğüme İlişkin Ağ Modeli

Şekil 1'den Bayes ağ modeli için ortak olasılık dağılımı $P(A,B,C,D,E)$'yi hesaplamak basittir. Bu hesaplama,

$$P(A,B,C,D,E)=P(A \backslash B)\ P(B \backslash C,E)\ P(C \backslash D)\ P(D)\ P(E)$$

ile yapılır.

Bayes ağlar; literatürde, fikir ağlar (belief networks), bayes fikir ağlar (bayesian belief networks), nedensel olasılıklı ağlar (causal probabilistic networks), olasılıklı etki diyagramları (probabilistic influence diagrams) ve olasılıklı neden-etki modelleri (probabilistic cause-effect models), olarak da adlandırılır. Bu çalışmada, Bayes ağ terimi kullanılmıştır.

Bu çalışmada, Bayes ağlarda olasılık hesaplamalarının elde edilmesi için, HUGIN'de kullanılan algoritma tanıtılacaktır. Algoritma, Bayes ağlar üzerinde doğrudan kullanılmamaktadır. Değişkenler kümesinin ağacı, birleşme ağacı olarak adlandırılır ve bunun üzerinde çalışılabilir. Burada, kümeler takımlardır. Takımlar, üçgenleştirilmiş grafikten elde edilebilmektedir. Üçgenleştirilmiş grafik, ağ üzerinde kurulan özel bir grafiktir. Her takım, değişkenlerin düzenleri üzerinde oluşan bir tabloya sahiptir. HUGIN programı, bu tablolar üzerindeki işlemlerin bir serisinden oluşur (HUGIN Systems, 1998).

2. BAYES AĞLARDA ÇIKARSAMALAR

Bayes ağlarda, bir değişken gözlenmiş ise başka değişkenler için bilgilere ihtiyaç vardır. Çıkarsama, başka değişkenlerin değerleri bilindiğinde, bir değişken için olasılık dağılımının güncelleştirilmesi ile son dağılımın elde edilmesidir (Stephenson, 2000). Bayes ağlarda, böyle çıkarsamalar yapmak için birleşme ağacı algoritması kullanılmıştır. Bu algoritmayı tanıtmadan önce, değişik Bayes ağ modelleri kullanılarak çıkarsamayı gösteren örnekler verilecektir. Bu örneklere olasılık hesaplamaları uygulanacak ve yapılan olasılık hesaplamaları, nedenselliği kurmak için kullanılacaktır. Bu örnekler, birleşme ağacı algoritmasının temelini oluşturur.

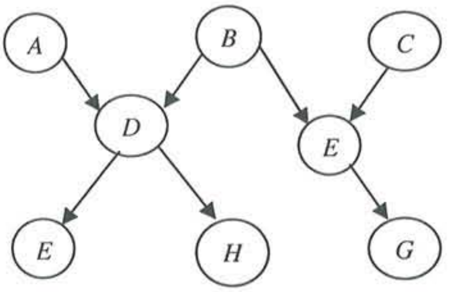
2.1. Fikir Takımlarının Yapıları

Bu kesimde bilgi alanındaki model, çoklu bağlantılı Bayes ağları gerektirir. Yön verilmiş döngüsel olmayan grafiklerde, birleşme ağaçlarının elde edilmesi için bir metot ve bilgi alanını modellemek için yeni bir grafik yapısı verilecektir. Bu yapı, birleşme ağacı olarak adlandırılır (Jensen,1996; Cowell, 1999).

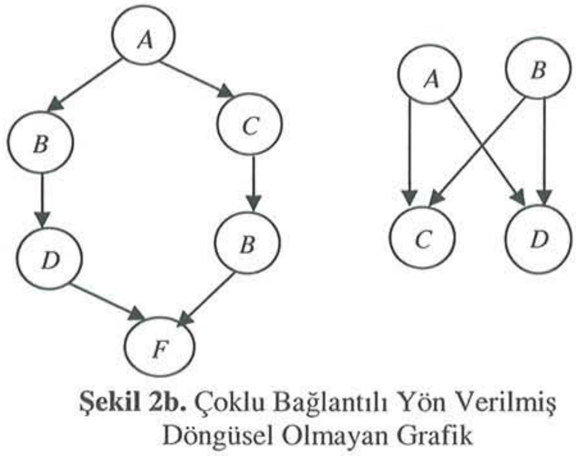
2.1.1. Tekli Bağlantılı Yön Verilmiş Döngüsel Olmayan Grafik

Bağlantıların yönlerini azaltarak aldığımız grafik ağaç ise, yön verilmiş döngüsel olmayan grafik, tekli bağlantılıdır denir (Jensen, 1996), (Şekil 2a ve 2b). Yani, düğümlerin her çifti arasında tek bir yol vardır.

Çoklu bağlantılı yön verilmiş döngüsel olmayan grafikte ise, düğümlerin her çifti arasında bir yoldan daha fazlası mümkündür.



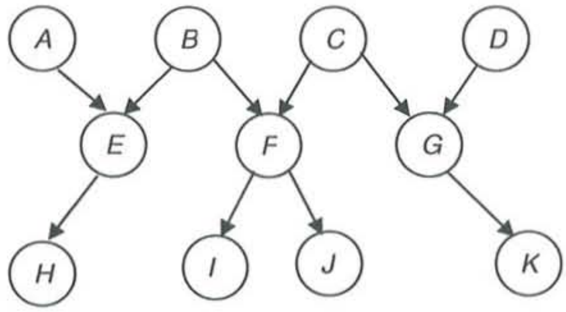
Şekil 2a. Tekli Bağlantılı Yön Verilmiş Döngüsel Olmayan Grafik



Şekil 2b. Çoklu Bağlantılı Yön Verilmiş Döngüsel Olmayan Grafik

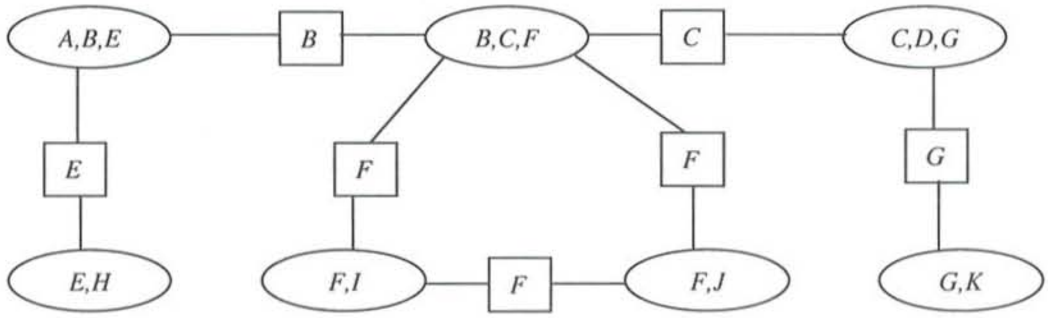
Tekli bağlantılı yön verilmiş döngüsel olmayan grafikler için, birleşme ağaçlarını kurmak kolaydır. Her *A* değişkeni, $eb(A)\neq\emptyset$ ile $eb(A)U\{A\}$ takımı kurulur. Her iki takım arasına, boş olmayan arakesit eklenir. Elde edilen grafiğe, birleşme grafiği denir (Jensen, 1996).

Tekli bağlantılı yön verilmiş döngüsel olmayan grafik için ağ modeli aşağıda verilmiştir (Şekil 3).



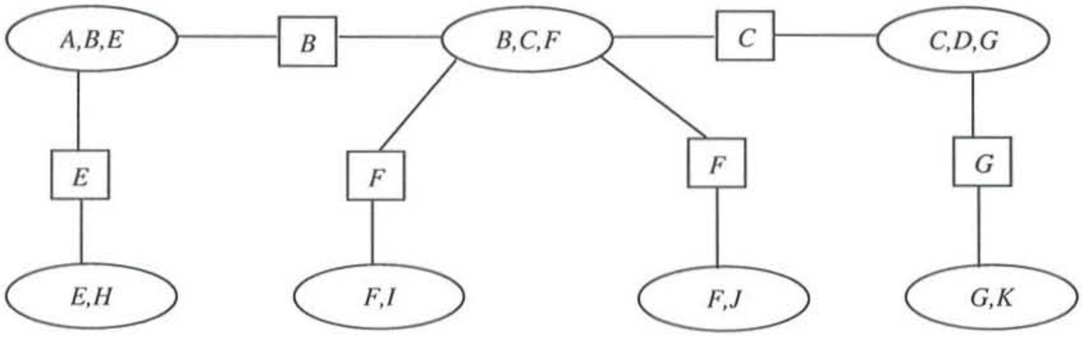
Şekil 3. Tekli Bağlantılı Yön Verilmiş Döngüsel Olmayan Grafik

Birleşme ağaçlarını kurabilmek için, ilk önce takımlar elde edilir. *E* düğümünün ebeveynleri $eb(E)=\{A,B\}$ ile $\{A,B,E\}$ takımı ve *F* düğümünün ebeveynleri $eb(F)=\{B,C\}$ ile $\{B,C,F\}$ takımları ele alınır. Bu iki takımın arakesiti *B* düğümü ile bağlanır. Daha sonra, *G* düğümü $eb(G)=\{C,D\}$ ile (C,D,G) takımı oluşturulur. (C,D,G) ve (B,C,F) takımları da *C* arakesiti ile bağlanır. $eb(H)=\{E\}$ ile (E,H) takımı ve $eb(I)=\{F\}$ ile $\{I,F\}$ takımı, $eb(J)=\{F\}$ ile (J,F) takımı ve $eb(K)=\{G\}$ ile (G,K) takımları elde edilir. Ortak düğüme sahip olan takımlar arakesitleri ile bağlanarak birleşme grafiği elde edilir (Şekil 4). Arakesitler, kare kutular ile gösterilmiştir.



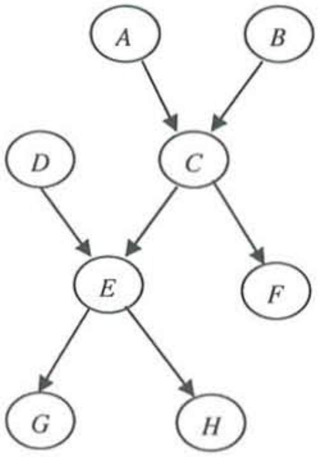
Şekil 4. Şekil 3 için Uygun Birleşme Grafiği

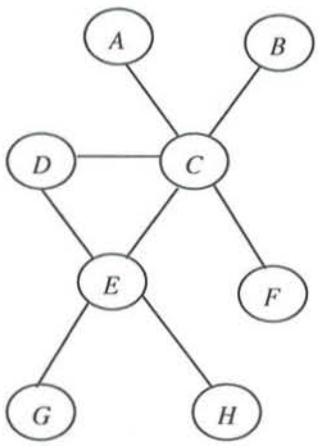
Elde edilen birleşme ağacı, döngülere sahip ise, döngü üzerindeki tüm arakesitler aynı değişkeni kapsar. Böylece, kenarların her biri döngüleri kırarak ortadan kaldırılır. Kenarların kaldırılması ile birleşme ağacı elde edilir (Jensen, 1996). Şekil 4'te (B,C,F), (F,I) ve (F,J) takımları, F düğümü ile bir döngü yaratmaktadır. Bu döngüyü ortadan kaldırmak için, (F,I) ve (F,J) takımları arasındaki *F* arakesiti ve kenarları ortadan kaldırılır. Böylece, Şekil 3 için uygun birleşme ağacı elde edilir (Şekil 5).



Şekil 5. Şekil 3 için Uygun Birleşme Ağacı

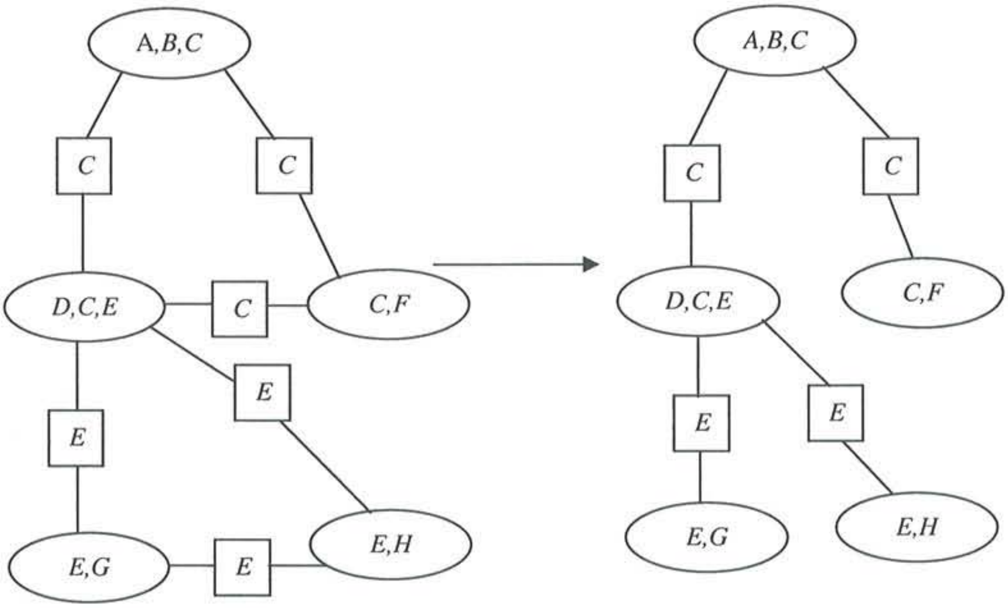
Birleşme ağacını kurabilmek için, aşağıda farklı bir metot verilecektir. İlk olarak moral grafik elde edilir ve elde edilen moral grafikten takımlar elde edilir. Birleşme ağacının yapısı, Şekil 6a, 6b, 6c ve 6d'de gösterilmiştir. Bu şekildeki Bayes ağ için uygun moral grafik elde edilmiştir. Bu moral grafikteki takımlar, *(A,B,C)*, *(D,C,E)*,*(C,F)*, *(E,G)* ve *(E,H)*'dır. İlk olarak, *(A,B,C)*, *(D,C,E)* ve *(C,F)* takımları ortak *C* düğümüne sahiptir. Böylece, bu takımlar, ayrı ayrı *C* arakesiti ile bağlanır. *(D,C,E)*, *(E,G)* ve *(E,H)* takımları da ortak *E* düğümüne sahiptirler. Bu takımlarda, *E* arakesiti ile ayrı ayrı bağlanırlar. Bu düzenleme sonucunda, birleşme grafiği elde edilmiş olur. Burada, birleşme grafiği, döngüye sahip olduğundan dolayı, döngüyü bozacak şekilde bağlantılardan biri ortadan kaldırılmalıdır. Böylece, birleşme ağacı elde edilmiş olur.





Şekil 6a. Çoklu Bağlantılı Yön Verilmiş Döngüsel Olmayan Grafik

Şekil 6b. Moral Grafik



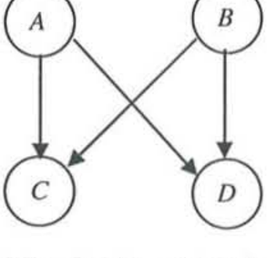
Şekil 6c. Birleşme Grafiği **Şekil 6d.** Birleşme Ağacı

2.1.2. Yapay Bağlantıların Eklenmesi ile Elde Edilen Birleşme Grafikleri

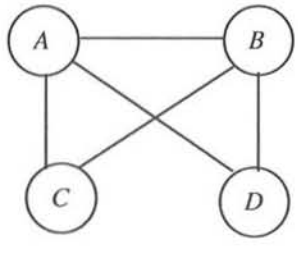
Verilen Bayes ağ, moral grafik haline getirildikten sonra, takımlar, bazen kolaylıkla görülmemektedir. Bundan dolayı, moral grafiğe, yapay bağlantılar eklenerek, grafik üçgenleştirilmiş grafik haline getirilir. Takımlar da, bu grafikten görülebilir. Takımlar sayesinde, değişkenler arasındaki koşullu bağımsızlık ilişkileri kolaylıkla görülmektedir (Edwards, 1995; Lauritzen, 1996).

Elde edilen moral grafik, üçgenleştirilmiş grafiktir. Bu grafikte uzunluğu 3'den büyük her döngü kirişe sahip olmalıdır. Üçgenleştirilmiş grafiği elde edebilmek için büyük ağlarda, yapay bağlantıların nasıl ekleneceği literatürde vardır. Yapay bağlantıların minimal sayısının bulunması problemi NP-hard olarak adlandırılır (Jensen vd., 1990). Bu çalışma da, amacımız, fikir takımlarının minimal kümesini alarak yapay bağlantıları eklemektir.

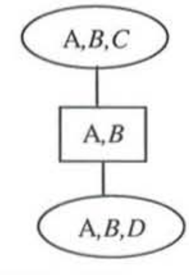
Aşağıda verilen ağ düşünülsün (Şekil 7a, 7b ve 7c). Burada iki takım arasındaki arakesit (A,B)'dir ve birleşme ağacı kolaylıkla elde edilebilir.



Şekil 7a. Çoklu Bağlantılı Yön Verilmiş Döngüsel Olmayan Grafik

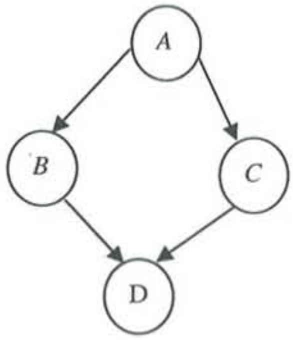


Şekil 7b. Moral Grafik

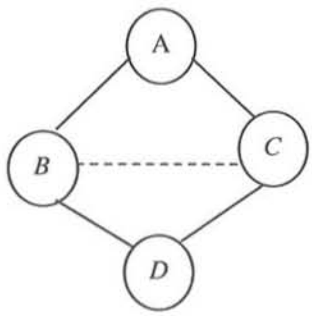


Şekil 7c. Birleşme Grafiği

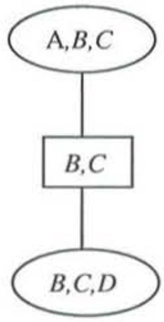
Şekil (8a, 8b ve 8c)'deki yön verilmiş döngüsel olmayan grafik düşünüldüğünde, birleşme ağacı doğrudan oluşmamaktadır. *B,C* düğümleri arasına bir yapay bağlantı eklenirse, üçgenleştirilmiş grafik elde edilir. Böylece, kolaylıkla birleşme ağacı sağlanır.



Şekil 8a. Yön Verilmiş Döngüsel Olmayan Grafik

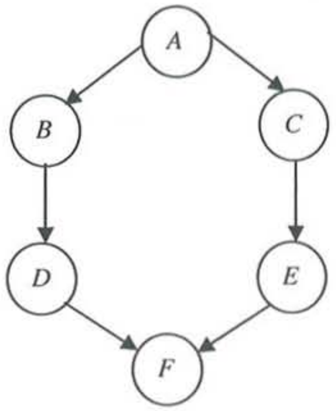


Şekil 8b. Moral Grafik

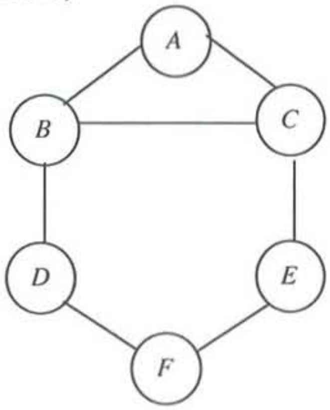


Şekil 8c. Birleşme Grafiği

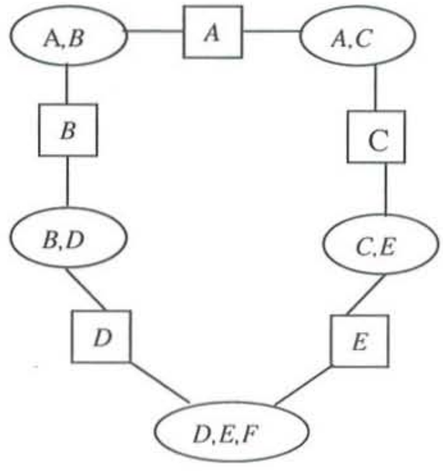
Aşağıda verilen grafikte, sadece, *D* ve *E* arasına bağlantı ekleyerek genişleme problemi ortadan kaldırılamamaktadır. Çünkü, moral grafikte, *D-E-C*-A-*B-D* döngüsü oluşmaktadır (Şekil 9a, 9b ve 9c).



Şekil 9a. Yön Verilmiş Döngüsel Olmayan Grafik

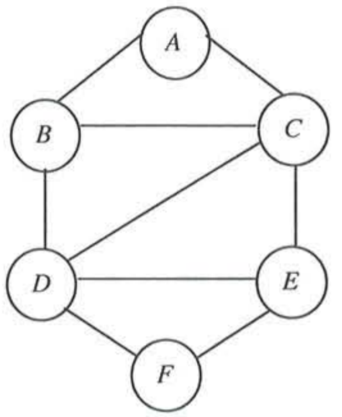


Şekil 9b. Moral Grafik

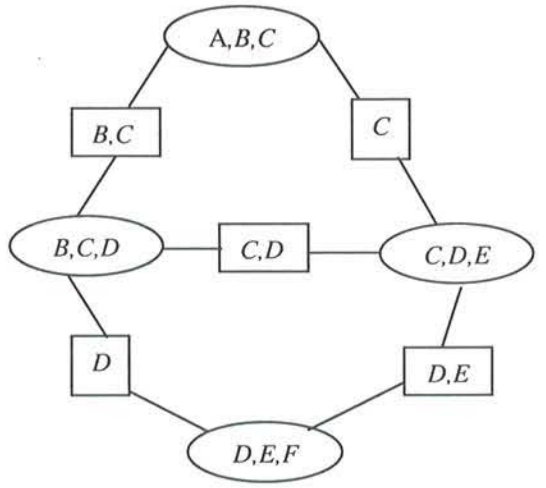


Şekil 9c. Birleşme Ağacı

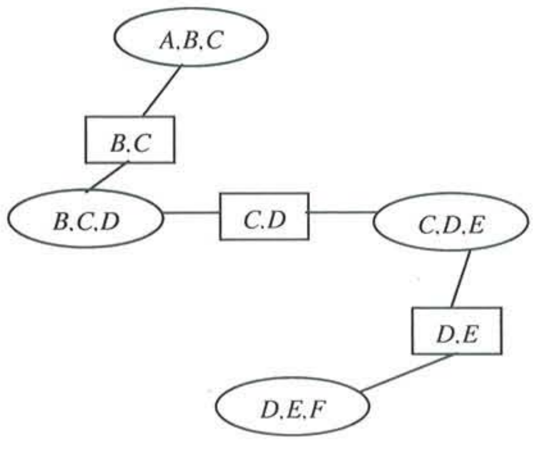
Yukarıda verilen probleme çözüm bulmak için, moral grafiğe başka yapay bağlantılar da eklenir. *C*, *D* ve *B*, *C* arasına bağlantılar eklenerek, birleşme ağacı Şekil 10a, 10b ve 10c'de görülmektedir.



Şekil 10a. Moral ve Üçgen Grafik



Şekil 10b. Birleşme Grafiği



Şekil 10c. Birleşme Ağacı

3. UYGULAMA

Bu çalışmada, Türkiye İstatistik Kurumu Demografi Şubesi tarafından elde edilen, 1996 yılı intihar verileri kullanılmıştır (Türkiye İstatistik Yıllığı, 1996). 1815 veri için intihar şeklini etkileyen nedenler, cinsiyet, medeni durum, intihar nedeni ve meslek grubu alınmıştır. Bu değişkenlerin kategori düzeylerinin sayısı ve düzeylerin tanımı aşağıdaki tabloda verilmiştir (Tablo 1). Ayrıca, kategorilerde gerekli birleştirmeler, CHAID (Chi-Squared Automatic Interaction Detector) analizi ile yapılmıştır (Hill vd., 1997). Burada, SPSS 8.0 for Windows CHAID 6.0 kullanılmıştır.

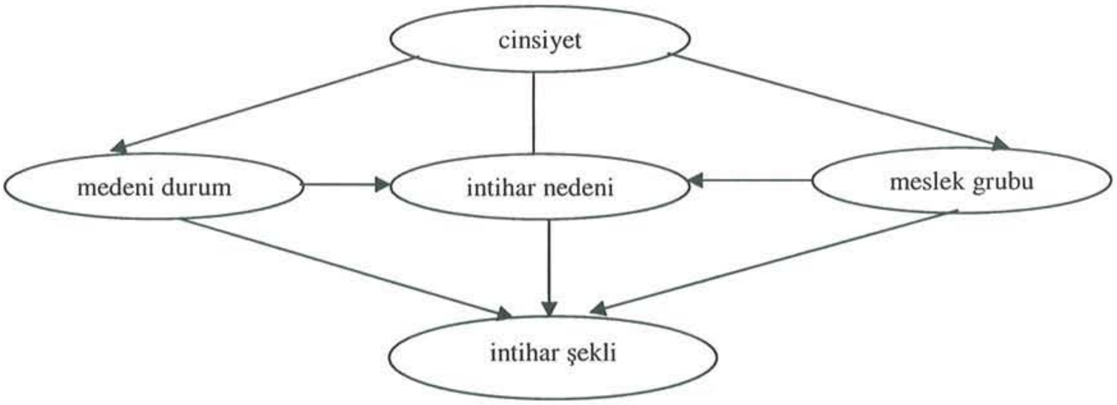
Tablo 1. Değişkenlerin Kategori Düzey Sayısı ve Tanımı

Değişkenler	Düzey sayısı	Düzeylerin tanımı	N
İntihar şekli	3	(1) Kendini asarak	900
		(2) Kimyevi madde kullanarak, havagazı, tüpgaz vb. kullanarak, kendini yakarak	270
		(3) Kendini yüksekten atarak, kendini suya atarak, ateşli silah kullanarak, kesici alet kullanarak, kendini tren vb. altına atarak ve diğer	645
İntihar nedeni	3	(1) Hastalık	664
		(2) Aile geçimsizliği, hissi ilişki ve istediği ile evlenememe	674
		(3) Geçim zorluğu, ticari başarısızlık, öğrenim başarısızlığı, diğer, bilinmeyen	477
Cinsiyet	2	(1) Erkek	1122
		(2) Kadın	693
Medeni durum	3	(1) Hiç evlenmedi	769
		(2) Evli	875
		(3) Eşi öldü ve boşandı	171
Meslek grubu	3	(1) İlmi teknik elemanlar, serbest meslek, idari personel ve benzeri çalışanlar, ticari ve satış personeli	270
		(2) Hizmet işlerinde çalışanlar, tarımcı, hayvancı, ormancı, balıkçı ve avcı, tarım dışı üretim ve ulaştırma makineleri kullananlar	446
		(3) Müteşebbis, direktör, üst kademe yöneticisi, diğer	1099

HUGIN çalıştırıldığında, aşağıda verilen uygun Bayes ağ modeli için, gerekli olan şekiller gösterilmiştir. Bu şekillerde, boş kutular, gözlenmiş değişkenleri, içi dolu kutulara karşılık gelen değerler ise, olasılık değerlerini gösterir. Ayrıca, şekillerin içerisinde yer alan "100,00" değeri, değişkenin seçilen düzeyinin biliniyor olması demektir. Değişkenlerde yapılan kısaltmalar aşağıda verilmiştir:

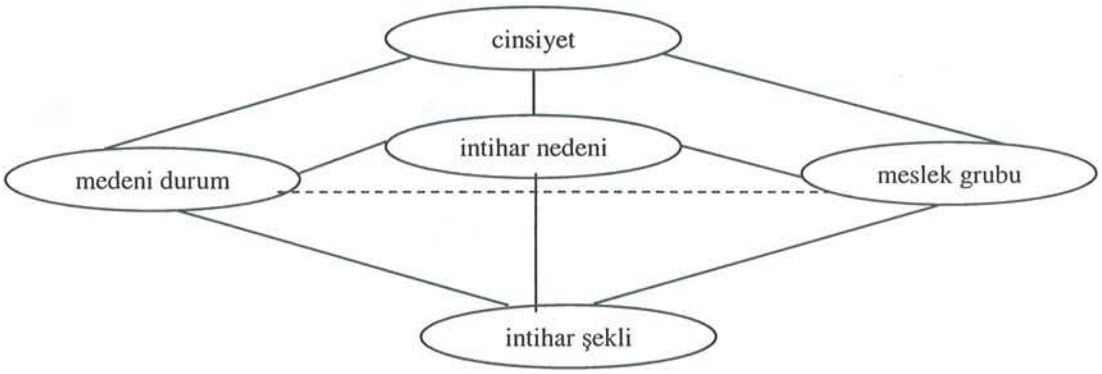
Cinsiyet: C
Medeni durum: MD
İntihar şekli: İŞ
İntihar nedeni: İN
Meslek grubu: MG

İlgilenilen değişkenlere ilişkin Bayes ağ modeli aşağıda verilmiştir (Şekil 11). Bu model, beş değişkenden oluşmaktadır: cinsiyet, medeni durum, intihar nedeni, meslek grubu ve intihar şekli. Bu modelde, cinsiyet, intihar nedeni, medeni durum ve meslek grubu değişkenlerini etkilemektedir. Medeni durum, meslek grubu ve cinsiyet değişkenleri de intihar nedenini etkiler. Ayrıca, medeni durum, intihar nedeni ve meslek grubu değişkenleri de intihar şeklini etkilemektedir.

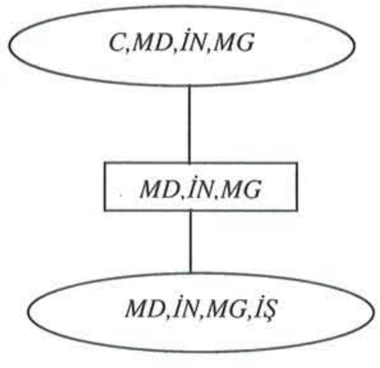


Şekil 11. Beş Düğümden Oluşan Bir Bayes Ağ Modeli

Model belirlendikten ve koşullu olasılıkların girişi yapıldıktan sonra, HUGIN paket programı çalıştırılmış ve modele uygun moral ve üçgen grafiği elde edilmiştir (Şekil 12). Moral grafikten elde edilen takımlar, $\{C,MD,IN,MG\}$ ve $\{MD,IN,MG,IŞ\}$ olmak üzere iki tanedir. Bu takımların arakesitleri olan $\{IN,MD,MG\}$ düğümleri alınarak, birleşme ağacı elde edilmiştir (Şekil 13).

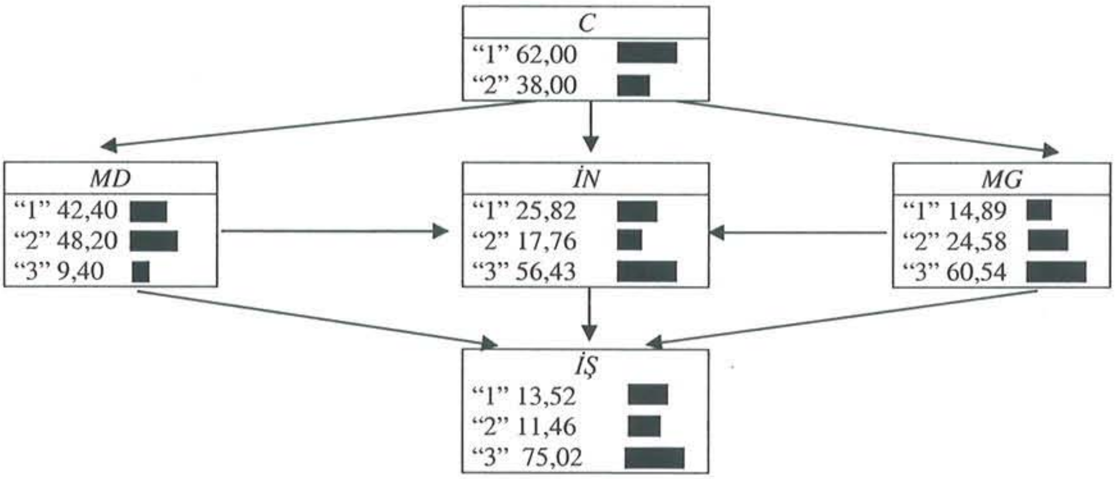


Şekil 12. Şekil 11 için Moral ve Üçgen Grafik



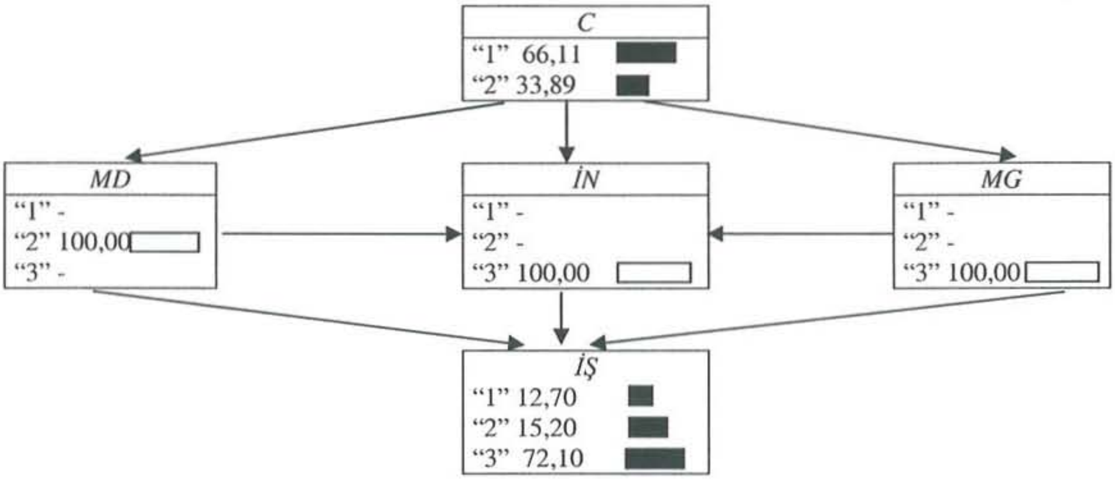
Şekil 13. Şekil 12'den Elde Edilen Birleşme Ağacı

Elde edilen birleşme ağacına göre, {*MD,IN,MG*} değişkenleri bilindiğinde, cinsiyet intihar şeklinden koşullu olarak bağımsızdır ve *C*⊥*İŞ* \ {*MD,İN,MG*} ile gösterilir. Acaba bu değişkenlerin, hangi düzeyleri için yukarıdaki koşul sağlanabilir? {*MD,İN,MG*} değişkenlerinin 27 tane mümkün durumu vardır. Bu mümkün durumların hepsi incelenmiş ve yorumlamaları sonuç ve tartışmalar bölümünde verilmiştir. Ancak, burada sadece tesadüfi olarak seçilen tek mümkün durum detayları ile açıklandı. Bayes ağ modeli için, beş düğüme ilişkin marjinal olasılıklar aşağıda verilmiştir (Şekil 14). Örneğin; 1815 bireyin erkek olma oranı 0,62 iken, kadın olma oranı 0,38'dir. 1815 bireyden tesadüfi seçilen bir bireyin kendini asarak intihar etmiş olma olasılığı 0,1352'dir. Evli olanların oranı ise 0,4820'dir.



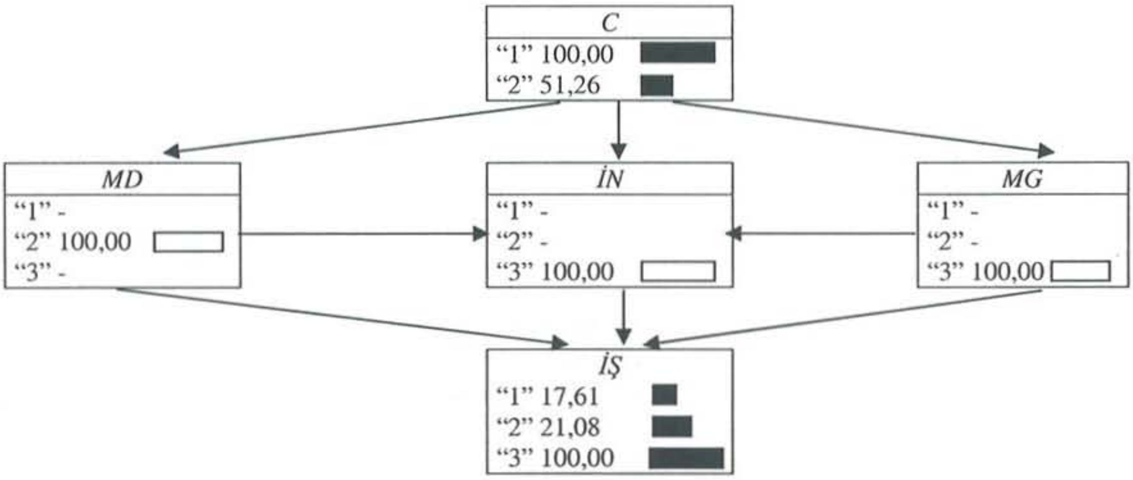
Şekil 14. Bayes Ağ Modeli için Marjinal Olasılıklar

İlk olarak, *İN*="3", *MD*="2" ve *MG*="3" değişkenlerinin düzeyleri bilindiği durum düşünülsün. *İN*="3", *MD*="2" ve *MG*="3" değişkenleri bilindiğinde, diğer değişkenlerin koşullu olasılıkları aşağıda gösterilmiştir (Şekil 15).

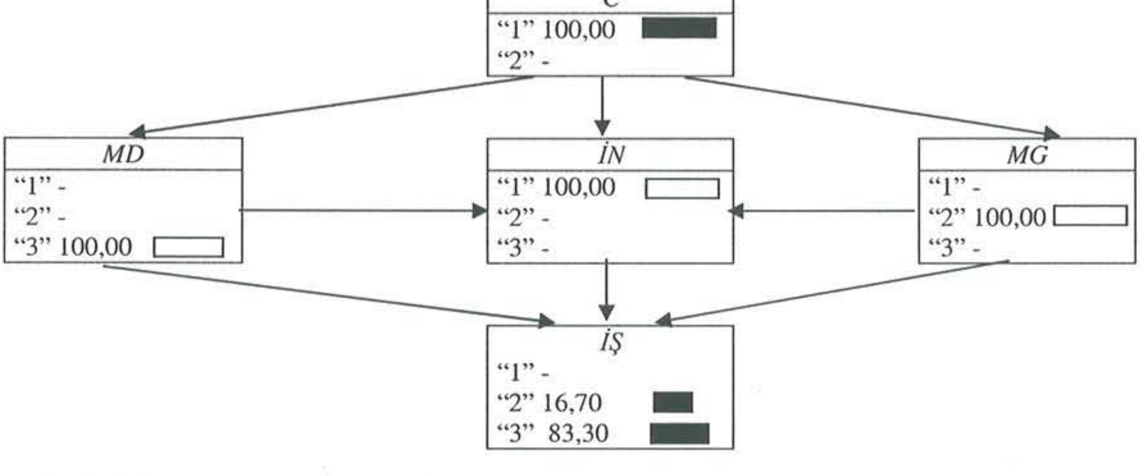


Şekil 15. Bayes Ağ Modeli için, *İN*="3", *MD*="2", *MG*="3" Olaylarının Genişlemesinden Sonra, Diğer Düğümlerin Koşullu Olasılıkları

Şekil 16'da, olaylar genişletildikten sonra, 1815 bireyden tesadüfi olarak seçilen bir bireyin erkek olma olasılığı 0,6611'dir. Bu düğüme ilişkin, ön olasılık 0,62 olduğundan, olaylar gözlendikçe, tesadüfi olarak seçilen bir bireyin erkek olma olasılığında artış görülmektedir. *İN*="3", *MD*="2" ve *MG*="3" olayları bilindiğinde, tesadüfi seçilen bir bireyin kendini yüksekten atarak vb. intihar şeklini seçenlerin olasılığı 0,7210'dur.



Şekil 16. Bayes Ağ Modeli için, *MD*="2", *İN*="3" ve *MG*="3" Olayları Genişletildiğinde, En Yüksek Olasılığa Sahip Düğümlerin Mümkün Birleşimleri



Şekil 17. Bayes Ağ Model, için, İN="1", MD="3", MG="2" Olaylarının Genişlemesinden Sonra, Diğer Düğümlerin Koşullu Olasılıkları

MD="3", *İN*="1" ve *MG*="2" olayları genişletildiğinde, diğer düğümlerin en yüksek olasılıklarının birleşimlerinin değeri 0,00255572'dir ve istenen olasılık,

$$P(C="1", İŞ="3" / MH="3", İN="1", MG="2") =$$
$$\frac{P(C="1", "İŞ="3", MH="3", İN="1", MG="2")}{P(MH="3", İN="1", MG="2")}$$
$$= 0{,}00255572/0{,}0030681$$
$$= 0{,}833$$

elde edilir. *MD*="3", *İN*="1" ve *MG*="2" değişkenlerinin düzeyleri bilindiğinde, *C*="1" ve *İŞ*="3" değişkenlerinin düzeylerini elde etme olasılığı 0,833 olarak elde edilmiştir. Tüm mümkün durumlar için koşullu olasılıklar belirlenebilir.

4. SONUÇ

Bu çalışmada amaç, kesikli değişkenler için Bayes ağları tanıtmak ve olasılık güncelleştirmesi için verilen bir problemin nasıl çözüleceğini ve yorumlanacağını açıklamaktır. Ayrıca, değişkenler arasındaki neden-etki ilişkilerine değinilmiş ve böylece bu değişkenler arasındaki ilişki modellenmiştir. Geleneksel metotlar kullanılarak, Bayes ağ modelleri için, gerekli olasılıklar hesaplanmış ve Bayes ağ modelindeki olaylar genişletilerek, gerekli olan olasılıklar tekrar incelenmiştir. Uygulamada yer alan, intihar verilerinde kullanılan değişkenler için, neden- etki değişkenleri belirlenmiş ve uygun Bayes ağ modeli kurulmuştur. Bu modele göre, etkiler gözlendiği zaman, farklı nedenler için çıkarımlar elde edilmiştir. İntihar örneğinde çıkarımlar ile ilgili bazı sonuçlar şöyledir:

Kurulan modele göre, {MD,İN,MG} değişkenleri verildiğinde {İŞ} ve {C} değişkenlerinin koşullu olarak bağımsız olduğu görülmüştür. {MD,İN,MG} değişkenlerin düzeyleri (27 mümkün durum) incelenmiştir.

Kurulan modele göre, genel olarak, bireylerin daha büyük olasılıkla erkek ve kendini yüksekten atarak, kendini suya atarak vb. intihar şeklini seçen bireyler olduğu gözlenmiştir. Bayes ağ modelinde, {İN="1",MD="1",MG="1"} değişkenlerinin düzeyleri bilindiğinde, cinsiyetinin erkek olma olasılığı %68.86'dır. Ancak, {İN="1",MD="3",MG="2"} değişkenlerinin düzeyleri bilindiğinde, cinsiyetinin erkek olma olasılığı %100.00'dır. {İN="2", MD="3",MG="1"} değişkenlerinin düzeyleri bilindiğinde, modelde cinsiyetin erkek olma olasılığı % 93.70'tir.

Koşullu bağımsızlıkların doğru belirlenmesi olasılıkları etkileyeceği için çok önemlidir. Bundan dolayı, koşullu bağımsızlıkların belirlenmesinde, konu ile ilgili uzmanların fikir dereceleri önemli rol oynar. Ayrıca, araştırmacı koşullarına uygun bir şekilde amacını belirlemelidir. Bu konu literatürde, uzman sistemler adı altında oldukça geniş bir yer kaplar. Modelin belirlenmesinde, uzman sistemlerin görüşleri belirlenmelidir.

KAYNAKLAR

COWELL, R.G., (1999), *Introduction to inference in Bayesian networks, in Learning in Graphical Models,* 9-26.

EDWARDS, D., (1995), *Introduction to Graphical Modelling*, Springler Verlag, New York.

FRIEDMAN, N. AND GOLDSZMIDT, M., (1996), *Learning Bayesian Networks with Local Structure*, Proceedings 12th Conference on Uncertainty in Artificial Intelligence, IFI-95-03109, 421-459.

HILL, D. A., DELANEY, L. M., RONCAL, S., (1997), *A Chi-Square Automatic Interaction Dedection, (CHAID) Analysis of Factors Determining Trauma Outcomes*, Journal of Trauma-Injury Infection and Critical Core, Vol. 42, ISS 1, 62-66.

HUGIN SYSTEMS, (1998), *Introduction to the HUGIN System,* HUGIN Expert Ltd., htttp://www.hugin.dk.

JENSEN, F.V., (1996), *An ıntroduction to Bayesian Networks*, UCL Press Ltd., London.

JENSEN, F.V., OLESEN, K.G. AND ANDERSEN, S.K., (1990), *An Algebra of Bayesian Belief Universes for Knowledge-Based Systems,* Networks, Vol. 20, 637-659.

LAURITZEN, S.L., (1996), *Graphical Models*, Oxford University Press, Oxford.

LAURITZEN, S.L. AND SPIEGELHALTER, D.J., (1988), *Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems*, J.R. Statist. Soc. B, 50(2), 157-224.

LIAROKAPIS, D., (1999), *An Introduction to Belief Networks,* htttp://www.cs.umb.edu/dimitris.

NIEDERMAYER, D., (1998), *An Introduction to Bayesian Networks and Their Contemporary Applications*, http://www.gpfn.sk.ca/~daryle/papers/ bayesian_networks/bayes html.

PEARL, J., (1986), *Fusion, Propagation and Structuring in Belief Nnetworks,* Artificial Intelligence, 29, 241-288.

SPIEGELHALTER, D.J., DAWID, A.P, LAURITZEN, S.L. AND COWELL, R.G., (1993), *Bayesian Analysis in Expert Systems*, Statistical Science, Vol.8, No.3, 219-283.

SPIEGELHALTER, D.J. AND LAURITZEN, S.L., (1990), *Sequetial Updating of Conditional Probabilities on Directed Graphical Structures*, Networks, 20, 579-605.

STEPHENSON, T.A., (2000), *An Introduction to Bayesian Network Theory and Usage,* IDIAP Research Report 00-03.

TÜRKİYE İSTATİSTİK YILLIĞI, (1996), *Demografi Şubesine Ait İntihar Verileri,* Devlet İstatistik Enstitüsü, Ankara.

PROPAGATION OF EVIDENCE USING JUNCTION TREE IN BAYESIAN NETWORKS

ABSTRACT

A Bayesian network is directed a cyclic graphs in which the nodes represent variables and the edges signify direct dependencies between variables. The strenghts of these edges dependencies are defined by conditional probabilities. In Bayesian networks, the evidence propagation is introduced by junction tree algorithm. Finally, strenghts of conditional probabilities are calculated by using application data and then the probabilities are examined. Here, SPSS software is used to constitute contingency tables and HUGIN software for probability calculus.

Key Words: *Bayesian Networks, Conditional Independence, Conditional Probability, Junction Tree.*

İstatistik Araştırma Dergisi Cilt 04 No 03 Sayfa 45-51 Aralık 2005

AYKIRI GÖZLEM SAYISININ BELİRLENMESİ

Ufuk EKİZ * Müslim EKNİ *

ÖZET

Clarke and Lewis (1998)'e göre aykırı gözlem, diğer gözlemlerle aynı merkezi parametreli fakat farklı varyanslı benzer bir dağılım tarafından üretilir. Bu tanımdan hareketle Wen-Liang Hung, Jong-Wuu Wu (2005), örnekteki aykırı gözlemlerin sayısını hata kareler toplamını en küçükleyerek belirleyen bir yöntem önermişlerdir. Bu yöntem Clarke ve Lewis tarafından tanımlanan R istatistiğine göre daha basit ve kolay hesaplanabilir. Ayrıca normal dağılımdan geldiği düşünülen örnekteki alt ve üst aykırı gözlemlerin sayısını belirlemekte gizleme ve yanılgıya-düşürme problemlerinden etkilenmediği söylenmektedir. Bu çalışmada, yöntemin ne kadar sağlıklı sonuçlar verdiğini görmek için bir simülasyon çalışması yapılmıştır. Sonuçlar örnek çapı büyüdükçe aykırı gözlem sayısını doğru belirleme oranının düştüğünü gösterdiğinden, yöntem gizleme ve yanılgıya-düşürme problemlerinden etkilenmektedir.

Anahtar Kelimeler : *Aykırı Gözlem, En Küçük Kareler, Gizleme, Monte Carlo, Sıra İstatistiği, Yanılgıya-Düşürme.*

1. GİRİŞ

Son yıllarda örnekte yer alabilecek aykırı gözlemlerin hangisi olduğu ya da bunların sayısının belirlenmesi problemleri ile pek çok yazar ilgilenmiştir (Guttman, I. (1973b), Pearson ve Sekar (1936), Cook ve Weisberg. (1982)). Aykırı gözlemlerle ilgili yapılmış çalışmaların kapsamlı bir özeti Barnett ve Lewis (1994)'te yer almaktadır.

Bu çalışmada, Clarke ve Lewis tarafından tanımlanmış olan problemle ilgilenilecektir. Bu problem aşağıdaki gibi tanımlanabilir.

Örneğin k tane üst aykırı gözlem içerdiği varsayılıyor olsun, tesadüfi örneğine ilişkin yokluk hipotezi,

* Gazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Teknikokullar, Ankara, TÜRKİYE, e-mail: ufukekiz@gazi.edu.tr, mekni@gazi.edu.tr

$$H_0 = \chi_I \approx N(\mu, \sigma^2 / \lambda_i), \quad i=1,2,\dots, n \tag{1.1}$$

şeklinde tanımlanmaktadır. Burada μ ve σ^2 sırasıyla konum ve yayılım parametrelerini, λ_i ise yayılımdaki değişim (shift in dispersion) parametresini ifade etmektedir. Karşıt hipotezde $X_1, X_2, ..., X_n$ tesadüfi örneğinde yer alan k tane tesadüfi değişkenin merkezi parametreleri birbirlerinden ve μ'den farklı normal dağılıma sahip olduğu şeklinde tanımlanmaktadır. Yani n çaplı örnekte yer alan gözlemlerden k tanesinden her biri, ortalamaları n-k tane gözlemin geldiği dağılımın ortalaması olan μ'den ve diğerlerinin sahip olduğu dağılımın ortalamasından farklı bir ortalamaya sahip normal dağılımdan geldiği şeklinde tanımlanmaktadır. Diğer bir ifadeyle karşıt hipotez,

$$H_k = \chi_r \approx N(\mu, \sigma^2 / \lambda r) \qquad \chi_t \approx N(\upsilon t, \sigma^2 / \lambda_t) \tag{1.2}$$

şeklinde tanımlanmaktadır. Burada r ve t indisleri r=1,2,...,n-k , t=1,2,...k değerlerini almaktadır (bu çalışmada örnekte yer alan gözlemlerden en büyük k tanesinin aykırı olabileceği önem taşımaktadır.). Ayrıca μ, σ, υ_t ve k $(0 \le k \le n - n/2 - 1)$ 'nın bilinmediği λ_r ve λ_t 'nin ise bilindiği varsayılmaktadır.

$X_1, X_2, ..., X_n$ tesadüfi örneğinin içerebileceği aykırı gözlem sayısını belirlemekte sıra istatistiklerinden faydalanarak yukarıdaki probleme çözüm getirebilecek en küçük karelere dayalı bir yöntem Wen-Liang Hung ve Jong-Wuu Wu (2005) tarafından ileri sürülmektedir. Bu yöntem aşağıdaki gibi tanımlanabilir. $X_1 \le X_2 \le ... \le X_n$, $X_1, X_2, ..., X_n$ tesadüfi örneğine ilişkin sıra istatistikleri olmak üzere, $X_{(n-k+1)}, X_{(n-k+2)}, ..., X_{(n)}$ sıra istatistiklerinin dağılımları tarafından üretilmiş $x_{(n-k+1)}, x_{(n-k+2)}, ..., x_{(n)}$ gözlem değerlerinin k tane üst aykırı gözlem olduğu varsayılsın. Bu durumda, hata kareler toplamının en küçüklenmesine dayalı en küçük kareler yöntemi ile k'nın değerinin ne olabileceğine karar verilebilir. Bu yöntemin uygulanışı diğer yöntemlere göre çok daha basit ve kolaydır. Ayrıca gizleme ve yanılgıya düşürme problemlerinden de etkilenmediği düşünülmektedir. Üst aykırı gözlem sayısının belirlenmesinin yanı sıra alt aykırı gözlem sayısının belirlenmesi veriye negatif dönüşümün uygulanması ile mümkün olur. Bölüm 2'de en küçük karelere dayalı olarak k'nın değerinin belirlenmesine ayrıntılı olarak yer verilecektir. Son bölümde simülasyon çalışması ile yöntemin doğru k sayısını belirleme oranı üzerinde durulacak ve bu oranın örnek çapının büyüklüğünden ve parametre tahmin değerlerinden nasıl etkilendiği tartışılacaktır.

2. EN KÜÇÜK KARELER

$X_1, X_2, ..., X_n$, Bölüm 1 'deki gibi tanımlanmış k tane aykırı gözlemi içeren tesadüfi bir örnek ve

$$\hat{Z}(X_{(i)}) = \frac{\sqrt{\lambda_i}\left(X_{(i)} - \mu\right)}{\sigma} \quad , \quad i = 1,2,\ldots,n-k \tag{2.1}$$
$$\hat{Z}(X_{(i)}) = \frac{\sqrt{\lambda_i}\left(X_{(i)} - \upsilon_i\right)}{\sigma} \quad , \quad i = n-k+1, n-k+2,\ldots,n$$

şeklinde bir tesadüfi değişken tanımlanmış olsun. Burada λ_i ve υ_i sırasıyla, $X_{(i)}$ sıra istatistiğinin dağılımına ilişkin yayılımdaki değişim ve konum parametrelerdir. F standart normal dağılımın birikimli dağılım fonksiyonu olmak üzere, $F^{-1}(i/(n+1))$, $i = 1,2,\ldots,n$, $\hat{Z}(X_{(i)})$ istatistiğine ilişkin dizinin yakınsama noktası olarak değerlendirilmektedir. En küçük karelerden hareketle k'nın değerine karar vermekte aşağıdaki kriterden yararlanılmaktadır..

$$\Phi_k\left(\mu,\sigma,\upsilon_{n-k+1},\ldots,\upsilon_n\right) = \sum_{i=1}^{n-k}\left\{F^{-1}\left(i/(n+1)\right) - \frac{\sqrt{\lambda_i}\left(X_{(i)} - \mu\right)}{\sigma}\right\}^2 + \sum_{i=n-k+1}^{n}\left\{F^{-1}\left(i/(n+1)\right) - \frac{\sqrt{\lambda_i}\left(X_{(i)} - \upsilon_i\right)}{\sigma}\right\}^2 \tag{2.2}$$

ve

$$\Phi_0\left(\mu,\sigma,\upsilon_{n-k+1},\ldots,\upsilon_n\right) = \sum_{i=1}^{n}\left\{F^{-1}\left(i/(n+1)\right) - \frac{\sqrt{\lambda_i}\left(X_{(i)} - \mu\right)}{\sigma}\right\}^2 \tag{2.3}$$

olmak üzere, $\hat{\Phi}_k\left(\hat{\mu},\hat{\sigma},\hat{\upsilon}_{n-k+1},\ldots,\hat{\upsilon}_n\right)$ 'nın en küçüklenmesi ile üst aykırı gözlem sayısı olan k' nın optimal çözümünü elde edebiliriz ($\Phi_k\left(\mu,\sigma,\upsilon_{n-k+1},\ldots,\upsilon_n\right)$'nın en küçüklenmesi ile $\mu,\sigma,\upsilon_{n-k+1},\ldots,\upsilon_n$ 'nin en küçük kareler tahminleri sırasıyla $\hat{\mu},\hat{\sigma},\hat{\upsilon}_{n-k+1},\ldots,\hat{\upsilon}_n$ olur). Yani $\hat{\Phi}_k\left(\hat{\mu},\hat{\sigma},\hat{\upsilon}_{n-k+1},\ldots,\hat{\upsilon}_n\right)$ k 'nın her değeri için elde edilecek ve en küçük $\hat{\Phi}_k\left(\hat{\mu},\hat{\sigma},\hat{\upsilon}_{n-k+1},\ldots,\hat{\upsilon}_n\right)$ değerinin elde edildiği k örneğin içerdiği üst aykırı gözlem sayısı olarak nitelenecektir.

Eğer X tesadüfi değişkeni normal dağılıma sahipse, (-X) tesadüfi değişkeni de normal dağılıma sahiptir. Bundan dolayı örneğin alt aykırı gözleme sahip olması durumunda, örneğe negatif bir dönüşüm uygulayarak yine aynı yöntemin uygulanması ile üst aykırı gözlem sayısı belirlenebilir. (-X) tesadüfi değişkeni üzerinden belirlenmiş üst aykırı gözlem sayısı, X tesadüfi değişkeni için alt aykırı gözlem sayısı anlamına gelmektedir. Dolayısıyla örnekte yer alabilecek üst ve alt aykırı gözlemlerin sayısına ilişkin bir tahmine, bu yöntem ile ulaşılabilir.

3. BENZETİM ÇALIŞMASI

Yöntemin örnekte yer alan aykırı gözlemlerin sayısını hangi oranda doğru olarak belirlediğini ortaya koymak amacına uygun bir Monte Carlo deney düzeni hazırlanmıştır. Bu deney düzenine göre, n çaplı bir örnek k tane üst aykırı gözlem içerecek şekilde normal dağılımdan üretilmektedir. Bu deney düzeninde n 'nin 10, 30, 50 ve 100 değerlerinden her biri için, k=0 (hiç üst aykırı gözlem olmaması), k=1 (örneğin bir tane üst aykırı gözlem içermesi), k=2 (örneğin iki tane üst aykırı gözlem içermesi), durumlarından her biri ayrı ayrı incelenmektedir. Örneğin, n=30 ve k=1 olması durumu için 500 adet örnek üretilecek ve birer tane üst aykırı gözlem içerdiği bilinen bu 500 örnekten kaç tanesinde, yöntemin gerçekten de bir tane üst aykırı gözlem tespit ettiği belirlenecektir. Böylece yöntemin üst aykırı gözlemlerin sayısını doğru olarak belirleme oranı elde edilecektir. Üretilecek örnekte yer alacak gözlemler ve k tane üst aykırı gözlem aşağıdaki gibi tanımlanmaktadır.

k=0 durumu için,

$\chi_i \approx N(1,1/\lambda_i)$, $\quad$ i = 1,2...,n

k=1 durumu için,

$\chi_i \approx N(1,1/\lambda_i)$ $\quad$ i = 1,2,..., n-1

$\chi_n \approx N(\upsilon_n,1/\lambda_i)$

k=2 durumu için,

$\chi_i \approx N(1,1/\lambda_i)$ $\quad$ i = 1,2,..., n-2

$\chi_{n-1} \approx N(\upsilon_{n-1},1/\lambda_{n-1})$ $\quad$ $\chi_n \approx N(\upsilon_n,1/\lambda_i)$, $\quad$ j=n-1, n

Farklı υ_j ve λ_j değerleri için, Monte Carlo simülasyon tekniğinin (tekrar sayısı 500) uygulanması ile yöntemin üst aykırı gözlem sayısı k'yı doğru belirleme oranları Tablo 1, Tablo 2, Tablo 3 ve Tablo 4' te yer almaktadır.

Tablo1. $\upsilon_n = 2$, $\lambda_1 = \lambda_2 = ... = \lambda_{n-1} = 5$, $\lambda_{n-1}, \lambda_n = 10$ için k' nın doğru belirlenme oranları

N	k=0	k=1	k=2
10	1	0.9458	1
30	1	0.7640	0.9380
50	1	0.7040	0.8220
100	1	0.5920	0.6640

Tablo 2. $\upsilon_n = 2$, $\lambda_1 = \lambda_2 = ... = \lambda_{n-1} = 5$, $\lambda_{n-1}, \lambda_n = 5$ için
k' nın doğru belirlenme oranları

n	k=0	k=1	k=2
10	1	0.8720	0.9760
30	1	0.7560	0.8715
50	1	0.6400	0.7640
100	1	0.5840	0.6980

Tablo 3. $\upsilon_n = 3$, $\lambda_1 = \lambda_2 = ... = \lambda_{n-1} = 5$, $\lambda_{n-1}, \lambda_n = 5$ için
k' nın doğru belirlenme oranları

n	k=0	k=1	k=2
10	1	1	1
30	1	1	1
50	1	0.9940	1
100	1	0.9960	1

Tablo 4. $\upsilon_n = 1.5$, $\lambda_1 = \lambda_2 = ... = \lambda_{n-1} = 5$, $\lambda_{n-1}, \lambda_n = 5$ için
k' nın doğru belirlenme oranları

n	k=0	k=1	k=2
10	0.9960	0.4920	0.5900
30	1	0.2960	0.2687
50	0.9980	0.1746	0.1701
100	1	0.1620	0.0924

4. SONUÇ

Tablo 1, Tablo 2, Tablo 3 ve Tablo 4'te yer alan sonuçlara bakıldığında, λ_i ve υ_n'nin farklı değerleri için örnek çapı arttıkça örnekte yer alan üst aykırı gözlem sayısı k'nın ileri sürülen yöntem tarafından doğru belirlenmesi oranı düşmektedir. Bunun sebebi örnekte bir tane aykırı gözlem varken, bu gözlemin $\hat{\Phi}_k(\hat{\mu}, \hat{\sigma}, \hat{\upsilon}_{n-k+1}, ..., \hat{\upsilon}_n)$ üzerindeki etkisinin örnek çapının büyüklüğüne bağlı olarak azalmasıdır. Bu durum ise yöntemin gizleme probleminden etkilenmediği söylemiyle ters düşmektedir. Yani gerçekte aykırı bir gözlem içeren büyük çaplı bir örnek üzerinde bu yöntemin uygulanması sonucunda, gerçekte aykırı olan gözlemin aykırı olmayan n-1 gözlem tarafından gizlenmesi ihtimali yükselmektedir. Ayrıca Tablo 3 ve Tablo 4'te yer alan sonuçlar karşılaştırıldığında, yöntem $\hat{\upsilon}_n$'nin $\mu=1$ değerinden uzak değerleri için yüksek, yakın değerler için düşük doğru belirleme oranı vermektedir. Bu ise yöntemi, aykırı gözlemin geldiği düşünülen dağılımın parametrelerinin, μ ve σ^2/λ_i büyük ölçüde farklı olmasına bağlı kılmaktadır. $\hat{\upsilon}_n$'nın μ'den çok büyük değerleri için örnek çapı çok büyük olsa da yöntemin doğruyu belirleme oranı yüksek olacaktır. Tablo 1, Tablo 2, Tablo 3 ve Tablo 4'te k=2 durumuna ilişkin sonuçlar, aykırı gözlem sayısının örnek çapına oranı yüksek oldukça, yöntemin doğru sonuç vermesi oranının yüksek olacağını göstermektedir.

KAYNAKLAR

BARNETT V., LEWIS T. (1994). *Outlier in Statistical Data, third ed., Wiley*, Chichester.

CLARKE B. R., LEWIS T., (1998). *An Outlier Problem in the Determination of Ore Grade, Journal of Applied Statistics* 25, 751-762.

COOK, R. D., and WEISBERG, S. (1982). *Residuals and Influence in Regression.* Chapman and Hall, New York.

GUTTMAN, I. (1973b). *Care and Handling of Univariate or Multivariate Outliers in Detecting Spuriosity- a Bayesian Approach. Technometrics*, 15, 723-738.

PEARSON, E. S, and CHANDRA SEKAR, C. (1936). *The Efficieny of Statistical Tools and a Criterion Forthe Rejection of Outling Observations. Biometrika*, 28, 308-320.

WEN-LIANG HUNG, JONG-WUU WU, (2005). *A Note on Determining the Number of Outliers in a Normal Sample with Unequal Variances by Least Squares Procedure. Applied Mathematics and Computation* 162 ,1007-1012.

DETECTING THE NUMBER OF OUTLIERS

ABSTRACT

According to Clarke and Lewis (1998) an outlier observation is generated by a similar distribution as of other observations, with the same location parameter but with different variance. Starting with this definition, Wen-Liang Hung, Jong-Wuu Wu (2005) have proposed a method based on minimizing square root error to determine the number of outliers in the sample. This method is more advantages compared to R statistic defined by Clarke and Lewis for its calculation is more simple and easier. Furtermore, it is said that this method is not affected from masking and swamping problems, in determining the number of lower and upper outlier observations in the sample generated from normal distribution. In this work, a simulation study is performed to detect how reliable results obtained by the method. The method is affected by the masking and swamping problems for the results showed that as the sample size is increased, the ratio of accurately determining the number of outlier observations decreases.

Key Words : *Least Square Method, Masking, Monte Carlo, Order Statistics, Outlier Observations, Swamping.*

İstatistik Araştırma Dergisi Cilt 04 No 03 Sayfa 52-62 Aralık 2005

REGRESYON ÇÖZÜMLEMESİNDE KAYIP VERİ SORUNU

Neslihan DEMİREL* Serdar KURT*

ÖZET

Kayıp veri çözümlemesinin konusu veri matrisindeki bazı değerlerin gözlenmemiş olmasıdır. Kayıp veri çözümlemesi özellikle uygulamalı istatistiğin çok önemli konularından birini oluşturmaktadır. Kayıp veriyi yok saymak, örneklemin rastgeleliğini bozarak yanlı parametre tahminleri elde edilmesine neden olabilmektedir. Regresyon analizi, tahmin amaçlı kullanılan önemli çok değişkenli istatistiksel analizlerin başında gelmektedir. Bu nedenle bu çalışmada, regresyon analizinde, bağımsız değişkenlerde kayıp veri mekanizması rassal kayıp (MAR) olacak şekilde, regresyon analizi varsayımlarının sağlandığı ve sağlanmadığı iki ayrı veri seti üzerinde benzetim çalışması yapılmıştır. Kayıp veri göz ardı edilebilir olduğunda model esaslı yöntemler arasında yer alan, EM algoritması ve çoklu atıf yöntemleri karşılaştırmalı olarak incelenmiştir. EM algoritmasının regresyon analizi varsayımlarının bozulmasından etkilenmediği, ancak çoklu atıf için, atıf sayısının arttırılması gerektiği sonucu elde edilmiştir.

Anahtar Kelimeler: *Çoklu Atıf EM Algoritması, Kayıp Veri, Regresyon Analizi.*

1. GİRİŞ

Çok değişkenli bir kitleden alınan örnekleme ait veri matrisinde gözlemler satırlarla, değişkenler sütunlarla temsil edilir. Kayıp veri çözümlemesinin konusu ise veri matrisindeki bazı değerlerin gözlenmemiş olmasıdır.

Kayıp veriyi yok saymak, örneklemin rastgeleliğini bozarak, yanlı tahminler elde edilmesine neden olabilmektedir. Kayıp veri çözümlemesi için geliştirilen yöntemlerden hangisinin uygun olduğuna karar verebilmek için, kayıp veri oluşumuna neden olan etkenlerin incelenerek, kayıp veri mekanizmasının bulunması gerekmektedir. Kayıp veri mekanizması, kayıp verilerin farklı nedenlerle ortaya çıkmasına bağlı olarak üç şekilde sınıflandırılır: tamamen rassal kayıp (missing completely at random – MCAR), rassal kayıp (missing at random – MAR) , rassal olmayan kayıp (not missing at random - NMAR).

* Dokuz Eylül Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Buca-İZMİR, TÜRKİYE
neslihan.ortabas@deu.edu.tr - serdar.kurt@deu.edu.tr

Y, nxK boyutlu tam veri matrisi olarak tanımlansın, $Y = (y_{ij})_{nxK}$. Burada $i = 1,...,n$ ile gözlem indisini, $j = 1,...,K$ ise değişken indisini temsil eder. Y'nin i. satırı $y_i = (y_{i1},...,y_{iK})$ içinde yer alan y_{ij}, Y_j değişkeninin i. gözlemine ait değeri gösterir. R, nxK boyutlu kayıp veri gösterge matrisi olarak tanımlansın, $R = (r_{ij})_{nxK}$. y_{ij} kayıp ise $r_{ij} = 1$, y_{ij} gözlenmiş ise $r_{ij} = 0$ değerlerini alır. Kayıp veri mekanizması, Y verildiğinde R'nin koşullu dağılımı, $f(R \mid Y, \phi)$, olarak tanımlanır. Bu fonksiyonda ϕ bilinmeyen parametreleri temsil eder. Üç farklı şekilde sınıflandırılan kayıp veri mekanizması bu tanımlamalardan hareketle aşağıdaki gibi özetlenir.

MCAR: Bir gözlemin kayıp değeri, gözlenenler ve kayıp gözlemlerin değerlerine bağlı değilse, tüm Y ve ϕ'ler için $f(R \mid Y, \phi) = f(R \mid \phi)$'dır.

MAR: $Y_{gözlenen}$, Y'nin gözlenen bileşenlerini, $Y_{kayıp}$ ise kayıp bileşenlerini göstersin. Bir gözlemin kayıp değeri, gözlenenlere bağlı, kayıp gözlemlerin değerlerine bağlı değilse, tüm $Y_{kayıp}$ ve ϕ'ler için $f(R \mid Y, \phi) = f(R \mid Y_{gözlenen}, \phi)$'dır.

NMAR: Bazı birimleri kayıp olan tek değişkenli rassal örneklem, bir veri seti olarak ele alınsın. $Y = (y_1,...,y_n)'$ gösteriminde rassal değişkenin i. birimin aldığı değer y_i'dir. Bu durumda $R = (R_1,...,R_n)$ gösteriminde birimler gözlenmiş ise $R_i = 0$, birimler gözlenmemiş ise $R_i = 1$ değerini alır. (y_i, R_i)'nin bileşik dağılımının karşılıklı birimler için bağımsız olduğunu varsayalım, böylece özellikle gözlenen birimin olasılığı diğer birimler için Y'nin ya da R'nin değerlerine bağlı olmaz. O zaman $f(Y, R \mid \theta, \phi) = f(Y \mid \theta) f(R \mid Y, \phi) = \prod_{i=1}^{n} f(y_i \mid \theta) \prod_{i=1}^{n} f(R_i \mid y_i, \phi)$ elde edilir. Burada θ bilinmeyen parametreleri temsil eder ve y_i'nin yoğunluğu $f(y_i \mid \theta)$ ile gösterilir ve ikili gösterge R_i için Bernoulli dağılımının yoğunluğu $f(R_i \mid y_i, \phi)$'dir. y_i kayıp ise olasılık $Pr(R_i = 1 \mid y_i, \phi)$ ile gösterilir. Eğer kayıp değer Y'den bağımsız ise olasılık $Pr(R_i = 1 \mid y_i, \phi) = \phi$ bir sabite eşit, bu sabit y_i'den bağımsız olduğundan, kayıp veri mekanizması MCAR olur. Eğer y_i'nin bazı kayıp değerlerine bağlı çıkarsa NMAR olur. (Little ve Rubin, 2002).

Kayıp veri oluşumuna neden olan etkenler incelenerek bulunan kayıp veri mekanizmasına hangi çözümlemenin yapılacağına karar vermek için Afifi ve Elashoff (1966), Hartley ve Hocking (1971), Orchard ve Woodbury (1972), Dempster, Laird ve Rubin (1977), Little ve Rubin (1983a), Little ve Schenker (1994) ve Little (1997) gibi araştırmacılara ait çalışmalar incelendiğinde, kayıp veri çözümlemesi yöntemlerini 4 ana başlık altında sınıflamak mümkündür: Tamamen kayıtlanmış birim esaslı yöntemler (procedures based on completely recorded units), ağırlıklandırılmış yöntemler (weighting procedures), atıf esaslı yöntemler (imputation-based procedures) ve model esaslı yöntemler (model-based procedures) (Little ve Rubin, 2002).

Tamamen kayıtlanmış birim esaslı yöntemde bazı değişkenlerin bazı birimleri kayıtlanmamış ise kayıtlanmamış birimlere ait satırları tüm değişkenlerden kaldırarak, çözümlemeye geri kalan tam veri seti ile devam edilir. Çözümlemeden birim çıkarmak oldukça kolay bir yöntem olmasına rağmen bu durumda parametre kestirimleri önemli yanlılık gösterebileceğinden, yöntemin etkili olmadığı belirtilmektedir (Little ve Rubin, 2002).

Ağırlıklandırılmış yöntem daha yaygın olarak anket çalışmalarında cevaplamamadan kaynaklanan kayıp gözlemler için kullanılmaktadır. Örneklemin alt kümelerinde gözlenen değerlere ağırlıklar verilerek, kayıp birimler için her alt kümenin ağırlığına karşılık gelecek şekilde değerler atanır (Little ve Rubin, 2002).

Atıf esaslı yöntemde, kayıp gözlemlerin yeri doldurularak, tamamlanan veri seti üzerinde standart yöntemleri uygulama esasına dayanır. Hot deck atfı, regresyon atfı, ortalama atfı, vb. olmak üzere farklı atıf yöntemleri vardır. Uygun atıf yönteminin seçilmesinde dikkatli karar vermek gerekir (Little ve Rubin, 2002).

Model esaslı yöntem, kısmen kayıp veri seti için bilinmeyen parametrelerin tahmin edilmesinde en çok olabilirlik tahmin yöntemi gibi yöntemlerin kullanılması esasına dayanır. Bu yöntemin avantajı esnek olması, model varsayımları altında sonuçların sergilenip, değerlendirilmesidir (Little ve Rubin, 2002).

2. KAYIP VERİLER İÇİN GELİŞTİRİLEN ÇÖZÜM YÖNTEMLERİ

Kayıp veri mekanizması belirlendikten sonra, 1. bölümde bahsedilen 4 ana başlık altında toplanan, kayıp veri için geliştirilen çözüm tekniklerinden uygun olan kullanılır. Bu yöntemler, tam gözlemlerin kullanılması (listwise data deletion), gözlemlerin ya da değişkenlerin silinmesi (casewise data deletion), çiftler bazında veri silme (pairwise data deletion), yerine ortalamayı koyma (mean substution), regresyon atfı (regression imputation), hot deck atfı (hot deck imputation), beklenti maksimizasyonu algoritması-em algoritması (expectation maximization algorithm) ve çoklu atıf (multiple imputation) olarak sıralanabilir.

Bu çalışmada ise kayıp veri mekanizması rassal kayıp (MAR) olan veri seti üzerinde çalışılmıştır. Kayıp veri göz ardı edilebilir olduğunda model esaslı yöntemler arasında yer alan ve bu çalışmada kullanılan EM algoritması ve çoklu atıf yöntemlerine kısaca değinilmiştir.

2.1 EM Algoritması

EM algoritması özellikle verinin bir kısmı kayıp olduğunda tahmin amaçlı kullanılan genel iteratif bir algoritmadır. Her bir iterasyon iki adımdan oluşur. Beklenti ve maksimizasyon. Bu adımlar aşağıdaki gibi tekrarlanır.

1. Kayıp değerler, tahmin değerleri ile yenilenir.
2. Parametreler tahmin edilir.
3. Yeni parametre tahminlerinin doğru olduğu varsayımı altında, kayıp değerler yeniden tahmin edilir.
4. Parametreler yeniden tahmin edilir

Bu iterasyon, ardışık iterasyondan elde edilen tahminlerin arasındaki fark önemsenmeyecek biçimde azalana kadar devam eder. Ortalama, standart sapma, korelasyon, kovaryans matrisi gibi özet istatistikler, standart doğrusal model yazılımlarından elde edilir (Schafer, 1997)

Çok değişkenli normal dağılan ve göz ardı edilebilen kayıp veri mekanizmasına sahip, kayıp gözlemleri bulunan bir örneklemin ortalama ve kovaryans matrisi en çok olabilirlik tahminleri ile elde edilir.

$Y=(Y_1,Y_2\ldots,Y_K)$, ortalaması $\mu=(\mu_1,\mu_2,\ldots,\mu_K)$ ve kovaryans matrisi $\sum=(\sigma_{jk})$ olan K-değişkenli normal dağılsın. $Y=(Y_{gözlenen},Y_{kayıp})$ yazıldığında gözlenen değerler seti, $Y_{gözlenen}$ ile, kayıp değerler seti $Y_{kayıp}$ ile gösterilir. $Y_{gözlenen\cdot}=(y_{gözlenen,1},y_{gözlenen,2},\ldots,y_{gözlenen,n})$ gösteriminde $y_{gözlenen,i}$: i. gözlem için gözlenen değişkenler setini temsil eder. $(i=1,\ldots,n)$

Gözlenen değerler için olabilirlik fonksiyonu;

$$l(\mu,\Sigma \mid Y_{gözlenen})=sabit-\frac{1}{2}\sum_{i=1}^{n}ln\left|\Sigma_{gözlenen,i}\right|-\frac{1}{2}\sum_{i=1}^{n}(y_{gözlenen,i}-\mu_{gözlenen,i})'\Sigma_{gözlenen,i}^{-1}(y_{gözlenen,i}-\mu_{gözlenen,i}) \tag{1}$$

burada $\mu_{gözlenen,i}$ ve $\Sigma_{gözlenen,i}$ Y'nin gözlenen bileşenlerinin i. gözlemi için ortalama ve kovaryans matrislerini temsil eder.

(1) eşitliğini EM algoritmasında en büyüklemek üzere, Y'nin kayıp gözlem içermediği varsayımı altında, Y üssel ailesi için Y_j ve Y_k $(j,k=1,\ldots k)$ rassal değişkenlerine ait yeterli istatistikleri

$$S=\left(\sum_{i=1}^{n}y_{ij};\ j=1,\ldots,K;\ \sum_{i=1}^{n}y_{ij}y_{ik};\ j,k=1,\ldots,K\right) \tag{2}$$

dir. t. iterasyonda EM, parametre tahmini için $\theta^{(t)}=(\mu^{(t)},\Sigma^{(t)})$ ile gösterilir. Algoritma E adımda

$$E(\sum_{i=1}^{n} y_{ij} \mid Y_{gözlenen}, \theta^{(t)}) = \sum_{i=1}^{n} y_{ij}^{(t)}, \quad j = 1,...,K$$

ulaşır ve

$$E(\sum_{i=1}^{n} y_{ij} y_{ik} \mid Y_{gözlenen}, \theta^{(t)}) = \sum_{i=1}^{n} (y_{ij}^{(t)} y_{ik}^{(t)} + c_{jki}^{(t)}); \quad j,k = 1,...,K$$

olur. Burada

$$y_{ij}^{(t)} = \begin{cases} y_{ij} & , y_{ij} \text{ gözlenmis ise;} \\ E(y_{ij} \mid y_{gözlenen,i}, \theta^{(t)}) & , y_{ij} \text{ kayıp ise;} \end{cases} \tag{3}$$

ve

$$c_{jki}^{(t)} = \begin{cases} 0 & , y_{ij} \text{ veya } y_{ik} \text{ gözlenmis ise;} \\ \text{cov}(y_{ij}, y_{ik} \mid y_{gözlenen,i}, \theta^{(t)}) & , y_{ij} \text{ ve } y_{ik} \text{ kayıp ise;} \end{cases} \tag{4}$$

olarak gösterilir.

Kayıp y_{ij} değerleri, $y_{gözlenen,i}$ değerlerinin bilinmesi durumunda y_{ij}'nin koşullu ortalaması ile yer değiştirir.

EM algoritmasının M adımında parametrelerin yeni tahminleri $\theta^{(t+1)}$ tahmin edilir. (Little ve Rubin, 2002). Bunlar,

$$\mu_j^{(t+1)} = n^{-1} \sum_{i=1}^{n} y_{ij}^{(t)}, \quad j = 1,..., K; \tag{5}$$

$$\sigma_{jk}^{(t+1)} = n^{-1} E(\sum y_{ij} y_{ik} \mid Y_{gözlenen}) - \mu_j^{(t+1)} \mu_k^{(t+1)}$$

$$= n^{-1} \sum_{i=1}^{n} [(y_{ij}^{(t)} - \mu_j^{(t+1)})(y_{ik}^{(t)} - \mu_k^{(t+1)}) + c_{jki}^{(t)}]; \quad j,k = 1,...,K$$

ile elde edilir.

2.2 Çoklu Atıf

Her kayıp değer yerine tek bir değer atanması yerine, çoklu atıfta her kayıp değere birkaç kez atama yapılarak, birkaç tam veri seti elde edilir. Analizler ayrı ayrı yapılırak, sonuçlar birleştirilir. Rubin'in notasyonuna göre, gözlenen değerler $Y_{gözlenen}$, kayıp değerler $Y_{kayıp}$ ile gösterilir. Kitle niceliği Q'nun sonsal yoğunluğu aşağıdaki gibi yazılır.

$$h(Q \mid Y_{gözlenen}) = \int g(Q \mid Y_{gözlenen}, Y_{kayıp}) F(Y_{kayıp} \mid Y_{gözlenen}) dY_{kayıp} \tag{6}$$

Burada $f(.)$ kayıp değerlerin sonsal yoğunluk fonksiyonu ve $g(.)$ θ'nın tam veri seti için sonsal yoğunluk fonksiyonudur. Çoklu atıflar, kayıp verinin sonsal dağılımlarından benzetim ile seçilir.

Tam veri istatistikleri olan $\hat{Q}$ ile U, s tam veriden elde edilen $\hat{Q}_{*1},...,\hat{Q}_{*s}$ ile $U_{*1},...,U_{*s}$ hesaplanır. Tekrarlı atıf tahminleri

$$\overline{Q}_s = \sum_{l=1}^{s} Q_{*l}/s \tag{7}$$

ve $\overline{Q}_s$ 'in varyans-kovaryansı

$$T_s = \overline{U}_s + \frac{s+1}{s} B_s \tag{8}$$

olur. Burada

$$\overline{U}_s = \sum_{l=1}^{s} U_{*l}/s \tag{9}$$

atıf-içi değişkenlik ve

$$B_s = \sum_{l=1}^{s} (Q_{*l} - \overline{Q}_s)(Q_{*l} - \overline{Q}_s)'/(s-1) \tag{10}$$

atıflar-arası değişkenlik olarak isimlendirilir.

s'nin büyük değerleri için tekrarlı-atıf çıkarımlarından $(Q-\overline{Q}_s)$ normal dağılır, varyans-kovaryans matrisi T_s ile gösterilir. $s=\infty$ olduğunda, $(Q-\overline{Q}_s)$'nin dağılımı

$$(Q-\overline{Q}_\infty) \sim N(0, T_\infty) \tag{11}$$

olur. Burada $T_\infty = \overline{U}_\infty + B_\infty$ dur. (Atkinson ve Cheng, 2000).

Çoklu atıf ve EM algoritması karşılaştırılmalı olarak özetlenirse, EM algoritmasında E adımında kayıp değerler atanır ve tamamlanmış veriyle M adımda parametreler tahmin edilir. Böylece, kayıp değerler elde edilerek parametrelerin en çok olabilirlik tahminleri bulunur. En çok olabilirliğin kayıp veri yaklaşımında çok önemli bir yeri olmasına rağmen, doğrusal ve logaritmik doğrusal modellerde bazı kısıtları vardır. Çoklu atıf ise en çok olabilirlik ile aynı özelliklere sahiptir. Aynı zamanda her tür veriye ve her tür modele uygulanabilir. Tek dezavantajı çoklu atıfın her kullanımında farklı tahminler elde edilebilir olmasıdır (McLachlan ve Krishnan, (1997)). Farklı araştırmacılar aynı veri için aynı yöntemle farklı değerler elde edebilir, istenen aradaki farkın önemsenmeyecek kadar az olmasıdır (Allison, 2002).

3. BENZETİM ÇALIŞMASI

Atkinson ve Cheng (2000) çalışmasında EM algoritması ve çoklu atıf yöntemlerini karşılaştırmak üzere bir benzetim çalışması uygulamıştır. Bu çalışmada 4 boyutlu çok değişkenli normal dağılan bir veri setinden 100 ve 200 birimlik örneklemler çekerek %10, %20, %30 ve %40 oranında rastgele kayıp gözlemler yaratmışlar ve yöntemleri karşılaştırmışlardır. Bu çalışmamızda, Atkinson ve Cheng'in (2000) çalışmasına ek olarak regresyon analizi varsayımından sapma olması durumunda yöntemlerin nasıl sonuçlar vereceği vurgulanmak istenmiştir. Çalışmada, ilk olarak Minitab paket programında simetrik ve çarpık olmak üzere iki ayrı kitle yaratıldı. Simetrik kitle için, üç bağımsız değişkenli $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$ regresyon modeli kuruldu. Bağımsız değişkenlerden oluşan X matrisi çoklu normal dağılımdan $MN(0, I_4)$ türetildi. Regresyon modelinin tüm parametrelerine 1 değeri verildi. Hata terimi 0 ortalamalı 1 varyanslı normal dağıldı. ($\varepsilon_i \sim N(0,1)$). İkinci uygulama için (çarpık kitle için) yine üç bağımsız değişkenli bir regresyon modeli kuruldu. Çarpık bir kitle yaratabilmek için X matrisi 2 serbestlik dereceli χ^2 dağılımdan türetildi. Yine regresyon modelinin tüm parametrelerine 1 değeri verildi. Bu kitlelerden 100 birimlik genişlikte örneklemler çekildi. Bu örneklemlerde her bir bağımsız değişken üzerinde eşit oranda kayıp sağlanması amacıyla tüm X matrisi üzerinde %12, %24 ve %36'lık kayıp değer oluşturuldu. Kayıp değerli örneklemlere EM algoritması ve çoklu atıf yöntemleri uygulandı. Çoklu atıflar arasında da karşılaştırma yapabilmek amacıyla 1'den 10'a kadar olan atıf sayıları arasından 2, 5 ve 10 atıf sayıları seçildi. Regresyon katsayıları ve modelin standart hatası ($\sqrt{Hata\ Kareler\ Ortalaması}$) karşılaştırma kriteri olarak kullanılmak üzere hesaplandı. Bu amaçla yapılan benzetim çalışmasının adımları aşağıdaki gibidir.

1. Simetrik bir kitle türetildi.
2. n=100 birimlik örneklem çekildi.
3. X matrisinde %12'si kayıp değer olarak atandı.
4. 2'li , 5'li ve 10'lu Çoklu Atıf ve EM algoritması uygulandı.
5. Regresyon katsayıları ve modelin standart hatası hesaplandı.
6. 2. adıma dönerek bu süreç 300 kez tekrarlandı.

Çalışma, 3. adımda %24 ve %36 kayıp değer olacak şekilde tekrar uygulandı. Aynı işlemlere 1. adımda türetilen kitle çarpık kitle olacak şekilde, tekrar devam edildi.

Minitab paket programında kitlenin türetilmesi, makro program aracılığı ile eksik veri türetilmesinde yararlanıldı. Bu örneklemler SOLAS paket programına aktarılarak çoklu atıflar, SPSS paket programına aktarılarak ise EM algoritması uygulandı. Uygulanan yöntemler sonunda kayıp değerlerin yerine tahmin değerleri konularak Minitab paket programında makro program aracılığıyla regresyon çözümlemesi yapılarak, regresyon katsayıları ve modelin standart hatası kayıt edildi. Yapılan çalışmadan elde edilen sonuçlar aşağıdaki tablolarda sunulmuştur.

Tablo 1. Simetrik Veri için 300 Tekrardan Sonra Elde Edilen Parametre Tahminlerinin Ortalaması, Standart Sapma Değerleri ile Modelin Standart Sapması.

Kayıp veri oranı %	Yöntem	$E(\hat{\beta}_0)$ $(S_{\hat{\beta}_0})$	$E(\hat{\beta}_1)$ $(S_{\hat{\beta}_1})$	$E(\hat{\beta}_2)$ $(S_{\hat{\beta}_2})$	$E(\hat{\beta}_3)$ $(S_{\hat{\beta}_3})$	Standart Hata
12	ÇA(2)	1.00234 (0.11360)	0.89951 (0.13070)	0.89484 (0.13485)	0.89360 (0.11900)	0.2679
	ÇA(5)	1.00614 (0.12378)	0.90321 (0.14728)	0.88632 (0.14077)	0.88617 (0.14478)	0.2636
	ÇA(10)	1.00861 (0.11324)	0.88136 (0.14065)	0.88074 (0.14200)	0.89486 (0.13295)	0.3066
	EM	0.99596 (0.09739)	0.97504 (0.11040)	0.97468 (0.10519)	0.97649 (0.10353)	0.2036
24	ÇA(2)	1.01883 (0.13665)	0.77319 (0.16179)	0.75485 (0.15405)	0.75536 (0.17101)	0.3489
	ÇA(5)	1.01902 (0.13837)	0.77400 (0.16151)	0.74761 (0.16345)	0.78096 (0.15766)	0.3433
	ÇA(10)	1.01577 (0.14104)	0.75798 (0.15866)	0.74450 (0.17125)	0.76661 (0.16511)	0.3591
	EM	1.01670 (0.12940)	0.76839 (0.25075)	0.75050 (0.26198)	0.76876 (0.25945)	0.6803
36	ÇA(2)	1.02200 (0.14122)	0.65712 (0.17048)	0.64930 (0.16580)	0.66161 (0.17325)	0.3673
	ÇA(5)	1.02815 (0.14659)	0.65937 (0.16465)	0.64781 (0.16662)	0.65456 (0.18045)	0.4222
	ÇA(10)	1.03076 (0.14862)	0.65609 (0.17962)	0.64791 (0.18280)	0.64215 (0.17473)	0.4154
	EM	1.01221 (0.13198)	0.76841 (0.14844)	0.76253 (0.14292)	0.77451 (0.15264)	0.3391

Kitle parametrelerinin değeri 1 olduğundan, Tablo 1 incelendiğinde $\hat{\beta}_i$ katsayılarının ortalamaları, 1'e en yakın değerleri ve modelin en küçük standart hatasını %12 ve %36'lık kayıplarda, EM algoritmasından elde etmiştir. %24'lük kayıpta ise 5'li çoklu atıf modelin standart hatasını en küçük vermiştir. Atkinson ve Cheng (2000) çalışmasında $\hat{\beta}_i$ katsayılarının ortalamalarının 1'e en yakın değerlerini çoklu atıfta bulmuştur. 5'li ve 10'lu çoklu atıfların, 2'li çoklu atıfa göre daha iyi sonuçlar verdiğini gözlemlemiştir. $\hat{\beta}_0$ değerlerinin ortalaması için 1'den büyük değerler, $\hat{\beta}_1, \hat{\beta}_2$ ve $\hat{\beta}_3$'te 1'den düşük değerler elde edilmiştir. Çoklu atıflar kendi içinde değerlendirilirse aralarında çok önemli bir fark gözlenmemekle beraber, genelde 5'li çoklu atıf kitle değerine daha yakın sonuç vermiştir.

Tablo 2. Çarpık Veri için 300 Tekrardan Sonra Elde Edilen Parametre Tahminlerinin Ortalaması, Standart Sapma Değerleri ile Modelin Standart Sapması.

Kayıp veri oranı %	Yöntem	$E(\hat{\beta}_0)$ $(S_{\hat{\beta}_0})$	$E(\hat{\beta}_1)$ $(S_{\hat{\beta}_1})$	$E(\hat{\beta}_2)$ $(S_{\hat{\beta}_2})$	$E(\hat{\beta}_3)$ $(S_{\hat{\beta}_3})$	Standart Hata
12	ÇA(2)	0.98865 (0.11860)	0.92644 (0.14967)	0.88079 (0.14408)	0.85505 (0.14429)	0.3788
	ÇA(5)	0.98816 (0.11946)	0.92963 (0.15634)	0.87729 (0.14150)	0.84995 (0.14253)	0.3843
	ÇA(10)	0.98227 (0.11797)	0.92023 (0.15610)	0.88627 (0.13554)	0.85671 (0.14329)	0.3734
	EM	0.99248 (0.11086)	0.96604 (0.13602)	0.90449 (0.13278)	0.88561 (0.12704)	0.3499
24	ÇA(2)	0.99684 (0.14136)	0.79486 (0.16362)	0.76363 (0.15342)	0.75197 (0.16239)	0.4417
	ÇA(5)	0.99279 (0.13227)	0.80806 (0.17032)	0.75760 (0.16906)	0.75450 (0.18137)	0.4479
	ÇA(10)	1.00350 (0.12876)	0.81109 (0.18312)	0.76535 (0.16704)	0.74960 (0.16559)	0.4391
	EM	0.995387 (0.14343)	0.77138 (0.34559)	0.73023 (0.32510)	0.729011 (0.30076)	0.8214
36	ÇA(2)	0.98886 (0.14097)	0.67761 (0.19223)	0.66218 (0.19182)	0.64490 (0.18511)	0.5248
	ÇA(5)	0.98766 (0.15672)	0.68323 (0.20389)	0.66941 (0.18420)	0.64564 (0.19014)	0.4848
	ÇA(10)	0.99866 (0.14883)	0.66894 (0.20154)	0.65555 (0.20469)	0.63507 (0.19116)	0.5589
	EM	0.98744 (0.14389)	0.80658 (0.15471)	0.78319 (0.14598)	0.76185 (0.15699)	0.4572

Kitle parametrelerinin değeri 1 olduğundan, Tablo 2 incelendiğinde $\hat{\beta}_i$ katsayılarının ortalamaları, 1'e en yakın değerleri değerleri ve modelin en küçük standart hatasını %12 ve %36'lık kayıplarda, EM algoritmasından elde etmiştir. %24'lük kayıpta ise 10'lu çoklu atıf modelin standart hatasını en küçük vermiştir. Çarpık kitlede tüm $\hat{\beta}_i$ değerlerinin ortalaması 1'den küçük değerler almıştır. Çoklu atıflar kendi içinde değerlendirilirse aralarında çok önemli bir fark gözlenmemekle beraber, genelde 10'lu çoklu atıf kitle değerine daha yakın sonuç vermiştir.

4. SONUÇ

Bir veri setinde bulunan kayıp veriyi yok saymak, örneklemin rastgeleliğini bozarak, yanlı tahminler elde edilmesine neden olabilmektedir. Bu nedenle, bu çalışmada, regresyon analizinde, bağımsız değişkenlerde ortaya çıkan rassal kayıplar EM algoritması ve çoklu atıf yöntemi ile giderilmeye çalışılmıştır. Bu amaçla yapılan benzetim çalışmasında Atkinson ve Cheng'in (2000) çalışması temel alınmış ve ek olarak, regresyon analizinde hataların normal dağılması gerektiği varsayımından sapma olması durumunda, yöntemlerin nasıl sonuçlar vereceği vurgulanmak istenmiştir.

Yöntemleri karşılaştırmak için, regresyon katsayılarının beklenen değerinin 1'e yakın, regresyon katsayılarının standart hatalarının ve modelin standart hatasının küçük çıkması kriterleri kullanıldığında simetrik kitle için %12 ve %36'lık kayıpta EM algoritmasının en iyi sonucu verdiği gözlenmiştir. %24'lük kayıpta ise 5'li çoklu atıf yöntemi modelin hatasını en küçük yapabilmiştir. Çarpık kitle için değerlendirme yapıldığında yine %12 ve %36'lık kayıpta EM algoritması en iyi sonucu verirken, %24'lük kayıpta ise 10'lu Çoklu Atıf modelin hatasını en küçük yapabilmektedir.

Sonuç olarak regresyon analizinde hataların normal dağılması gerektiği varsayımından sapma olması durumunda, EM algoritması bu durumdan etkilenmemektedir. Ancak çoklu atıf için atıf sayısının arttırılması gerektiği söylenebilir.

KAYNAKLAR

AFIFI, A.A. and ELASHOFF, R.M. (1966). *Missing Observations in Multivariate Statistics: Review of the Literature,* J. Am. Statist. Assoıc. 61, 595-604.

ALLISON, P.D. (2002). *Missing Data, Sage Publications,* USA

ATKINSON, A.C. and CHENG, T-C. (2000). *On Robust Linear Regression with Incomplete Data,* Computational Statistics & Data Analysis, 33, 361.

DEMPSTER, A.P., LAIRD, N.M. and RUBIN, D.B. (1977). *Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion),* J. Roy. Statist. Soc. B39, 1-38.

HARTLEY, H.O. and HOCKING, R.R. (1971). *The Analysis of Incomplete Data ,* Biometrics 14, 174-194.

LITTLE, R. J. A. (1997). *Biostatistical Analysis with Missing Data,* in Encyclopedia of Biostatistics (P. Armitage and T. Colton, eds.) London: Wiley.

LITTLE, R. J. A. and RUBIN, D. B. (1983a). *IncompleteData,*in Encylopedia of Biostatistics *(P. Armitage and T. Colton, eds.) London: Wiley.*

LITTLE, R. J. A. and RUBIN, D. B. (2002), 2nd Ed., *Statistical Analysis with Missing Data,* A John Wiley & Sons, Inc. USA.

LITTLE, R.J.A. and SCHENKER, N. (1994). *Missing Data in Handbook for Statistical Modeling in the Social and Behavioral Sciences (G. Arminger, C.C. Clogg, and M.E. Sobel, eds.)* pp.39-75. NewYork: Plenum.

MCLACHLAN, G.J. and KRISHNAN, T. *(1997. The EM Algorithm and Extensions. New York: Wiley.*

ORCHARD, T. and WOODBURY, M.A. *(1972). A Missing Information Principle: Theory and Applications,* Proceedings of the 6th Berkeley Symposium on Mathematics, Statistics, and Probability, Volume 1, 697-715.

RUBIN, D.B. *(1976). Inference and Missing Data. Biometrika 63, 581-592.*

SCHAFER, J.L. *(1997). Analysis of Incomplete Multivariate Data,* Chapman & Hall, USA.

THE PROBLEM OF MISSING DATA IN REGRESSION ANALYSIS

ABSTRACT

The subject of missing data analysis consists of a data matrix in which some of the values in the matrix are not observed. Missing data analysis is one of the most important topics in applied statistics. It destroys the randomness of the sample and causes serious bias in the parameter estimate. The regression analysis is one of the most important procedures used for estimation in multivariate statistical analysis. For this reason, in this study, missing data mechanism designed by missing at random (MAR) for independent variables in regression analysis in two different data sets; one that verifies, one that violates regression assumptions; is used. When missing data can be ignored, model based methods that EM algorithm and multiple imputation method are compared. EM algorithm is not affected by the violation of regression assumptions but for multiple imputation number of imputations needs to be increased.

Key Words: *EM Algorithm, Missing Data, Multiple Imputation, Regression Analysis.*

İstatistik Araştırma Dergisi Cilt 04 No 03 Sayfa 63-75 Aralık 2005

2005 YILI İÇİN ÇOKLU AZALIM TABLOSU ÖNGÖRÜMLEMESİ

Banu ÖZGÜREL*

ÖZET

Çoklu azalım tablosu, sosyal güvenlik sistemine dahil olan belli bir grubun, yaşlara ait birden çok nedenden dolayı azalım olasılıklarının yer aldığı bir tablodur. Bu azalımlar, sisteme giren bir kişinin sistemden ayrılma nedenleri olarak tanımlanabilir. Örneğin, işten ayrılma, ölüm, malulen emeklilik azalım olarak tanımlanabilir. Bu çalışmada, sistemden ayrılışlar çeşitli ölüm nedenleriyle gerçekleşmektedir ve 2005 yılına ilişkin bir çoklu azalım tablosu oluşturulmuştur. Ayrıca 2005 yılına ait belirlenen nedenlerdeki ölüm sayıları, TÜİK'ten alınan 1961 - 2000 yılları arasındaki, belirli nedenlere ve yaşlara bağlı ölüm sayıları verileri öngörümlenerek (forecasting) bulunmuştur. Öngörümleme için gerekli olan modeller ise zaman serisi konusunun önemli bir alt başlığı olan Box–Jenkins metodu kullanılarak elde edilmiştir. Box - Jenkins modelleri zamana bağlı bir serinin otokorelasyon yapısını açıklayan ve böylelikle geleceğe ilişkin tahminler yapmaya imkan veren modellerdir. Sonuç olarak, elde edilen modeller ve çoklu azalım tablosu yardımıyla, özel ve sosyal güvenlik sigortalarının özellikle sağlık sigortası boyutu için sonucu ölüm olabilecek 3 azalım nedeninin toplum üzerine etkisi önceden tahminlenmeye çalışılmıştır.

Anahtar Kelimeler: *Box–Jenkins Modelleri, Çoklu Azalım Tablosu, Öngörümleme, Zaman Serisi Analizi*

1. GİRİŞ

Ölüm olayları nüfusun hacmine ve bünyesine etki eden üç farklı faktörden birini oluşturduğu için, ölüm olaylarının demografik bakımdan büyük bir önemi vardır. Bir nüfus kitlesine giriş doğum ve içe doğru göç ile, çıkış ise ölüm ve dışa doğru göç ile mümkündür. Göç, nüfus kitleleri arasında bir geçiş olduğu için nüfus kitlesinin hacmine çok büyük etkisi olmaz. Tabii faktörler olarak doğumların nüfusun hacmini artırıcı etkisi ne kadar önemli ise, ölümlerin azaltıcı rolü de o kadar büyüktür ve incelemeye değerdir.

* Dokuz Eylül Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, 35160 Buca/İZMİR, TÜRKİYE banu.ozakcan@deu.edu.tr

Ayrıca ölümler, sosyal ve iktisadi gelişmenin, tıbbi ilerlemenin, medeniyetin ve refah seviyesinin de demografik bir göstergesidir. Çünkü toplumların medeniyet ve refah seviyesi yükseldikçe ölüm oranları gerilemektedir.

2. YÖNTEM

Sosyal sigortalarda, bir veya birkaç nüfus sayımı kayıtlarından ve ölüm istatistiklerinden, nüfusun tamamı için derlenmiş tabloların kullanılması sağlanır. Bu tablolar, kadınlar ve erkekler için ayrı ayrı düzenlenir ve analiz tek azalmaya bağlı olarak yapıldığında yani tek bir ölüm nedeni ile azalma oluyorsa tek azalımlı tablo veya hayat tablosu, birden fazla azalım nedeni ile gruptan ayrılma varsa (Örneğin, sistemden ayrılma, malulen emeklilik, ölüm vb. gibi) bu durumda çoklu azalım tablosu olarak adlandırılır.

İnsan hayatına ait riskler, sigortacılık sektörüyle paylaşılarak azaltılmaktadır. Özellikle hayat sigortalarında ve sağlık sigortalarında prim hesaplamaları, hayat tablolarına ve çoklu azalım tablolarına dayanmaktadır. Çoklu azalım tablosu belli bir grubun, yaşlara göre, birden çok nedenden dolayı azalım olasılıklarının yer aldığı bir tablodur. Bu çalışmada öngörümleme yapılarak ileri tarihli bir yıl için çoklu azalım tablosu oluşturulmaya çalışılmıştır.

Gelecekte ne olacağı varsayımlarına iyi bir temel, geçmişte ne olduğuna ilişkin bilgidir. Bu bilgileri elde edebilmek amacıyla zaman serisinin özelliklerini ortaya koymak ve aynı zamanda ileriyi öngörümlemek amacını taşıyan ARIMA (p,q,d) modellerinden yararlanılmıştır. Durağanlık varsayımları sağlandığında çok tutarlı sonuçlar veren Box – Jenkins ve ARIMA (p,q,d) modelleri bu öngörümleme için uygun yöntemler olacaktır. Çünkü TÜİK verileri genellikle logaritmik veya karesel dönüşümle durağanlığı neredeyse kesin olan verilerden oluşmaktadır. Uygulamada en güçlü olan bu yöntemin karmaşık dönüşümler kullanılarak durağanlaştırılan serilerle yapılan ARIMA modellerinin geçerliliği de tartışmaya açıktır. Ancak bu araştırma için durağanlık küçük ve basit dönüşümlerle sağlanabildiğinden en güçlü ve kararlı analiz olan Box – Jenkins modellemesi bu araştırma için uygun görülmüştür.

3. VERİLER

Bu çalışmada TÜİK'in 1961 – 2000 yıllarına ait, belirli nedenlere ve yaşlara bağlı ölüm sayıları verileri kullanılmıştır[1]. Bu veriler 50 ölüm nedeni ve 10 yaş grubunu içermektedir. Ancak ölüm nedenleri bazı yıllarda değişiklik gösterdiği için bu 50 ölüm nedeni 46 ortak neden altında toplanmıştır (Ek-1)[1]. Örneğin, 1960'lı yıllarda vitaminsizlik nedeniyle ölümler yaşanırken daha sonraki yıllarda ölüm yaşanmaması nedeniyle bu neden ortadan kalkmıştır. Bunun gibi durumlardan dolayı 50 ölüm nedeni 46 nedene indirilerek, 40 yıllık veriler tekrar düzenlenmiştir. 1961-2000 yılları için veriler; erkek, kadın ve toplam olmak üzere gruplandırılmıştır. (Ek – 2)[2]

[1] 2000 yılından sonraki veriler henüz düzenlenmediği için ulaşılamamıştır.

[2] 1961-2000 yıl arasındaki tüm yıllar için erkek ve kadın olmak üzere toplam 80 adet ölüm sayıları tablosu hazırlanmıştır. Ancak sayfa kısıtı nedeniyle bu çalışmada yer almamaktadır. 2000 yılının 46 ölüm nedenine ilişkin ölüm sayıları (erkekler) tablosu örnek olarak verilmiştir.

4. UYGULAMA

1961 – 2000 yılları arasında, 46 ölüm nedenine ilişkin ölüm sayıları verilerinden yararlanarak; öncelikle yılları göz ardı ederek her yaş grubu için her nedene ait ölüm sayıları 40 yıl için toplanarak, erkekler ve kadınlar için çoklu azalım tablosu hazırlanmıştır.[3] Bu tablo yardımıyla her ölüm nedeni için her yaş aralığına ilişkin o nedene ait ölüm olasılıkları elde edilmiştir.

Daha sonra ayrı ayrı, her nedene ilişkin veriler Zaman Serisi Diyagramında (TSPlot) çizdirilip, serinin 40 yıl için genel seyri incelenmiştir. Box – Jenkins modellerinin en önemli varsayımı olan durağanlık koşulunu sağlamak amacıyla, gerek görüldüğü taktirde, verileri trendden ve dalga genişliklerinin değişkenliğinden arındırmak için fark alma ve dönüşüm yöntemleri (Log, 2. dereceden kök vb. gibi) uygulanmıştır. Dönüştürülmüş veriler arasından dalga genişlikleri değişkenliği ve trendi en az olan veriler için TAC ve TPAC'ın kestiricisi olan SAC ve SPAC çizdirilerek durağanlığı en iyi sağlayan veri seti ile model kurma aşamasına geçilmiştir. ARIMA modelleri uygulanarak varsayımları en iyi sağlayan ve parametreleri en anlamlı çıkan model öngörümleme aşaması için kullanılmıştır. Eğer birden fazla durağan seri varsa ve hangi serinin çalışma için seçileceği konusunda kararsız kalındıysa her bir durağan seri için ayrı ayrı ARIMA modelleri kurularak, en uygun modelde karar kılınmıştır (Ek-3)[4]. Elde edilen tüm modeller Ek – 4'teki tabloda yer almaktadır.

Bu deneysel aşamalardan sonra, geçerli modellerin hepsiyle bir çoklu azalım tablosu oluşturmaktansa, kadın ve erkekler için, modelleme yapılabilen ve ortak olan ölüm nedenleri ile öngörümleme aşamasına geçilmiştir. Belirlenen ölüm nedenleri; serebro vasküler hastalıklar, motorlu taşıt kazaları ve kronik romatizmal kalp hastalığıdır. Belirlenen ARIMA modelleri ile, 2000 yılı zaman merkezi (time origin) alınarak 2001, 2001 yılı zaman merkezi alınarak 2002, 2002 yılı zaman merkezi alınarak 2003, 2003 yılı zaman merkezi alınarak 2004 ve 2004 yılı zaman merkezi alınarak 2005 yılları için ölüm sayıları öngörümlenmiştir.

İlk aşamada oluşturduğumuz 40 yıllık çoklu azalım tablolarından yararlanılarak, 2005 yılı için her bir yaş aralığındaki ölüm olasılıklarının ortalaması hesaplatılmıştır. Belirlediğimiz ölüm nedenleri için 2005 yılına ait tüm yaş aralıklarına ilişkin beklenen ölüm sayıları bulunmuştur. Böylelikle 2005 yılı için seçilen tüm nedenler ve her yaş aralığına ait çoklu azalım tablosu oluşturulmuştur. Bu tablo Ek.5a ve Ek.5b'de yer almaktadır.

---

[3] 1961-2000 yıl arasındaki tüm yıllar için erkek ve kadın olmak üzere toplam 80 adet çoklu azalım tablosu hazırlanmıştır. Ancak sayfa kısıtı nedeniyle bu çalışmada yer almamaktadır.

[4] 46 nedene ilişkin tüm incelemeler yapılmıştır. Ancak sayfa kısıtı nedeniyle örnek olması için kronik romatizmal kalp hastalığı için Box-Jenkins modelinin MINITAB çıktısı Ek-3't verilmiştir

5. SONUÇ

Gelecekte incelenecek bir grup kişinin ölüm olasılığının ne olacağı bilinmemesine karşın, gelecekteki ölüm olasılığına ilişkin varsayımlar yapılarak, özel ve sosyal sigortalara ilişkin sorunlar çözümlenebilir. Bunun için, ölüm olasılıklarına ayanan bir tablo yardımıyla sayısal sonuçlar elde edilebilir. Bu çalışmada, çoklu azalım tablosu ile özel ve sosyal sigortalara ilişkin sorunların çözümüne yardımcı olabilecek bir uygulama yapılmaya çalışılmıştır.

Bazı hastalıklar insanın yaşamı için büyük bir önem taşımaktadır. Özellikle ölümle sonuçlanabilecek hastalıkların toplum üzerindeki etkisinin önceden tahminlenmesi, o hastalık için alınacak önlemler ve yapılacak çalışmalar için temel oluşturur. Bu çalışmada, ölümle sonuçlanabilecek 3 azalım nedeninin toplum üzerine etkisi önceden tahminlenmeye çalışılmıştır. Örneğin 2000 yılında motorlu taşıt kazalarından dolayı 25-34 yaşları arası ölen erkeklerin sayısı 360, ölme olasılığı ,12644 iken; 2005 yılı için 25-34 yaş aralığında ölen erkeklerin sayısı 438 (yaklaşık), ölme olasılığı 0,18796 olarak tahminlenmiştir. Ancak bu tahminlerin tutarlılığını ve geçerliliğini henüz kontrol edemiyoruz. Çünkü TÜİK veri tabanında bu bilgiler yer almamaktadır. İleri bir araştırma olarak önerimiz, 2005 yılı verileri elde edildikten sonra bu yöntemin geçerliliği ve tutarlılığı incelenebilir.

Ek 1. 46 Ölüm Nedeni

1	Kolera	24	Kronik Romatizmal Kalp Hastalığı
2	Tifo	25	Hipertansiyon
3	Basili Dizanteri	26	Kalp Hastalıklarının Diğer Şekilleri
4	Solumun Sistemi Tüberkülozu	27	Serebro Vasküler Hastalık
5	Geç Etkileri Dahil Tüberküloz Çeşitleri	28	Grip
6	Veba	29	Pnömöni
7	Difteri	30	Bronşit,Anfizem,Astma
8	Boğmaca	31	Mide Ülseri
9	Streptokoklu Anjin ve Kızıl	32	Apandisit
10	Meningokok Enfeksiyonları	33	Bağırsak Tıkanması Fıtık ve Enterit
11	Akut Poliyomiyelit	34	Karaciğer Sirozu
12	Çiçek	35	Nefrit ve Nefroz
13	Kızamık	36	Prostat Hiperplazisi
14	Tifüs ve Diğer Rickettsia Hastalıkları	37	Gebelik Doğurma Komplikasyonları
15	Sıtma	38	Doğuştan Gelen Anomaliller
16	Frengi ve Sekelleri	39	Doğum Travmatizması, Güç Doğum
17	Bütün Diğer Enfeksiyon Parazit Hastalıları	40	Perinatal Mortalitenin Diğer Sebepleri
18	Lenf ve Kan Yapıcı Dokuların Urları Dahil Habis Urlar	41	Semptomlar ve İyi Tanımlanmayan Haller
19	Selim Urlar ve Tabiatı Bilinmeyen Urlar	42	Bütün Diğer Hastalıklar
20	Şeker Hastalığı	43	Motorlu Taşıt Kazaları
21	Anemiler	44	Bütün Diğer Kazalar
22	Menenjit	45	Kendini Öldürme ve Yaralama
23	Akut Romatizma	46	Bütün Diğer Dış Sebepler

Ek 2. 2000 Yılının 46 Ölüm Nedenine İlişkin Ölüm Sayıları (Erkekler)

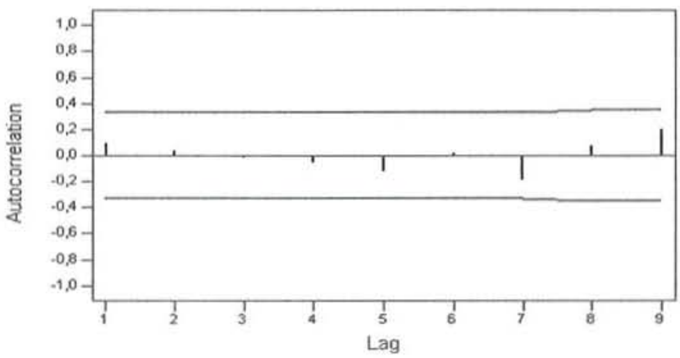
	Yaş										
Neden	**0**	**1-4**	**5-14**	**15-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**75+**	**?**	**Toplam**
1	0	0	0	0	0	0	0	0	0	0	0	**0**
2	3	0	0	0	0	0	0	0	0	1	0	**4**
3	1	0	0	0	0	0	0	0	2	2	0	**5**
4	4	5	4	9	35	81	117	103	115	47	0	**520**
5	2	2	7	0	5	4	8	3	7	5	0	**43**
6	0	0	0	0	0	0	0	0	0	0	0	**0**
7	2	0	0	0	0	1	0	0	0	0	0	**3**
8	1	0	0	0	0	1	0	0	0	0	0	**2**
9	0	0	0	0	0	0	0	2	0	0	0	**2**
10	1270	206	46	42	29	45	68	91	100	67	0	**1964**
11	2	3	2	1	0	1	0	0	3	0	0	**12**
12	0	0	0	0	0	0	0	0	0	0	0	**0**
13	3	1	1	2	0	0	0	0	1	0	0	**8**
14	0	0	0	0	0	0	1	0	0	0	0	**1**
15	0	0	0	0	0	0	0	0	0	0	0	**0**
16	0	0	0	0	0	0	1	0	0	2	0	**3**
17	39	36	37	14	18	34	44	66	67	18	0	**373**
18	38	63	140	242	301	953	2480	3958	5073	2290	0	**15538**
19	0	0	0	0	0	0	1	3	0	1	0	**5**
20	6	4	3	10	26	70	186	346	556	332	0	**1539**
21	9	9	4	13	4	4	10	6	17	10	0	**86**
22	50	29	25	14	17	12	12	10	13	6	0	**188**
23	0	0	0	0	1	0	0	0	0	0	0	**1**
24	0	2	1	3	9	19	51	60	98	46	0	**289**
25	0	6	0	6	10	26	90	248	481	298	0	**1165**
26	108	284	205	332	455	1663	4154	7066	11013	8794	0	**34074**
27	61	42	71	130	169	337	742	1384	2490	1706	0	**7132**
28	0	0	0	0	0	0	0	1	0	3	0	**4**
29	444	187	45	32	22	26	51	94	173	132	0	**1206**
30	28	12	6	6	5	11	20	60	143	81	0	**372**
31	0	0	19	56	80	105	133	192	227	190	0	**1002**
32	0	0	1	0	0	2	0	0	0	1	0	**4**
33	172	58	10	15	2	13	14	18	40	42	0	**384**
34	5	7	5	15	16	78	170	215	201	56	0	**768**
35	1	2	4	2	1	3	8	5	9	4	0	**39**
36	0	0	0	0	0	0	2	2	2	4	0	**10**
37	0	0	0	0	0	0	0	0	0	0	0	**0**
38	1822	122	0	0	0	0	0	0	0	0	0	**1944**
39	825	0	0	0	0	0	0	0	0	0	0	**825**
40	3402	0	0	0	0	0	0	0	0	0	0	**3402**
41	96	171	129	259	271	380	451	625	1298	3993	0	**7673**
42	240	201	153	231	279	573	1318	2182	3741	2922	0	**11840**
43	22	64	156	327	360	362	276	215	177	103	0	**2062**
44	108	106	149	458	413	402	290	247	189	132	0	**2494**
45	0	0	22	267	239	169	135	93	56	23	0	**1004**
46	1	0	6	91	80	93	49	37	19	2	0	**378**
Toplam	**8765**	**1622**	**1251**	**2577**	**2847**	**5468**	**10882**	**17332**	**26311**	**21313**	**0**	**98368**

Ek 3. Kronik Romatizmal Kalp Hastalığı İçin Box – Jenkins Modeli

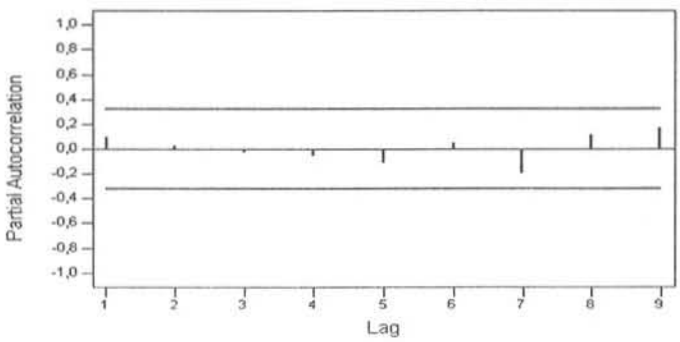
ARIMA Model : log+2fark for 24 – Kronik Romatizmal Kalp Hastalığı

```
Estimates at each iteration
Iteration        SSE     Parameters
     0        3,63914    0,100
     1        3,25946    0,250
     2        2,98154    0,400
     3        2,76930    0,550
     4        2,59381    0,700
     5        2,44394    0,850
     6        2,36428    0,927
     7        2,33073    0,956
     8        2,32638    0,966
     9        2,32634    0,963
    10        2,32629    0,963
Relative change in each estimate less than  0,0010
Final Estimates of Parameters
Type           Coef      SE Coef         T        P
MA   1       0,9628      0,0789      12,21    0,000
Number of observations:  38
Residuals:     SS =  2,32284  (backforecasts excluded)
               MS = 0,06278  DF = 37
Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag                 12        24        36       48
Chi-Square        10,1      23,4      28,2        *
DF                  11        23        35        *
P-Value          0,520     0,437     0,786        *
```

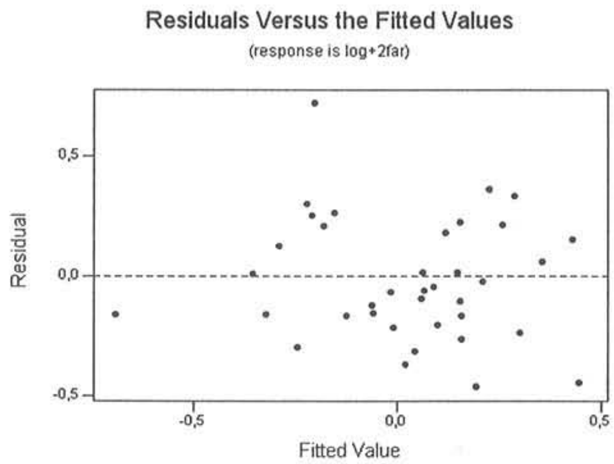
ACF of Residuals for log+2far
(with 95% confidence limits for the autocorrelations)

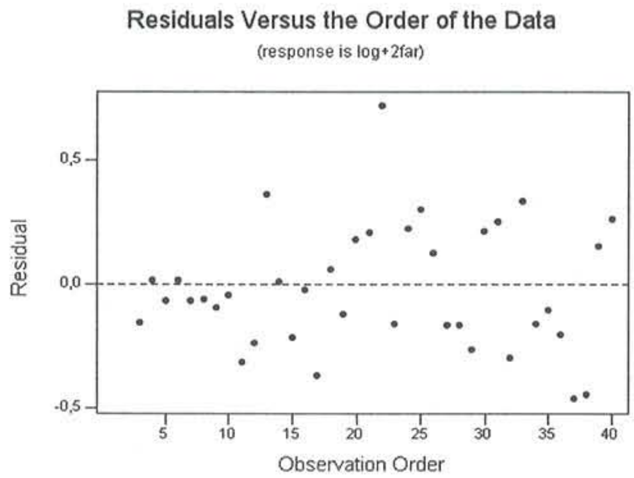


PACF of Residuals for log+2far
(with 95% confidence limits for the partial autocorrelations)



Ek 3. Kronik Romatizmal Kalp Hastalığı İçin Box – Jenkins Modeli (devam)





Ek 4. 46 Ölüm Nedeni için Elde Edilen Modeller ve Varsayımlar

Ölüm nedeni	Cinsiyet	Dönüşüm	Model	Sabit terim [1]	Model geçerliliği [2]	Varyans Homojenliği Varsayımı [3]	Otekorelasyonlu Olmaması Varsayımı [4]	Gözlemler arası bağımsızlık varsayımı [5]	Artıkların normal dağılması Varsayımı [6]
1	K	0	0	0	0	0	0	0	0
	E	0	0	0	0	0	0	0	0
2	K	kok 2 fark	ma1	0	1	1	1	1	1
	E	kok 1 fark	ma1	0	1	0	1	1	0
3	K	1. fark	ma1	0	1	1	1	1	1
	E	1. fark	ma1	0	1	1	1	1	1
4	K	log 2 fark	ar2	0	1	1	1	1	1
	E	2. fark	ma1	1	1	0	1	1	0
5	K	log 2 fark	ma1	0	1	1	0	1	1
	E	1. fark	ma1	0	1	0	0	1	0
6	K	0	0	0	0	0	0	0	0
	E	0	0	0	0	0	0	0	0
7	K	kok 2 fark	ma1	0	1	0	1	1	1
	E	kok 2 fark	ar2	0	1	1	1	1	1
8	K	kok 2 fark	ma1	0	0	1	1	1	1
	E	kok 1 fark	ar1	1	0	1	0	1	1
9	K	kok 2 fark	ar3	0	1	1	0	1	1
	E	1. fark	ar3	0	1	1	1	1	1
10	K	kok 2 fark	ma1	0	1	0	1	1	0
	E	0	0	0	0	0	0	0	0
11	K	4kok 1 fark	ma1	0	1	1	1	1	1
	E	4kok 1 fark	ar2	0	1	1	1	1	1
12	K	0	0	0	0	0	0	0	0
	E	0	0	0	0	0	0	0	0
13	K	log 2 fark	ar1	0	1	1	1	1	0
	E	log 1 fark	ar1	0	1	0	0	1	1
14	K	0	0	0	0	0	0	0	0
	E	0	0	0	0	0	0	0	0
15	K	0	0	0	0	0	0	0	0
	E	0	0	0	0	0	0	0	0
16	K	kok 2 fark	ma1	0	1	1	1	1	1
	E	kok 1 fark	ar1ma1	0	1	1	1	1	1
17	K	log 2 fark	ma1	0	1	1	0	1	1
	E	2. fark	ma1	0	1	0	0	0	0
18	K	2 fark	ma1	0	1	1	1	1	0
	E	1. fark	ma1	1	1	0	1	1	0
19	K	kok 1 fark	ar 3	0	1	1	1	1	0
	E	kok 1 fark	ma1	0	1	0	1	1	0
20	K	log 1 fark	ma1	1	1	1	1	1	1
	E	kok 1 fark	ma1	1	1	0	1	1	0

1) Sabit Terim (0: yok; 1: var)
2) Model Geçerliliği (0: sağlanamadı; 1: sağlandı)
3) Varyans Homojenliği Varsayımı (0: sağlanamadı; 1: sağlandı)
4) Artıkların Otokorelasyonlu Olmaması Varsayımı (0: sağlanamadı; 1: sağlandı)
5) Gözlemler Arası Bağımsızlık Varsayımı (0: sağlanamadı; 1: sağlandı)
6) Artıkların Normal Dağılması Varsayımı (0: sağlanamadı; 1: sağlandı)

Ek 4. 46 Ölüm Nedeni için Elde Edilen Modeller ve Varsayımlar (devam)

Ölüm nedeni	Cinsiyet	Dönüşüm	Model	Sabit terim [1]	Model geçerliliği [2]	Varyans Homojenliği Varsayımı [3]	Otekorelasyonlu Olmaması Varsayımı [4]	Gözlemler arası bağımsızlık varsayımı [5]	Artıkların normal dağılması Varsayımı [6]
21	K	4kok 1 fark	ma1	0	1	0	1	1	0
	E	2. fark	ma1	0	1	0	1	1	0
22	K	log 2 fark	ma1	0	1	1	0	1	1
	E	1. fark	ma1	0	1	0	0	1	0
23	K	kok 2 fark	ma1	0	0	1	0	1	1
	E	kok 1 fark	ar1ma1	0	1	1	0	1	1
24	K	log 2 fark	ma1	0	1	1	1	1	1
	E	4kok 2fark	ma1	0	1	1	1	1	1
25	K	log 2 fark	ar3	0	1	0	1	1	1
	E	2. fark	ma1	0	1	0	1	1	0
26	K	log 1 fark	ma1	1	1	0	1	1	0
	E	log 2 fark	ma1	0	1	0	0	0	0
27	K	log 2 fark	ma1	0	1	1	1	1	1
	E	log 2 fark	ar2	0	1	1	1	1	0
28	K	4kok 1 fark	ar1	0	1	1	1	1	0
	E	kok 1 fark	ma1	0	1	1	1	1	1
29	K	log 2 fark	ma1	0	1	1	1	1	1
	E	log 2 fark	ma1	0	1	1	0	1	1
30	K	kok 2 fark	ar4	0	1	1	1	1	1
	E	2. fark	ar3	0	1	1	1	1	1
31	K	log 2 fark	ma1	0	1	1	0	1	1
	E	2. fark	ar2	0	1	1	1	1	1
32	K	kok 1 fark	ma1	1	1	1	1	1	1
	E	kok 1 fark	ar3	0	1	1	1	1	1
33	K	log 1 fark	ar1	1	1	0	1	1	0
	E	1. fark	ar1ma1	1	1	1	1	1	1
34	K	kok 2 fark	ar4	0	1	1	1	1	1
	E	log 2 fark	ar2	0	1	0	0	0	0
35	K	kok 2 fark	ma1	0	1	1	0	1	1
	E	log 2 fark	ma1	0	1	0	1	1	0
36	K	0	0	0	0	0	0	0	0
	E	1. fark	ma1	1	1	1	1	1	1
37	K	kok 2 fark	ma1	0	1	1	1	1	1
	E	0	0	0	0	0	0	0	0
38	K	log 2 fark	ar3	0	1	0	0	1	1
	E	1. fark	ma1	0	1	0	1	1	0
39	K	kok 2 fark	ma1	0	1	0	1	1	1
	E	2. fark	ma1	0	1	0	1	1	0
40	K	kok 2 fark	ma1	0	1	0	0	1	0
	E	2. fark	ma1	0	1	0	1	1	0
41	K	log 2 fark	ma1	0	1	1	1	1	1
	E	log 2 fark	ma1	0	1	1	1	1	1
42	K	log 2 fark	ma1	0	1	0	1	1	0
	E	1. fark	ma1	1	1	0	1	1	0
43	K	log 2 fark	ma1	0	1	1	1	1	1
	E	log 2 fark	ma1	0	1	0	1	1	1
44	K	0	0	0	0	0	0	0	0
	E	0	0	0	0	0	0	0	0
45	K	log 2 fark	ma1	0	1	0	0	0	0
	E	log 2 fark	ma1	0	1	1	1	1	1
46	K	log 2 fark	ma1	0	1	1	0	1	1
	E	log 2 fark	ma1	0	1	0	1	1	0

Ek 5a. 2005 Yılı Seçilen Ölüm Nedenleri ve Her Yaş Aralığına ait Öngörümlenen Ölüm Sayıları

	ERKEK			KADIN		
	Kronik romatizmal kalp	Serebro vasküler	Motorlu taşıt kazaları	Kronik romatizmal kalp	Serebro vasküler	Motorlu taşıt kazaları
2005 için öngörümlenen ölüm sayısı	386,28	8402,06	2333,53	244,63	8263,55	937,625
Yaş / Ölüm Sayıları	$d_x(24)$*	$d_x(27)$	$d_x(43)$	$d_x(24)$	$d_x(27)$	$d_x(43)$
0	0,548518	103,7654	29,19246	0,787709	60,32392	24,82831
1-4	0,853679	97,37988	99,80508	1,103281	78,9169	76,48207
May.14	4,229766	185,2654	278,3668	5,961633	115,2765	152,5516
15 -24	5,91781	261,3041	385,2658	7,669151	115,6071	117,9157
25-34	11,04375	300,6257	438,6103	10,55823	156,8422	112,0837
35-44	32,8647	441,6123	395,58	15,88872	293,356	108,005
45-54	73,43569	938,8462	268,4726	28,3624	678,5201	99,3695
55-64	121,0061	1840,891	188,7126	57,06729	1464,632	98,64753
65-74	87,99845	2373,498	139,9651	61,61007	2446,093	85,21136
75+	47,32316	1797,369	78,42994	54,73107	2801,591	53,07895
Bilinmeyen	1,058407	61,41906	31,12929	0,8929	52,39091	9,45126
Toplam	386,28	8401,976	2333,53	244,6324	8263,55	937,625

* 24 nolu ölüm nedeni olan kronik romatizmal kalp hastalığı nedeni ile ölenlerin sayısı

Ek 5b. 2005 Yılı Seçilen Ölüm Nedenleri ve Her Yaş Aralığına ait Çoklu Azalım Tabloları Öngörümleri (Öngörümlenen Ölüm Olasılıkları)

	ERKEK			KADIN		
	Kronik romatizmal kalp	**Serebro vasküler**	**Motorlu taşıt kazaları**	**Kronik romatizmal kalp**	**Serebro vasküler**	**Motorlu taşıt kazaları**
2005 için öngörümlenen ölüm sayısı	386,28	8402,06	2333,53	244,63	8263,55	937,625
Yaş / Ölüm Olasılığı	$q_x(24)$**	$q_x(27)$	$q_x(43)$	$q_x(24)$**	$q_x(27)$	$q_x(43)$
0	0,00142	0,01235	0,01251	0,00322	0,0073	0,02648
1-4	0,00221	0,01159	0,04277	0,00451	0,00955	0,08157
May.14	0,01095	0,02205	0,11929	0,02437	0,01395	0,1627
15 -24	0,01532	0,0311	0,1651	0,03135	0,01399	0,12576
25-34	0,02859	0,03578	0,18796	0,04316	0,01898	0,11954
35-44	0,08508	0,05256	0,16952	0,06495	0,0355	0,11519
45-54	0,19011	0,11174	0,11505	0,11594	0,08211	0,10598
55-64	0,31326	0,2191	0,08087	0,23328	0,17724	0,10521
65-74	0,22781	0,28249	0,05998	0,25185	0,29601	0,09088
75+	0,12251	0,21392	0,03361	0,22373	0,33903	0,05661
Bilinmeyen	0,00274	0,00731	0,01334	0,00365	0,00634	0,01008
Toplam	1	0,99999	1	1,00001	1	1

** 24 nolu ölüm nedeni olan kronik romatizmal kalp hastalığı nedeni ile ölme olasılığı

KAYNAKLAR

BOWERMAN B.L. AND O'CONNELL R.T. (1993), *Forecasting and Time Series*, California Belmond, Duxbury Press.

BOWERS, GERBER, HICKMAN, JONES AND NESBITT (1986), *Actuarial Mathematics,* the Society of Actuaries.

BENJAMIN AND POLLARD (1992), *The Analysis of Mortality and Other Actuariel Statistics,* Butterworth – Heinemann Ltd.

WEI, W.S. (1990), *Time Series Analysis: Univariate and Multivariate Methods, Addison – Wesley,* Redwood City

FORECASTING THE MULTIPLE DECREMENT TABLE FOR 2005

ABSTRACT

A Multi – Decrement table is a probability table of more than one decrement causes versus ages for a given population is included in social security system. These decrements are described as leaving causes of the person included in the system. For example, withdrawal, death and disabled are described as leaving causes. In this research, leaving causes arise with various death causes and a Multi-decrement table is built using the deaths by selected causes and by age groups in 2005. The number of deaths in year 2005 is forecasted from TURKSTAT "Deaths by selected causes and by age groups for year 1961 - 2000" data. The models needed for forecasting is Box – Jenkins models which are the subtitle of Time Series Analysis. Box – Jenkins Models are explaining the structure of the autocorrelation of a time series that allows us to forecast. In conclusion, 3 decrement causes which can be in consequence of death try to estimate the influence of special and social security insurance which for especially health insurance on population.

Key Words: *Box–Jenkins Models, Forecasting, Multi Decrement Table, Time Series Analysis.*

İstatistik Araştırma Dergisi Cilt 04 No 03 Sayfa 76-91 Aralık 2005

EKK VE BAZI DAYANIKLI TAHMİNCİLERİN DERİNLİKLERİNİN KİRLENMEYE KARŞI DEĞİŞİMLERİNİN İNCELENMESİ

Enis SINIKSARAN* M. Hakan SATMAN* Y. Barış ALTAYLIGİL*

ÖZET

Bu çalışmada, veri kirliliği karşısında EKK ve bazı dayanıklı (robust) regresyon yöntemlerinin derinliklerinin davranışları incelendi. Yapılan Monte Carlo simülasyonları sonucunda bazı tahmincilere ilişkin örüntüler tespit edildi. Çalışmada ayrıca Bootstrap yöntemi ile elde edilen derinlik dağılımlarına dayanarak, EKK tahmincilerinin hipotez testlerinin derinliğe dayalı olarak yapılabileceği gösterildi.

Anahtar Kelimeler : *Aykırı Değerler, Bootstrap, Dayanıklı Yöntemler, En Derin Regresyon, P-Değeri Regresyon Derinliği.*

1. GİRİŞ

Derinlik, DR (Deepest regression) dışında herhangi bir dayanıklı yöntemin ve EKK'nın amaç fonksiyonunda yer alan bir kavram değildir. Bu anlamda regresyon derinliğinin, DR dışındaki tahmincilerin kalitelerine ilişkin bir ölçüt olup olamayacağı konusunda, literatürde net bir yanıt bulunmamaktadır. Bu çalışmada temel olarak bu sorunun yanıtı arandı.

Çalışma ana hatlarıyla şöyledir: Bu bölümün ardından gelen bölümde regresyon derinliği kavramı ve DR tahmincisi tanıtıldı. 3. bölümde aykırı değerlere karşı dayanıklı olan DR, LMS (Least Median Squares), LTS (Least Trimmed Squares) ve Huber'in M tahmincileriyle birlikte EKK ve LAD (Least Absolute Deviation) tahmincilerinin çeşitli oranlardaki kirlenmeler karşısında derinliklerinin nasıl değiştiği Monte Carlo simülasyonlarına dayanarak incelendi. Kirlenmenin yönü ve derecesine göre değişik örüntüler saptandı. Çalışmanın 4. bölümünde regresyon derinliğinin bootstrap dağılımları elde edildi ve bu dağılımlara dayanarak EKK parametrelerine ilişkin hipotez testlerinin yapılabilme olanakları araştırıldı. Klasik hipotez testlerinin p-değerleri ile karşılaştırıldığında, oldukça yakın sonuçlar elde edildiği görüldü.

* İstanbul Üniversitesi İktisat Fakültesi, Ekonometri Bölümü, İstanbul, TÜRKİYE
esiniksaran@istanbul.edu.tr

2. REGRESYON DERİNLİĞİ VE EN DERİN REGRESYON

Derinlik kavramının, ilk kez "yarı-uzay derinliği" olarak Tukey tarafından ortaya atılmasının ardından (Tukey, 1975) sorgulayıcı istatistikte konum, basıklık ve çarpıklık ölçüsünden (Liu, Parelius, Singh, 1990), çanta grafiği (bag plot) ile aykırı değer teşhisçisine (Rousseeuw, Ruts ve Tukey, 1999) ve kalite kontrol grafiklerinden (Liu ve Singh, 1993) asimptotik p-değerlerinin hesabına (Liu ve Singh, 1997) kadar pek çok alanda uygulanmıştır. (Karabulut ve Öztürk, 2003) ise yarı uzay derinliği kavramını dengeli bootstrap güven aralıklarının oluşturulmasında kullanmışlardır.

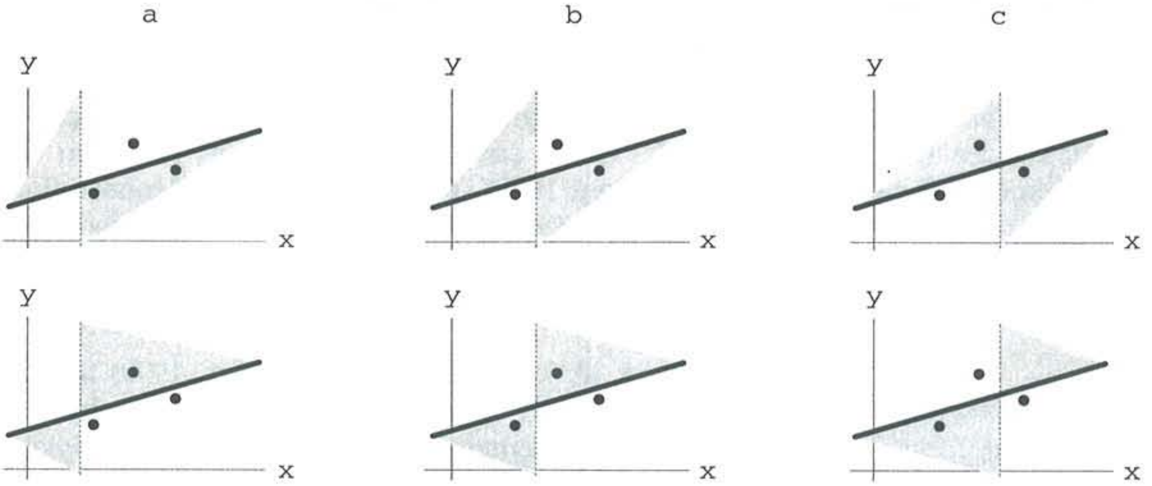
Tukey'in tanımına dayanarak ilk kez (Rouseeuw ve Hubert, 1999) tarafından sunulan regresyon derinliği kavramı (rdepth) ise Daniels'ın (Daniels, 1954) çalışması ile yakından ilişkilidir. Regresyon derinliği kavramı bir tahmincinin (θ) bir veriyle (Z_n) olan ilişkisini temsil eder. Burada θ, n birimlik bir örneklemden herhangi bir regresyon yöntemi ile elde edilmiş parametre tahmincilerinin oluşturduğu bir vektördür. Örneğin, $y = \beta_0 + \beta_1 x + \varepsilon$ basit doğrusal regresyon modelinin tahmin denklemi $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ ise, $\theta = (\hat{\beta}_0, \hat{\beta}_1)$ ve $Z_n = \{(x_i, y_i), i = 1, 2, \ldots, n\} \subset R^2$ olacaktır. Bu uzay için regresyon derinliğinin hesabında (Rousseeuw ve Hubert, 1999) bir algoritma önermişlerdir. Buna göre; verideki gözlemler x değerleri aynı olanlardan sadece biri alınarak $x_1 \leq x_2 \leq \ldots \leq x_n$ şeklinde dizildikten sonra n tane referans doğrusu x-eksenini $\{x_1-1, (x_1+x_2)/2, \ldots, (x_{n-1}+x_n)/2\}$ değerlerinden dik olarak kesecek şekilde belirlenir. Her bir v referans doğrusunun solunda kalan pozitif kalıntılar L^+, sağında kalan negatif kalıntılar R^- olarak gösterilmek üzere

$$L^+(v) = \#\{j; x_j \leq v \ ve \ r_j \geq 0\} \quad ve \quad R^-(v) = \#\{j; x_j > v \ ve \ r_j \leq 0\} \tag{1}$$

değerleri hesaplanır. Benzer şekilde L^- ve R^+ değerleri de hesaplandıktan sonra bir doğrunun regresyon derinliği:

$$rdepth(\theta, Z_n) = \min_{1 \leq i \leq n}(\min\{L^+(x_i) + R^-(x_i), L^-(x_i) + R^+(x_i)\}) \tag{2}$$

eşitliğini sağlayacak şekilde bulunur.



Şekil 1. Regresyon Derinliğinin Araştırılması

(2) eşitliği başka bir biçimde açıklanırsa; bir doğrunun regresyon derinliği, doğruyu uyumsuz (non-fit) yapabilmek için veriden atılması gereken minimum gözlem sayısıdır. Burada uyumsuz bir doğrunun geometrik yeri $x = c$ doğrusudur. Geometrik olarak açıklanırsa, bir doğru gözlemlerle çakışmayan her bir kaldıraç noktasına göre ve uyumsuz bir doğruya (x eksenine dik bir doğruya) dönüşecek şekilde çevrildiğinde üzerinden geçtiği nokta sayılarından minimum olanı bu doğrunun derinliğini verecektir denilebilir. Şekil 1'de 3 gözlemden oluşan bir verinin serpilme diyagramı ve bu veriye ilişkin bir regresyon doğrusu görmekteyiz. Kesikli olarak gösterilen v referans doğrularının, regresyon doğrusu ile kesim noktaları kaldıraç noktası olarak düşünülürse, regresyon doğrusunun her bir kaldıraç noktasına göre saat yönünde (üsttekiler) ve saatin tersi yönünde (alttakiler) döndürüldüğü durumlar a, b ve c olarak 3 bölgede gösterilmiştir. Regresyon doğrusu dönerken taradığı alanlar da Şekil 1'de görülmektedir. Eşitlik (2) uygulanırsa;

a: $(L^+ + R^-) = 0 + 2 = 2$, $(L^- + R^+) = 0 + 1 = 1 \Rightarrow \min(2, 1) = 1$

b : $(L^+ + R^-) = 0 + 1 = 1$, $(L^- + R^+) = 1 + 1 = 2 \Rightarrow \min(1, 2) = 1$

c : $(L^+ + R^-) = 1 + 1 = 2$, $(L^- + R^+) = 1 + 0 = 1 \Rightarrow \min(2, 1) = 1$

$rdepth(\theta, Z_n) = \min(1, 1, 1) = 1$ yazılabilir.

Bir verideki gözlemlerin tamamı regresyon doğrusunun üzerinde yer alırsa, regresyon doğrusu döndürüldüğünde bu noktaların üzerinden geçmiş olarak düşünüleceği için regresyon derinliği "n" olacaktır. Dolayısıyla bir doğrunun regresyon derinliği için $0 \leq \text{rdepth}(\boldsymbol{\theta}, Z_n) \leq n$ eşitsizliği yazılabilir.

Bağımsız değişken sayısı arttırıldığı zaman ise regresyon derinliği tanımlanırken doğruların yerini düzlemler (iki bağımsız değişkende) ve hiperdüzlemler (ikiden fazla bağımsız değişkende) alacaktır. Bu çalışmada regresyon derinliğini hesaplamada 2 boyut için yazılan ve ekte sunulan Mathematica kodları, 3 ve 4 boyut içinse Rousseeuw vd.'nin yazdığı Fortran yazılımları olan Medsweep, Rdepth3 ve Rdepth4[†] kullanılmıştır. (Bakınız Ek.1)

Bir regresyon tahmincisi θ'nın derinliği, uyum kalitesi olarak düşünülürse, maksimum derinliği verecek olan tahminciyi aramak da doğal olacaktır. Amaç fonksiyonu derinliğe dayanan ve en derin regresyon (DR) olarak isimlendirilen bu regresyon Rousseeuw ve Hubert (1999) tarafından aşağıdaki gibi tanımlanmış ve hesaplanması için algoritmalar sunulmuştur:

$$\mathrm{DR}(Z_n) = \arg\max_{\theta} \mathrm{rdepth}(\theta, Z_n)$$

Burada θ, maksimum derinliğe sahip tahmincidir. θ, birden fazla olduğu zaman ise bu tahmincilerin ortalamasının alınması önerilmektedir. Maksimum derinliğe ilişkin alt ve üst sınır ise $Z_n \subset R^p$ için

$$\left\lceil \frac{n}{p+1} \right\rceil \leq \max_{\theta} \mathrm{rdepth}(\theta, Z_n) \leq \left[\frac{n+p}{2} \right]$$

eşitsizliği ile belirlenir.

Amaç fonksiyonunda derinliğin yer almadığı EKK, LTS, LAD tahmincilerinin derinliklerine ilişkin alt sınırların ise birincisi için 1, diğerleri için p olduğu yukarıda söz edilen çalışmada ispatlanmıştır. Ancak literatürde bu tahmincilerin ve LMS, M gibi diğer dayanıklı tahmincilerin regresyon derinliklerine, özellikle kirlenme altındaki davranışlarına ilişkin ampirik bulgular henüz yer almamıştır. Bundan sonraki bölümde bu konuya ilişkin Monte Carlo simülasyonlarından elde edilen sonuçlar ve yorumları sunulacaktır.

3. DERİNLİK HESAPLAMALARINA İLİŞKİN MONTE CARLO SİMÜLASYONLARI

Çalışma tek, iki ve üç bağımsız değişkenli lineer regresyon modelleri için düşünüldü:

Model 1 : $y = \beta_0 + \beta_1 x + \varepsilon$ (3)

Model 2 : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ (4)

[†] Söz konusu programlar http://www.agoras.ua.ac.be adresinden indirilebilir.

Model 3 : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$ (5)

Simülasyonlarda temiz veri şöyle üretildi: X matrisinin sütunları; ortalamaları 7, varyansları 16 ve kovaryansları 0 olan çok değişkenli normal dağılımdan; hata vektörü ε ise standart normal dağılımdan elde edildi. Parametre vektörü $\beta = [\beta_0, \beta_1, ..., \beta_k]' = [5, 5, ..., 5]'$ şeklinde seçilirken y değerleri de $y = X\beta + \varepsilon$ denkleminden elde edildi. Bu şekilde üretilen regresyon verisi klasik varsayımları karşılamaktadır.

x yönünde veri kirletmek için, kirlenme oranı kadar seçilen rastlantısal gözlemlere şu dönüşüm uygulandı:

$$x_{ij}^{yeni} = \mu_j^x + \delta\sigma_j^x + Uniform(0,2) \quad (6)$$

Eşitlik (6)'da x_{ij}, X matrisinin i. satır ve j. sütun elemanını; δ, x yönünde kaç standart sapma sapılacağını, μ_j^x kirletilecek olan gözleme ait j. değişkenin ortalamasını, σ_j^x ise standart sapmasını göstermektedir. 0 ve 2 parametreli Uniform dağılımdan çekilecek rastlantısal sayıları simgeleyen *Uniform (0, 2)* terimi ise simülasyonlarda başvurulan tekrarlı yeniden örnekleme süreçlerinde X matrisinin herhangi bir alt kümesinin determinantının sıfırdan farklı çıkması için eklenmiştir.

Benzer şekilde y yönünde (dikey) kirlenme için Eşitlik (7)'deki dönüşüm uygulandı:

$$y_i^{yeni} = \mathbf{x}_i\boldsymbol{\beta} + \delta\sigma_\varepsilon = \mathbf{x}_i\boldsymbol{\beta} + \delta \quad (7)$$

Burada y_i, $\mathbf{y}$ vektörünün i. elemanını; $\mathbf{x}_i$, X matrisinin i. satırını; δ, i. kalıntının kaç standart sapma sapacağını göstermektedir. Eşitlik (6) ve Eşitlik (7), aynı gözlemler için uygulandığında hem x hem de y yönünde kirlenmeler elde edildi.

Simülasyonlar 3 farklı şekilde uygulandı. Birincisinde, derinliğin, örneklem büyüklüğü ve parametre sayısı ile ilişkisini çıkarmak için n = 10 birimden, n = 100 birime kadar örneklemler m = 25 kez çekildi ve her 3 model için EKK tahminlerinin derinlik medyanları hesaplandı. Bu deney temiz veriye uygulandı ve elde edilen sonuçlar Tablo 1'de verildi.

İkinci deneyde Model 2 ve Model 3 için n = 100 birimlik örneklemler m = 25 kez çekildi. Her bir örneklem için EKK, LMS, LTS, LAD, Huber'in M ve DR tahmincileri ve karşılık gelen derinlikleri ve bu derinliklerin medyan ($B_{1/2}$) ve MAD ($\sum_{i=1}^{m}|x_i - B_{1/2}|/m$) değerleri hesaplandı. Bu süreç temiz veri (c = 0) ve c = {0.10, 0.20, 0.40} düzeylerinde kirletilmiş veri için tekrarlandı. Her bir kirlenme düzeyi x ve y yönünde ve hem x, hem y yönünde ortalamadan 3 ve 5 standart sapma olacak şekilde tekrarlandı. Böylece derinliğin x yönündeki aykırı değerler (kaldıraç değerler) ile y yönündeki aykırı değerlere (dikey aykırı değerler) ve her iki yöndeki aykırı değerlere (bir kısmı potansiyel iyi huylu kaldıraç olabilecek) karşı değişimini inceleme olanakları elde edildi. Bu simülasyonlardan elde edilen sonuçlar Tablo 2 ve Tablo 3'te verilmiştir.

Üçüncü deneyde Model 2 yukarıda betimlenen kısıtlarla c = 0.01' den c = 0.49' a kadar 0,01'lik kirlenme artışlarıyla uygulandı ve her bir tahminci için hesaplanan derinliklerin grafikleri elde edildi. Bu grafikler Şekil 2, Şekil 3 ve Şekil 4'te görülmektedir. Şekillerdeki her bir grafiğin başlığında yer alan iki rakamdan ilki x yönündeki, ikincisi y yönündeki sapma miktarını diğer bir deyişle Eşitlik (6) ve (7)'deki δ değerini göstermektedir.

Üç deneyden elde ettiğimiz sonuçlara dayanarak aşağıdaki çıkarsamalar yapılabilir:

1) Bekleneceği ve Tablo 1'den açıkça görüleceği gibi, örneklem büyüklüğü arttırıldıkça derinlik artmaktadır. Parametre sayısı ile derinlik arasında ise ters yönlü bir ilişki olduğu yine anlaşılmaktadır.

2) Temiz veride derinliği en yüksek regresyon bekleneceği gibi DR'dir. İkinci derin regresyon ise LAD çıkmıştır. Bu sonuç kanımızca sürpriz sayılmamalıdır. Bilindiği gibi DR ve LAD tek değişkenli veriler için verilen 2 farklı medyan tanımının genelleştirilmiş halleridir. DR, medyanın gözlemler sıralandığında ortaya gelen terim olma tanımına, LAD ise $\sum_{i=1}^{n}\left|x_i - B_{1/2}\right|$ toplamını minimum yapma tanımına dayanmaktadır. Tek değişkenli verilerde en derin konum parametresi olan medyan, çok değişkenli versiyonlarında bu özelliğini korumuştur. Derinlik sıralamasında EKK ve Huber'in M tahmincileri DR ve LAD'dan sonra gelirken, LMS ve LTS daha sonra yer almışlardır. LMS ve LTS ayrıca en büyük saçılım gösteren (MAD'ı en büyük) tahminciler olarak görülmektedir.

3) Tüm tahminciler kirlenmeden, yönüne ve derecesine bağlı olarak farklı şekillerde etkilenmişlerdir. DR ve LAD, y yönünde kirlenmelerden hemen hiç etkilenmezlerken, x yönünde bir yerel minimum yapacak şekilde etkilenmektedirler. LMS, LTS ve Huber 'in M regresyonu kirlenmeye karşı birbirlerine benzer davranışlar sergilemişlerdir. Her üçü de gerek x, gerekse y yönünde kirlenme altında derinliklerini monoton azalan bir karakterde kaybetmektedir, EKK ise gerek x, gerekse y yönünde derinliğini çok küçük yüzdelerde hızla kaybetmekte; yerel bir minimumdan sonra kirlenme yüzdesi arttırıldıkça derinliği de artmaktadır. Bu özellik EKK'deki kadar belirgin olmasa da x yönündeki kirlenmelerde LAD'da ve bir miktar da DR'de görülmektedir. Bu sonuçlar, derinliği her tahminci için bir uyum iyiliği ölçüsü olarak kullanmanın sakıncalı olabileceğine işaret etmektedir.

Tablo 1. Model 1, 2 ve 3 için Örneklem Büyüklüğü ve Derinliklerin Medyanları

n	p = 2	p = 3	p = 4
10	3	2	2
15	5	3	2
20	7	5	4
25	9	7	6
30	11	9	8
35	13	11	10
40	15	13	12
45	17	15	13
50	19	18	15
55	22	19	17
60	24	21	20
65	26	24	21
70	28	25	23
75	30	28	25
80	33	29	27
85	36	32	31
90	38	35	32
95	40	36	34
100	42	38	37

Tablo 2. Model 2 için Tahmincilerin Derinliklerinin Medyan ve MAD Değerleri

Kısaltmalar: Kirlenme Yüzdesi - X Yönünde Sapma - Y Yönünde Sapma

n=100	%10 - 3 - 0		%10 - 5 - 0		%10 - 0 - 3		%10 - 0 - 5		%10 - 3 - 3		%10 - 3 - 5		%10 - 5 - 3		%10 - 5 - 5	
m=25	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD
OLS	9	0,64	9	0,76	10	0,96	10	0,96	7	1,48	7	1,12	8	1,72	7	1,92
DR	40	1,48	40	0,96	43	1,60	43	1,64	43	1,44	42	1,84	42	1,72	43	1,44
LMS	31	2,52	32	2,68	34	2,88	36	2,92	37	2,52	34	2,84	34	3,08	35	3,08
LTS	32	2,00	31	2,92	34	2,96	35	2,64	34	3,20	35	3,00	34	2,56	34	2,60
LAD	31	2,64	10	2,84	39	1,64	39	1,40	35	2,44	37	2,52	30	4,36	31	5,16
HuberM	31	3,52	32	2,72	37	2,88	36	3,12	34	2,44	34	2,68	34	3,48	35	2,16

n=100	%20 -3 - 0		%20 - 5 - 0		%20 - 0 - 3		%20 - 0 - 5		%20 - 3 - 3		%20 - 3 - 5		%20 - 5 - 3		%20 - 5 - 5	
m=25	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD
OLS	16	1,08	14	1,04	17	1,48	18	1,36	12	0,92	12	1,20	12	2,32	11	2,48
DR	31	1,12	31	1,60	44	1,12	43	1,52	41	2,48	40	2,40	39	2,56	38	2,84
LMS	28	3,48	29	2,48	35	3,00	31	2,72	34	2,80	33	2,88	33	3,72	32	2,36
LTS	28	2,88	29	2,08	34	2,84	35	3,00	34	2,92	35	2,36	34	2,32	34	2,52
LAD	17	0,84	15	1,00	39	1,72	39	1,76	33	4,60	30	5,24	20	5,24	21	5,60
HuberM	29	2,24	26	1,92	32	3,28	35	2,68	33	2,56	33	2,96	35	3,80	33	3,96

n=100	%40 - 3 - 0		%40 - 5 - 0		%40 - 0 - 3		%40 - 0 - 5		%40 - 3 - 3		%40 - 3 - 5		%40 - 5 - 3		%40 - 5 - 5	
m=25	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD
OLS	23	1,72	23	1,48	36	1,28	33	1,.88	22	1,48	21	1,80	21	1,64	22	1,68
DR	37	2,00	35	2,28	44	1,20	43	1,72	36	2,80	37	1,88	36	2,04	36	2,52
LMS	20	1,96	20	1,80	28	3,48	27	3,32	28	2,24	31	3,32	30	2,92	30	2,40
LTS	22	1,52	22	1,40	27	2,16	28	2,60	30	1,92	31	2,36	32	2,08	31	2,64
LAD	26	1,64	22	1,48	39	1,64	39	1,16	27	1,12	27	1,84	25	1,28	25	0,96
HuberM	21	1,80	21	1,68	28	3,32	27	3,12	31	2,88	31	2,64	30	2,88	30	2,60

n=100	%0 - 0 - 0	
m=25	Med.	MAD
OLS	39	1,12
DR	43	1,60
LMS	35	2,56
LTS	35	1,92
LAD	40	1,40
HuberM	37	2,12

Tablo 3. Model 3 için Tahmincilerin Derinliklerinin Medyan ve MAD Değerleri

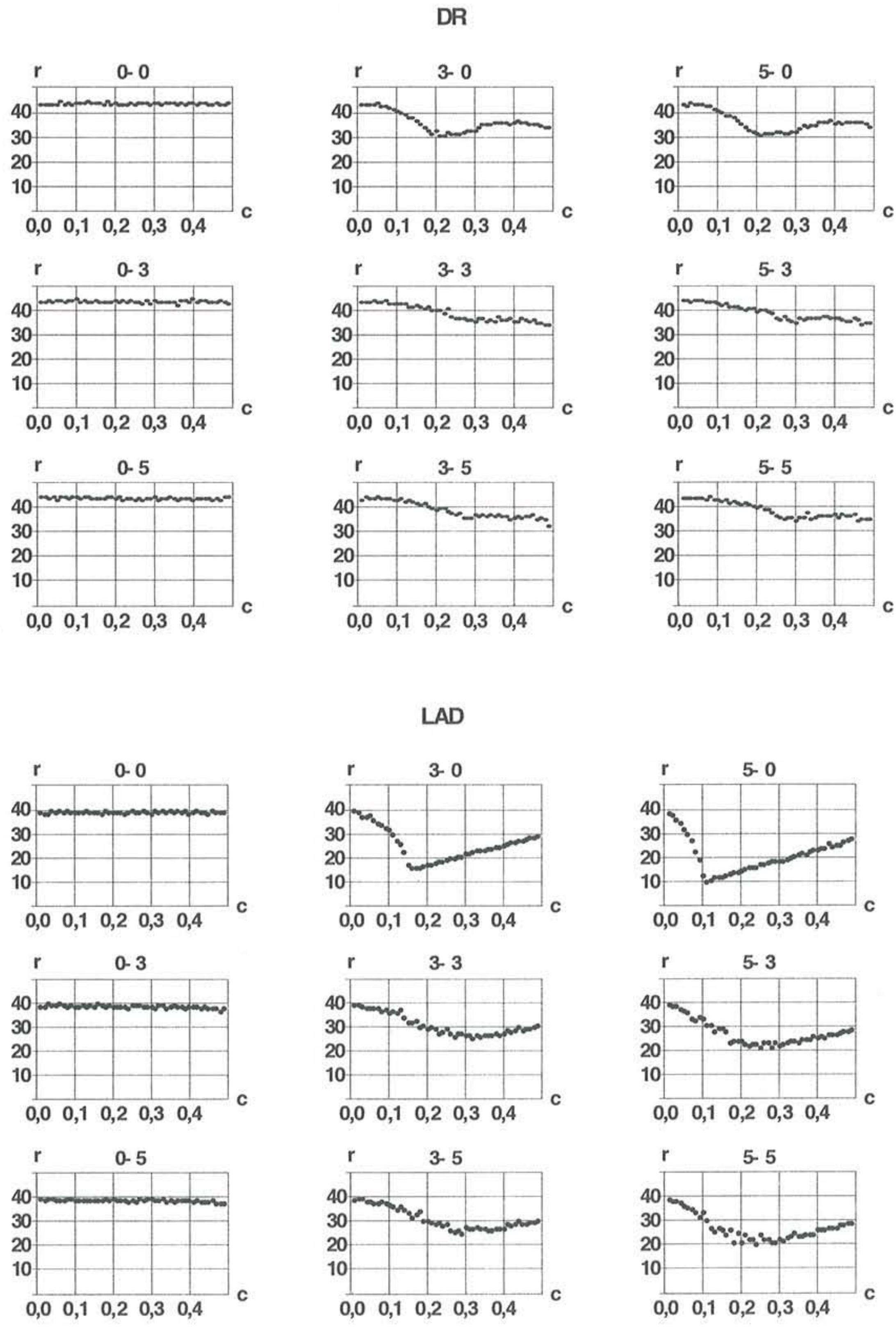
Kısaltmalar: Kirlenme Yüzdesi - X Yönünde Sapma - Y Yönünde Sapma

n=100	%10 - 3 - 0		%10 - 5 - 0		%10 - 0 - 3		%10 - 0 - 5		%10 - 3 - 3		%10 - 3 - 5		%10 - 5 - 3		%10 - 5 - 5	
m=25	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD
OLS	9	0,56	8	0,92	9	1,08	9	0.92	6	0,96	6	1,32	6	1,40	6	1,80
DR	40	1,12	39	1,32	43	1,36	43	1,28	42	1,84	42	1,72	42	2,08	42	1,28
LMS	29	2,44	32	2,36	32	3,48	32	2,60	31	2,56	29	2,88	31	1,92	30	2,28
LTS	29	2,04	27	2,76	32	2,32	31	2,96	32	2,40	30	2,76	31	2,12	31	2,56
LAD	24	2,84	9	1,04	35	1,60	35	1,68	32	2,12	32	2,76	26	7,56	22	6,80
HuberM	28	2,60	29	2,64	31	3,28	31	2,40	29	3,68	32	3,04	32	2,80	30	3,08

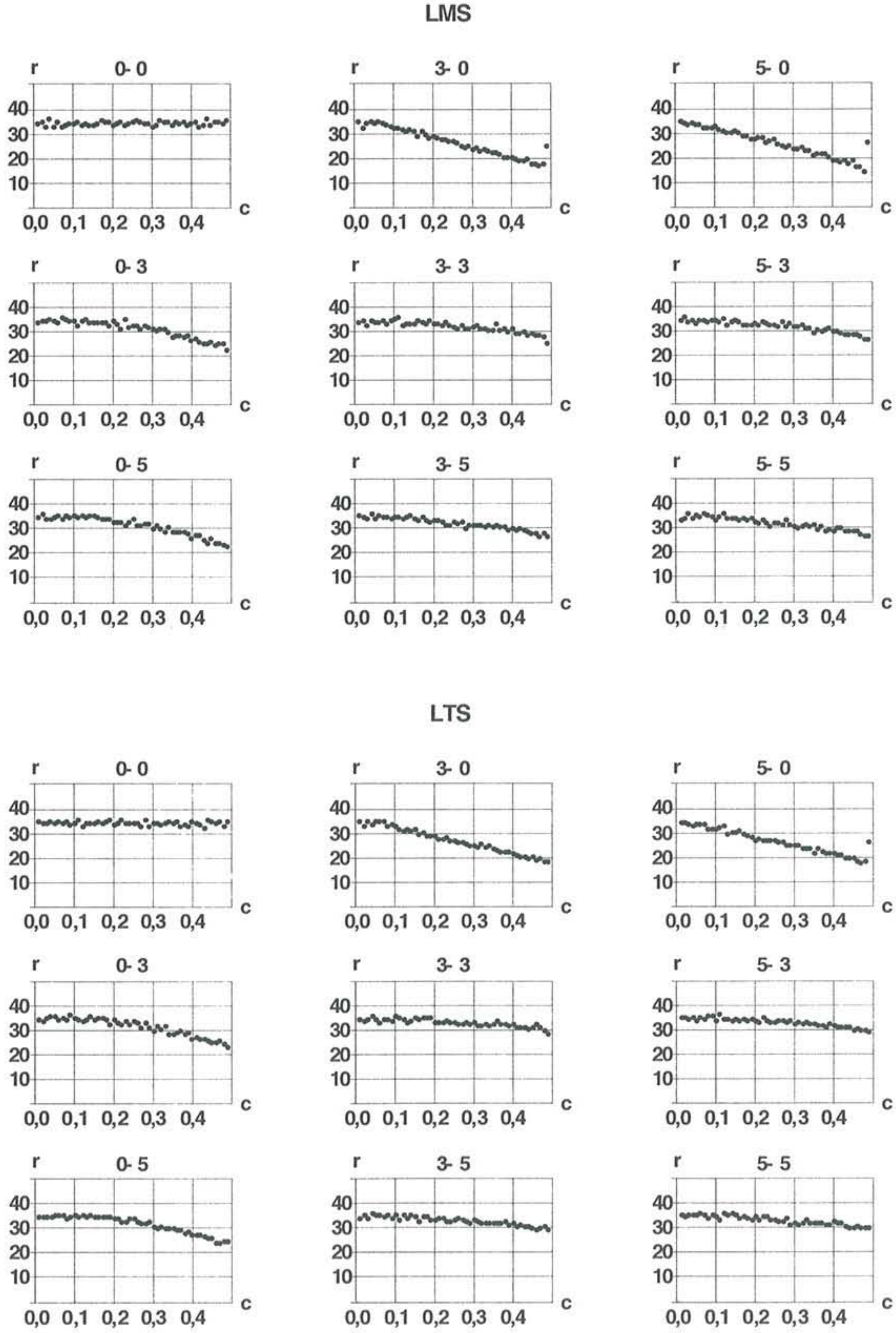
n=100	%20 - 3 - 0		%20 - 5 - 0		%20 - 0 - 3		%20 - 0 - 5		%20 - 3 - 3		%20 - 3 - 5		%20 - 5 - 3		%20 - 5 - 5	
m=25	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD
OLS	15	1,28	14	1,20	17	0,88	18	0,80	11	1,12	11	1,00	10	1,32	11	1,40
DR	34	1,64	34	1,52	43	1,44	43	0,96	39	2,64	38	2,92	38	2,52	36	2,48
LMS	25	2,76	26	1,64	31	2,84	31	3,44	31	2,64	28	2,12	30	3,40	28	2,76
LTS	25	2,40	27	1,56	32	2,12	31	2,16	31	2,32	30	1,88	31	2,60	30	3,00
LAD	16	1,28	13	1,56	35	1,36	35	1,40	26	4,76	20	4,60	17	3,28	16	2,32
HuberM	27	2,72	26	2,36	32	3,00	30	3,04	30	2,00	32	2,96	30	3,48	30	2,88

n=100	%40 - 3 - 0		%40 - 5 - 0		%40 - 0 - 3		%40 - 0 - 5		%40 - 3 - 3		%40 - 3 - 5		%40 - 5 - 3		%40 - 5 - 5	
m=25	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD	Med.	MAD
OLS	24	1,64	23	1,76	33	1,56	32	1,60	21	0,84	21	1,40	20	1,12	22	1,48
DR	37	2,00	37	2,20	43	2,00	43	1,48	38	2,04	37	1,88	38	2,04	38	2,08
LMS	18	2,08	18	1,56	24	2,72	25	2,96	27	2,76	28	3,36	25	3,08	26	1,92
LTS	19	1,68	18	2,00	25	2,20	26	2,28	27	2,64	28	2,76	27	2,40	26	2,04
LAD	24	2,32	23	2,56	35	1,52	34	1,80	25	1,52	25	1,60	23	1,32	24	1,48
HuberM	17	2,28	18	1,92	27	2,88	26	4,12	27	2,04	26	2,56	27	2,52	27	2,84

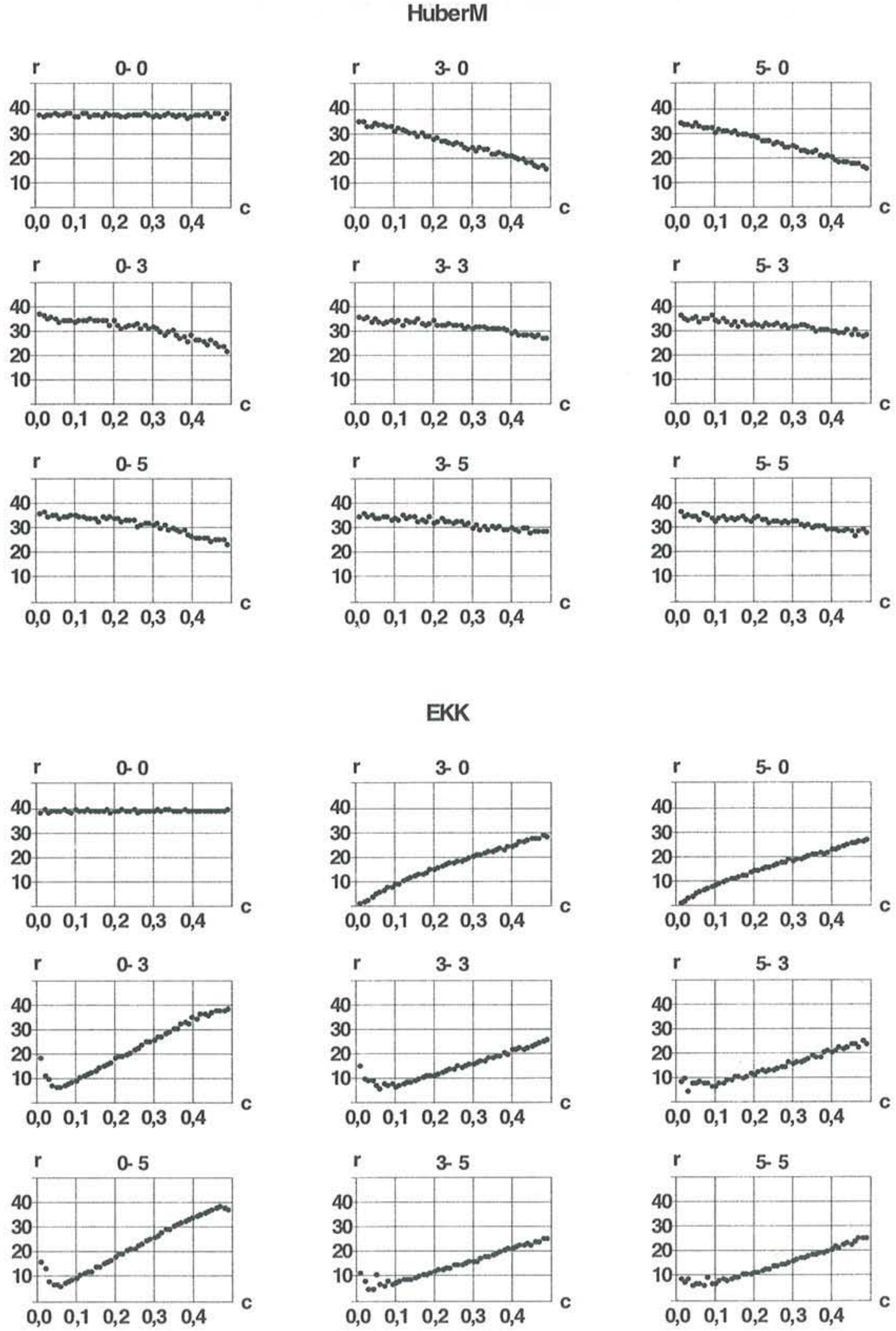
n=100	%0 - 0 - 0	
m=25	Med.	MAD
OLS	37	1,28
DR	44	1,48
LMS	32	3,36
LTS	32	2,96
LAD	35	1,44
HuberM	34	1,84



Şekil 2. DR ve LAD Tahmincilerinin Çeşitli Kirlenme Yüzdelerinde (c) Derinlikleri (r)



Şekil 3. LMS ve LTS Tahmincilerinin Çeşitli Kirlenme Yüzdelerinde (c) Derinlikleri (r)



Şekil 4. Huber'in M ve EKK Tahmincilerinin Çeşitli Kirlenme Yüzdelerinde (c) Derinlikleri (r)

4. DERİNLİK DAĞILIMLARI İLE HİPOTEZ TESTLERİ

Van Aelst vd., (2002) DR'nin parametrelerinin anlamlılık testlerini gerçekleştirmede maksimum derinliğin dağılımının kullanılabileceğini önermişlerdir. 2 boyut için (tek bağımsız değişkenli model) böyle bir dağılım Daniels (1954)'te elde edilmiştir. 2'den büyük boyut için ise dağılımların simülasyonlarla elde edilebileceği yine Aelst'in çalışmasında önerilmiştir.

Bir önceki bölümde tahmincilerin derinliklerinin kirlenmeden etkilendikleri, açık örüntüler sergiledikleri görüldü. Dolayısıyla Van Aelst vd. (2002)'de DR'ye uygulanan parametrelerin anlamlılık testlerine EKK uygulanabileceği düşüncesi akla gelebilir. Düşüncenin araştırılması amacıyla Eşitlik (4)'teki 2 bağımsız değişkenli model için n = 100 birimlik örneklemler çekilip, her bir örneklemden de B=1000 bootstrap örneklemi elde ederek, H_0: $\beta_i = 0$ hipotezi altındaki derinliklerin dağılımları bulundu. Dağılımların kantillerinden ilgilenilen hipotezlerin p-değerleri hesaplandı ve bunlar klasik yolla yani t dağılımından hesaplanan p-değerleriyle karşılaştırıldı. m = 25 kez uygulanan bu sürecin sonunda % 90' ın üzerinde tutarlılıklar elde edildi. Diğer bir deyişle klasik yöntem ve derinlik dağılımları ile uygulanan süreçle hipotez testi sürecinde aynı sonuca varma olasılığı 0,90'nın üzerindedir. Sonuçlar Tablo 4'te verilmiştir.

Uygulanan süreç, EKK'nın dağılımsal varsayımlarının ihlal edildiği ve kaldıraç etkisi yapabilecek aşırı değerlerin olduğu verilerde dayanıklı bir yöntem olarak önerilebilir.

Tablo 4. Klasik Süreçten ve Derinliklerin Bootstrap Dağılımlarından Bulunan p-Değerleri

$H_0:\beta_0=0$, Klasik	$H_0:\beta_0=0$, Derinlik	$H_0:\beta_1=0$, Klasik	$H_0:\beta_1=0$, Derinlik	$H_0:\beta_2=0$, Klasik	$H_0:\beta_2=0$, Derinlik
0.35607	0.25	0.41051	1.	0.0014605	0.
0.90508	0.99	0.28324	0.	1.1564×10^{-10}	0.
0.016943	0.	0.064383	0.	0.15362	0.96
0.47282	0.6	6.7121×10^{-6}	0.	0.012055	0.
0.10975	0.	0.094342	0.06	0.72028	0.92
0.0013437	0.	0.026793	0.	0.43660	0.08
0.0060544	0.	0.050335	0.	0.061832	0.01
0.46853	0.91	9.8772×10^{-10}	0.	0.000097474	0.
0.46265	0.95	0.23471	0.8	0.084624	0.57
0.0032205	0.	0.16828	0.1	0.86380	1.
0.0065012	0.	0.0090960	0.	0.041791	0.25
0.11605	0.05	0.79132	0.56	0.0013340	0.
0.18035	0.09	3.2195×10^{-9}	0.	0.0067276	0.
0.19853	0.14	0.00071839	0.	0.54661	0.42
0.22275	0.17	0.0066673	0.	0.52604	0.43
0.91249	0.88	0.00051203	0.05	0.18752	0.42
0.037204	0.	0.00031639	0.	0.12748	0.05
0.26202	0.5	0.17236	0.37	0.24331	0.68
0.75644	0.76	0.043975	0.04	0.12043	0.33
0.66189	0.27	0.91368	0.93	0.0014065	0.
0.77260	0.99	0.040525	0.01	0.043852	0.03
0.51636	0.4	0.027517	0.01	0.0018272	0.01
0.24201	0.01	0.0033037	0.	0.16190	0.37
0.0013644	0.	0.010118	0.	0.0078881	0.
0.29193	0.15	0.0056876	0.	0.99860	0.92

5. SONUÇ

Bu çalışmada çeşitli tahmincilerin derinliklerinin kirlenme karşındaki değişimleri incelendi ve derinliğin, tahmincilerin uyum iyiliklerinin (goodness of fit) bir ölçüsü olabilme olanakları araştırıldı. Çalışmadaki simülasyonlar, kirlenme karşısında, ilgilenilen 6 tahmincinin derinliklerinin belirgin örüntüler sergilediğini gösterdi. LMS, LTS ve Huber'in M tahmincileri kirlenme arttırıldıkça derinliklerini kaybederken, EKK'de sezgiye ters gelebilecek ters yönlü bir ilişki görüldü. EKK'deki kadar belirgin olmasa da yalnızca x yönündeki kirlenme karşısında LAD ve hafif ölçüde DR'de benzer bir özellik görüldü. Bu sonuçlar, her tahminci için derinliğin bir uyum ölçüsü olamayacağına ancak derinlik ve uyumun bazı durumlarda, birlikte incelenmeye değer olduğuna işaret edebilir.

Çalışmada ayrıca bootstrap yöntemiyle elde edilen derinlik dağılımları ile hipotez testi sürecinin uygulanma olanakları araştırılmış, klasik hipotez testi süreciyle karşılaştırıldığında anlamlı sonuçlar elde edilmiştir.

Ek 1 . 2 Boyutlu Veride Regresyon Derinlik Hesabının Mathematica Programı

```
RDepth2::usage =
  "RDepth2[data,parameters], nx2 boyutlu bir matris ile verilen veri seti ve 2x1
    boyutlu bir parametre vektörü için regresyon derinliği hesaplar";
Off[General::"spell1"];
RDepth2[data_, parameters_] :=
  Module[{ind, dep, n, distx, ones, xmatrix, e, vj, j, kaldirac, splus, sminus, gplus,
    gminus, rdepths, depth2},
   (* Design Control *)
   If [ MatrixQ[data] == False, {Print["Veri seti nx2 boyutlu bir matris olmalıdır."]; Abort[];}];
   If [Dimensions[data][[2]] ≠ 2 , {Print["Veri matrisi 2 sütundan oluşmalıdır."]; Abort[];}];
   If [Length[parameters] ≠ 2 || VectorQ[parameters] == False,
    {Print["Parametreler 2x1 boyutlu bir vektör olmalıdır"]; Abort[];}];
   If [Length[data] < Length[parameters],
    {Print["Parametre sayısı, gözlem sayısından fazla olamaz."]; Abort[];}];
   (* Variable Definition *)
   ind = Transpose[data][[1]];
   dep = Transpose[data][[2]];
   n = Length[ind];
   distx = Union[ind];
   ones = Table[1, {Length[ind]}];
   xmatrix = Transpose[Table[{ones, ind}]];
   (* Residuals and Leverages Calculation *)
   e = dep - xmatrix.parameters;
   vj = Table[0, {Length[distx]}];
   vj[[1]] = distx[[1]] - 1;
   For [j = 2, j ≤ Length[distx], j++,
    vj[[j]] = (distx[[j - 1]] + distx[[j]]) / 2;
   ];
   rdepths = Table[0, {0}];
   (* All Possible Depths for All Leverages *)
   For [j = 1, j ≤ Length[vj], j++,
    kaldirac = vj[[j]];
    splus = 0; sminus = 0; gplus = 0; gminus = 0;
    For [i = 1, i ≤ n, i++,
     If [ind[[i]] < kaldirac && e[[i]] ≥ 0 , {splus++;}];
     If [ind[[i]] < kaldirac && e[[i]] < 0 , {sminus++;}];
     If [ind[[i]] > kaldirac && e[[i]] ≤ 0 , {gminus++;}];
     If [ind[[i]] > kaldirac && e[[i]] > 0 , {gplus++;}];
    ];
    rdepths = Append[rdepths, Min[splus + gminus, sminus + gplus]];
   ];
   (* Minimum of all depths is Rdepth *)
   depth2 = Min[rdepths];
   Return [depth2];
  ];
```

KAYNAKLAR

DANIELS, H.E. (1954), *A Distribution-Free Test for Regression Parameters*, Annals of Mathematical Statistics, 25,499-513.

KARABULUT, İ., ÖZTÜRK, F. (2003), *Lineer Modellerde Yarı Uzay Derinliğine Dayalı Dengeli Bootstrap Güven Bölgeleri*, İstatistik Araştırma Dergisi, Cilt 2, No:3, 63-72.

LIU, R., PARELIUS, J. , SINGH, K. (1990), *Multivarite Analysis By Data Depth : Descriptive Statistics, Graphics and Inference*, the Annals of Statistics, Vol.27, No.3, 783-858.

LIU, R., SINGH, K. (1993), *A Quality Index Based on Data Depth and Multivariate Rank Tests*, Journal of the American Statistical Association, Vol.88, 257-260.

LIU, R., SINGH, K. (1997), *Notions of Limiting P-Values on Data Depth and Bootstrap*, Journal of the American Statistical Association, Vol.91, 266-277.

ROUSSEEUW, P. J., HUBERT, M. (1999), *Regression Depth*, Journal of the American Statistical Association, Vol.94, 388-402.

ROUSSEEUW, P.J., RUTS, I., TUKEY, J.W. (1999), *The Bagplot: A Bivarite Boxplot*, the American Statistican, Vol.53, No.4, 382-387.

TUKEY, J.W. (1975), *Mathematics and Picturing Data*, Proceedings of the 1974 International Congress of Mathematics, 523-531.

VAN AELST, S., ROUSSEEUW, P.J., HUBERT, M., STRUYF, A. (2002), *The Deepest Regression Method*, Journal of Multivariate Analysis, Vol.81, 138-166.

A STUDY FOR EXAMINING THE BEHAVIORS OF REGRESSION DEPTHS OF OLS AND SOME ROBUST REGRESSION METHODS UNDER CONTAMINATION

ABSTRACT

In this paper, we examine the behaviors of the regression depths of OLS and some robust regression methods under contamination. Upon Monte Carlo simulations, we determine some patterns for some estimators. Using bootstrap method, we also show that the distributions of regression depth can be utilized in hypothesis testing of regression parameters of OLS.

Key Words: *Bootstrap, Outliers, P-Value, Regression Depth, Robust Methods, The Deepest Regression.*

ENDEKS-INDEX						
MAKALE İSMİ	NAME OF ARTICLE	YAZAR ADI *NAME OF CONTRIBUTOR*	CİLT *VOLUME*	NO *NUMBER*	SAYFA *PAGE*	BASIM YILI *YEAR*
Bernoulli Sayıları ve Lojistik Dağılış Fonksiyonu Üzerine Bir Çalışma	*A Study on Bernoulli Numbers and Logistic Distributiın Function*	DÜNDAR Samim	01	01	1	2002
Deney Tasarımında Kesirli Çoketkenli Tasarımlar ve Uygulamaları	*Fractional Factor Designs and Their Applications in Experimental Design*	KADILAR Cem MULUK Zehra	01	01	11	2002
Üç Yönlü Tablolarda χ^2 İstatistiğinin Kullanılması	χ^2 Statistical Analysis Used of Three-Dimensional Table	AKYOL Mehmet GÜRBÜZ Fikret	01	01	23	2002
Bir Yatırım Problemine Analitik Hiyerarşi Prosesi Yönteminin Uygulanması	*An Application of the Analytic Hierarchy Process for an Investment Problem*	UYAR Yavuz DİZDAR N. Ercüment KURT Mustafa	01	01	39	2002
Matematiksel Finans	*Mathematical Finance*	ÖNALAN Ömer	01	01	53	2002
Çok Değişkenli Kalite Kontrol Yaklaşımlarının Bir Değerlendirmesi	*An Assessment of Multivariate Quality Control Approaches*	BAKIR M. Akif KARACA Sezar	01	01	67	2002
Ardışık Örnekleme Planında CUSUM Kontrol Kartlarının Kullanımı	*The Use of CUSUM Control Charts in Squential Sampling Plans*	HAMURKAROĞLU Canan BACANLI Sevil	01	01	87	2002

ENDEKS-INDEX						
MAKALE İSMİ	NAME OF ARTICLE	YAZAR ADI NAME OF CONTRIBUTOR	CİLT VOLUME	NO NUMBER	SAYFA PAGE	BASIM YILI YEAR
Oransal ve Çarpımsal Tahmin Ediciler	*Ratio Type and Product Estimators*	KADILAR Cem Hülya ÇINGI	01	01	101	2002
Basılı Türkçe'nin Önemli Bazı İstatistiksel Özellikleri	*Some Important Statistical Properties of Printed Turkish*	DALKILIÇ E. Mehmet DALKILIÇ Gökhan	01	01	113	2002
Tiroit Bezi Verilerinin Bayes ve En Yakın K-Komşu Gibi Eğiticili Yöntemlerle Sınıflanması	*Classification of Thyroid Gland Data by Supervised Methods Like Bayes and K-Nearest Neighbour*	ALBAYRAK Songül	01	01	131	2002
Hotelling T^2 Testinde Kovaryans Matrislerinin Heterojenliğinin I. Tip Hata Üzerine Etkisi: Bir Simülasyon Çalışması	*Heterogenity of Covariance Matrices Effect on Type-I Error in Hotelling-T^2 Test: A Simulation Study*	KANIK E. Arzu ÇAMDEVİREN Handan GÜRBÜZ Fikret	01	03	1	2002
Farklı Ortalama Vektörü ve Farklı Kovaryans Matrisi Koşullarında Dört Değişkenli Lojistik Regresyon Modeli ve Diskriminant Analizine Ait Doğru Sınıflandırma Olasılıklarının Simülasyon Tekniği Yardımıyla Karşılaştırılması	*Probabilities of Correctly Classifying Belong to Logistic Regression Model and Discriminant Analysis with four Variables Compares with Simulation Technique in Condition that Different Mean Vector and Different Covariance Matrix*	ÇAMDEVİREN Handan KANIK E. Arzu GÜRBÜZ Fikret	01	03	13	2002
Ortogonal Düzenler	*The Orthogonal Arrays*	BAYRAK Hülya ALHAN Aslıhan	01	03	25	2002
Varyansı Bilinmeyen Normal Dağılımlı Kitlenin Ortalamasının Ardışık Tahmini	*Sequential Estimation of the Mean of a Normal Population with Unknown Variance*	KOÇBERBER Güler Hülya ÇINGI	01	03	35	2002

ENDEKS-INDEX

ENDEKS-INDEX

ENDEKS-INDEX

ENDEKS-INDEX

ENDEKS-INDEX

MAKALE İSMİ	NAME OF ARTICLE	YAZAR ADI *NAME OF CONTRIBUTOR*	CİLT *VOLUME*	NO *NUMBER*	SAYFA *PAGE*	BASIM YILI *YEAR*
Kümelenmiş Verilerde Bağımlı İki Oranın Karşılaştırılması	*Comparison of Two Correlated Proportions for Clustered Data*	GÖKMEN Derya GENÇ Yasemin ATAKURT Yıldır YAĞMURLU Banu	03	03	21	2005
Lojistik ve Cox Regresyon Modellerinin İncelenmesi ve Karşılaştırılması	*Review of Logistic and Cox Regression Models*	SERTKAYA Durdu	03	03	31	2005
Üniversitelerarası Kurul Yabancı Dil Sınavı (ÜDS) Sonuçlarının Loglineer Analiz ile İncelenmesi	*Evaluation of the Results of the UDS by Loglinear Models*	ALTAŞ Dilek YILDIRIM Z. Esen	03	03	43	2005
Çok Değişkenli Veri Kümeleri Üzerinde Tanımlı Aykırı Değer Belirleme Tekniklerinin Simülasyon Çalışması ile Karşılaştırılması	*A Comparison of the Multiple Outlier Detection Method for Multivariate Data by Simulation Study*	KIRAL Gülsen BİLLOR Nedret	03	03	55	2005
Türkiye'de 1980-1985 Dönemi Bölgelerin İl Merkezi, İlçe Merkezi ile Bucak ve Köyler Bakımından Daimi İkametgaha göre İç Göçler	*The Internal Migration by the Permanent Residence Among the Regions in Turkey in View of Centers of Province, District and Subdistrict and Villages, in the Period 1980-1985*	ÇELEBİOĞLU Salih	03	03	67	2005
Whittaker Düzeltme Yöntemi ve Türkiye Ölüm Oranlarına Bir Uygulaması	*Whittaker Graduation Method and an Application to Turkish Mortality Rates*	TUZGÖL Hatice GÜNAY Süleyman	03	03	85	2005

ENDEKS-INDEX						
MAKALE İSMİ	NAME OF ARTICLE	YAZAR ADI *NAME OF CONTRIBUTOR*	CİLT *VOLUME*	NO *NUMBER*	SAYFA *PAGE*	BASIM YILI *YEAR*
Otoregresif Modellerin Bayes Analizinin Hava Kirliliği Verilerine Uygulaması	*Bayesian Analysis of Autoregresive Models with an Application to Air Pollution Data*	CENGİZ M. Ali EĞRİOĞLU Erol	04	01	1	2005
Meta-Analizi ve Bir Uygulama	*Meta Analysis and an Application*	SERTKAYA Durdu EROĞLU Aydan	04	01	13	2005
Konjoint Model Uygunluğu Üzerinde Değişken Etkilerinin Sınıflama Regresyon Ağaçları Analizi ile İncelenmesi	*Examining the Effects of the Variables on Conjoint Model Fitness with Classification and Regression Trees Analysis*	SARAÇLI Sinan DOĞAN İsmet	04	01	27	2005
Çoklu Doğrusal Bağlantı Durumunda Ridge Regresyon ve Temel Bileşenler Regresyon Yöntemlerinin Benzetim Çalışması ile Karşılaştırılması	*The Comparison of Ridge Regression and Principal Components Regression Methods in the Problem of Multicollinearity by Simulation*	ORTABAŞ Neslihan KURT Serdar	04	01	35	2005
Türkiye için İhracat Beklenti Endeksi Oluşturulması	*Building up an Export Expectation Index for Turkey*	ATABEK Aslıhan ŞAHİNÖZ Saygın COŞAR E. Evren	04	01	43	2005
Yapay Sinir Ağları ile Lojistik Regresyon Analizinin Karşılaştırılması	*Comparison of Artificial Neural Networks and Logistic Regression Analysis*	KURT İmran TÜRE Mevlüt	04	01	57	2005
Basit Doğrusal Otoregresif Modeller Sisteminde Parametre Tahmini ve Hipotez Testi: Simetrik İnovasyonlar	*Estimation of Parameters and Hypothesis Testing in the System of Simple Autoregressive Models: Symmetric Innovations*	TÜRKER Özlem AKKAYA D. Ayşen	04	01	75	2005

ENDEKS-INDEX

Not: İstatistik Araştırma Sempozyumu Bildirileri'nin yer aldığı *Özel Sayılar*, Endekse dahil edilmemiştir.

1: İstatistik Araştırma Dergisi, Ağustos 2005 sayısında hatalı basım olması nedeniyle, düzeltilmiş makale, Aralık 2005 sayısında sayfa 76'da tekrar yayınlanmaktadır.

ANNOUNCEMENT BOARD

Name of the Activity	Date	Place	Information	E-mail	Website
11th Annual Berlin SCORUS (Standing Committee on Regional and Urban Statistisc) Conference	January 8-10 2007	Berlin, Germany	Dr. Elsner	profelsner@aol.com	www.scorusnet.com
The 4th Meeting of the EMR-IBS (Eastern Mediterranean Region of the International Biometric Society)	January 22-26 2007	Eliat, Isreal	David Zucker	mszucker@mscc.huji.ac.il	www.congress.co.il/emr-ibs2007
32nd CEIES (Committee on Statistical Information in the Economic and Social Spheres) Seminar	February 5-6 2007	Arhus, Denmark	CEIES Secretariat	estat-ceies@ec.europa.eu	http://forum.europa.eu.int
28th Linz Seminar on Fuzzy Set Theory	February 6-10 2007	Linz, Austria	Erich Peter Klement	ep.klement@iku.at	www.flll.jku.at
Workshop on Statistical Extremes and Environmental Risk	February 15-17 2007	Lisbon, Portugal	M. İsabel Frage Alves	isabel.alves@fc.ul.pt	http://seer2007.fc.ul.pt

ANNOUNCEMENT BOARD

Name of the Activity	Date	Place	Information	E-mail	Website
International Conference on Applied Statistics for Development in Africa	February 26- March 2 2007	Cotonou, Benin	Eugène Azatassou	sada2007@univ-pau.fr	http://lma.iniv-pau.fr
The 4th International Conference SETIT (Sciences of Electronic, Technology of Information and Telecommunications), 2007	March 25-29 2007				www.setit.rnu.tn
3rd Brazilian Conference on Statistical Modelling in Insurance and Finance	March 25-29 2007	Maresias, Brazil		ubatuba@ime.usp.br	www.ime.usp.br
The First Joint Statistical Conference	March 27-30 2007	Biefield, Germany		dagstat2007@uni-bielefeld.de	www.statistik2007.de
Joint ENBIS (European Network of Business and Industry Statistics) DEINDE 2007 Conference, Computer Experiments versus Physical Experiments	April 11-13 2007	Torino, Italy	Ennio Davide Isaia	deinde07@econ.unito.it	http://web.econ.unito.it

ANNOUNCEMENT BOARD					
Name of the Activity	**Date**	**Place**	**Information**	**E-mail**	**Website**
32nd Annual Spring Lectures in the Mathematical Sciences on the theme:" Spatial and Spatio Temporal Statistics	April 12-14 2007	Arkansas, USA	Victor De Oliveira	vdo@uark.edu	http://comp.uark.edu
The 8th Seminaire European de Statistique on Statistics for Stochastic Differential Equation Systems	May 6-12 2007	La Mangadel Mar Menor, Spain	Mathieu Kessier	semstat@upct.es	www.dmae.upct.es
Second ICCRA 2 (International Conference on Cancer Risk Assessment)	May 25-27 2007	Santorini, Greece	Christos P. Kitsos	xkitsos@teiath.gr	
XIIth ASMDA (International Conference on Applied Stochastic Models and Data Analysis)	May 29-June 1 2007	Crete, Greece	Christos H. Skiadas	skiadas@asmda.com	www.asmda.com
The Second Baltic-Nordic Conference on Survey Sampling	June 2-7 2007	Kuusamo, Finland			www.mathstat.helsinki.fi

ANNOUNCEMENT BOARD

Name of the Activity	Date	Place	Information	E-mail	Website
The 6th International Workshop on "Objective Bayesian Analysis"	June 8-12 2007	Roma, Italy	M.M. Barbieri	brunero.liseo@uniroma1.it	http://3w.eco.uniroma1.it
The 35th Annual Meeting of the Statistical Society of Canada	June 8-12 2007	Nevfoundland, Canada	Brajendra Sutradhar	bsutradh@math.mun.ca	
The Third International Conference on Establishment Surveys (ICES-III)	June 18-21 2007	Quèbec, Canada		ices3@census.gov	www.amstat.org
The International Environmetrics Society North American Regional Meeting	June 19-21 2007	Washington, USA			www.stat.washington.edu
The ISF (International Symposium on Forecasting) is the Premier Forecasting Conference	June 24-27 2007	New York City, USA	Marriott Marquis		www.forecasters.org

ANNOUNCEMENT BOARD

Name of the Activity	Date	Place	Information	E-mail	Website
The 2007 Taipei International Statistical Symposium and ICSA International Conference	June 25-27 2007	Taipei, Taiwan	Chi-Lun Cheng		http://wiki.stat.ucia.edu
The Bi-annual Conference of the European Survey Research Association ESRA 2007	June 25-29 2007	Prague, Chech Republic	Martin Zeleny	esra2007@vse.cz	http://esra2007vse.cz
International Workshop on New Directions in Monte Carlo Methods	June 25-29 2007	Fleurance, France	Douch Randal	adapmc07@ens.fr	www.adamc07.enst.fr
The 5th International Conference on Multiple Comparison Procedures	July 9-11 2007	Vienna, Austria			www.mcp-conference.org
Sixth International Conference on Lattice Path Combinatorics and Applications	July 12-14 2007	Tennessee, USA	Anant Godbole	godbolea@etsu.edu	www.etsu.edu

ANNOUNCEMENT BOARD					
Name of the Activity	**Date**	**Place**	**Information**	**E-mail**	**Website**
Joint Statistical Meeting	July 29- August 2 2007	Utah, USA			www.amstat.org
The 28th Annual Conference of the International Society for Clinical Biostatistics	July 29- August 2 2007	Alexanroupoli, Greece	Vana Sypsa	vsipsa@cc.uoa.gr	www.iscb2007.gr
Joint Statistical Meeting	August 3-7 2007	Denver, USA			www.amstat.org
32nd Conference on Stochastic Process and their Applications	August 5-11 2007	Illinois, USA		spa07@math.uiuc.edu	www.math.uiuc.edu
Satelite Summer School on Levy Process	August 9-12 2007	Sandbjerg, Denmark			www.math.ku.dk

ANNOUNCEMENT BOARD					
Name of the Activity	**Date**	**Place**	**Information**	**E-mail**	**Website**
The Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy	August 11-17 2007	Coventry, U.K.	Janet Ainley	janet.ainley@warwick.ac.uk	http://srtl.stat.auckland.ac.nz
The 5th International Conference on Levy Processs	August 13-17 2007	Copenhagen, Holland			
The Annual Meeting of the TIES (International Environmetrics Society) Satellite Meeting	August 16-20 2007	Mikulov, Czech Republic			www.math.muni.cz
ISBIS (International Society for Business and International Statistics)	August 18-20 2007	Canada	Bovas Abraham	babraham@uwaterloo.ca	
The International Microsimulation Association, 1st General Conference	Augustt 20-22 2007	Vienna, Austria		IMA2007@euro.centre.org	www.microsimulation.org

ANNOUNCEMENT BOARD					
Name of the Activity	**Date**	**Place**	**Information**	**E-mail**	**Website**
International Statistical Institute, 56th Biennial Session	August 22-29 2007	Lisbon, Portugal	ISI Permanent Office	isi@cbs.nl	www.isi2007.com.pt
The IACS (International Association for Statistical Computing) for Data Mining, Learning and Knowledge Extraction	August 30-September 1 2007	Aveiro, Portugal	Carlos Ferreira & Paula Brito		www.mat.ua.pt
Small Area Estimation, 2007	September 3-5 2007	Granada, Spain	Carmen Batanero	batanero@ugr.es	www.ugr.es
Fourth International Meeting on Statistical Implicative Analysis	October 18-21 2007	Spain	Regis Gras	rencontreASI4@polytech. univ-nantes.fr	www.asi4.uji.es

ANNOUNCEMENT BOARD					
Name of the Activity	**Date**	**Place**	**Information**	**E-mail**	**Website**
Fourth International Conference on Agriculture Statistics	October 22-24 2007	Beijing, China		icas4@stats.gov.cn	www.stats.gov.cn
International Conference on Multiple Decisions and Related Topics in Honor of D.Y. Huang	December 28-30 2007	Taipei County, Taiwan	Ming-Chung Yang & Sheng-Tsaing Tseng	yang@stat.ncu.edu.tw sttseng@stat.nthu.edu.tw	
Joint ICMI/IASE (The International Commission on Mathematical Instruction/ Information Assuranca Support Environment) Study	June 30-July 4 2008	Monterrey, Mexico		2007symp@stat.sinica.edu.tw	www.stat.sinica.edu.tw
XXIVth International Biometric Conference 2008	July 13-18 2008	Dublin, Ireland		sinead@conferencepartners.ie	www.conferencepartners.ie

ANNOUNCEMENT BOARD					
Name of the Activity	**Date**	**Place**	**Information**	**E-mail**	**Website**
Fifth European Congress of Mathematics	July 14-18 2008	Amsterdam, Holland			www.emis.de
The Seventh Joint Meeting of the Bernouilli Society and the Institute of Mathematical Statistics	July 14-19 2008	Singapore		wc2008@ims.nus.edu.sg	www.ims.nus.edu.sg
COMPSTAT 2008: International Conference on Computational Statistics	August 24-29 2008	Porto, Portugal		compstat08@fep.up.pt	www.fep.up.pt
The 4th World Conference on Computational Statistics and Data Analysis of the IASC	December 5-8 2008	Yokohama, Japan		iasc2008@ism.ac.jp	www.iasc-ars.org/

METİN HAZIRLAMA KALIBI

1. Araştırma, yazılar, kaynaklar, tablo ve şekiller ile birlikte en az 2 en çok 15 sayfa olmalıdır.

2. Gönderilecek araştırma PC ortamında Word 7.0 veya daha yukarı versiyonları ile Times New Roman font ortamında yazılmalıdır.

3. Araştırma A4 normundaki beyaz kağıda sol ve üstten 3,5 cm, sağ ve alttan 2,5 cm boşluk bırakılarak yazılmalıdır.

4. Araştırmanın Türkçe ve İngilizce başlıkları metne uygun olmalıdır. Araştırmanın başlıkları büyük harflerle 14 punto harf büyüklüğünde koyu olarak yazılmalı ve özet büyük harflerle ortalı, 12 punto harf büyüklüğünde koyu olarak yazılmalıdır.

5. Yazarın adı ve soyadı, unvan belirtilmeden başlığın iki satır altından ortalı olarak ad küçük, soyad büyük harfli olarak yazılmalıdır. İki veya daha fazla yazar olması durumunda, yan yana kolon (sütun) açılarak yazılmalıdır.

6. Yazarın adresi dip not şeklinde verilerek yıldız (*) ile gösterilmelidir. Birden fazla yazar söz konusu olduğunda, yazışmaların hangi yazar ve adresle yapılacağı ise parantez içinde (haberleşme adresi) yazılarak verilmelidir. Dipnot vermek gerektiğinde de yıldız (*) kullanılmalıdır. Yazar(lar)ın adresi ve dipnot ilgili sayfanın altına Times New Roman font ve 10 punto harf büyüklüğü kullanılarak yazılmalıdır.

7. Çalışma herhangi bir kurumun desteği ile gerçekleştirilmişse, kurumun adı ilk sayfa altında dipnot olarak yazılmalıdır.

8. Araştırma bölümleri; Türkçe özet, Araştırma metni, Kaynaklar ve İngilizce özet (Abstract) şeklinde olmalıdır.

- Türkçe özet, yazar isminden sonra üç satır boşluk bırakılarak yazılır. 200 kelimeyi geçmeyecek şekilde soldan 5,5 cm ve sağdan 4,5 cm boşluk bırakılarak, 11 punto harf büyüklüğü kullanılarak, italik olarak yazılmalıdır.

- Araştırma metni 12 punto harf büyüklüğü kullanılarak bir satır aralığında ve paragraflar arasında bir satır boşluk bırakılmalıdır. Paragraflar ve formüller bir tab içeriden yazılmalıdır. Birinci derece bölüm başlıkları büyük harfle, ikinci derece alt bölüm başlıklarında her sözcüğün ilk harfi büyük, diğerleri küçük harfle, üçüncü ve daha alt derece alt bölüm başlıklarının yalnız ilk harfi büyük, diğerleri küçük harfle yazılmalıdır. Bütün bölüm başlıkları koyu olarak yazılmalıdır, tablo ve şekillere başlık ve sıra numarası bölüm numarası içermeksizin verilir. Tablo ve şekil başlık ve sıra numaraları yarım satır aralıklı tablolarda üstte, şekillerde altta yer almalıdır.

- Kaynaklara göndermeler metin içinde açılan ayraçlarla yapılmalıdır. Ayraç içindeki sıra şöyledir: Yazar(lar)'ın soyadı ve kaynağın yılı. Örneğin; ...kanıtlanmıştır (Rao, 1974)., ...(Grossman ve Weiss, 1983)., ...(Baumal, 1952; Tobin, 1956)., ... (Winebrake vd., 1995)., ...Rao (1974) kanıtlamıştır vb. şeklinde gösterilmelidir.

  Çalışmada gönderme yapılan bütün kaynaklar, kaynaklar listesinde belirtilmeli; çalışmada yararlanılmayan kaynaklar, kaynaklar listesinde yer almamalıdır. Kaynaklar araştırma metninin sonunda yazarının soyadına göre alfabetik sırada ve 11 puntoda kaynaklar arasında bir satır boşluk bırakılarak yazılmalıdır. Bunların yazım şekli aşağıda gösterildiği gibi standart formda olmalıdır:

Örnekler:

Kitap

BRUBAKER, S., (1967). *Trends in the World Aluminium Industry*, Baltimore, Maryland: John Hopkins Press.

Araştırma

RAO, J.N.K., (1994). *Estimating Totals and Distribution Function Using Auxiliary Information at the Estimation Stage*, Journal of Official Statistic, 10, 153 – 165.

Derleme

ARTHUR, W.B., (1988). *Competing Technologies: An Overview*, G.Dosi, C. Freeman, R. Nelson, G. Silverberg ve L. Soete (der.), Technical Change and Economic Theory içinde Londra:Pinter, 590-607.

Internet

SUTCLIFFE, M.J., Wo, Z.G. and OSWALD, R.E., (1996). *Three-DimensionalMmodels of Non-NMDAglutamatereceptors,* Erişim: [http://neon.chem.le.ac.uk/cornell/ Sutcliffe_BJ/Sutcliffe_BJ.html]. Erişim Tarihi: 22.12.1996.

- Araştırmanın İngilizce dilde özeti araştırmanın sonunda verilmelidir. Araştırmanın İngilizce adı üstten 2 satır boşluk bırakılarak ortalı, büyük harflerle, 14 punto harf büyüklüğünde koyu olarak yazılmalı, Abstract büyük harflerle ortalı, 12 punto harf büyüklüğünde koyu olarak yazılmalıdır. İngilizce özet soldan 5,5 cm ve sağdan 4,5 cm boşluk bırakılarak 200 kelimeyi geçmeyecek şekilde 11 punto harf büyüklüğünde italik olarak, araştırmanın İngilizce adından sonra 3 satır boşluk bırakılarak yazılmalıdır.

- Anahtar kelimeler (Key words) her iki özetin bir satır altına, Anahtar Kelimeler ve Key Words koyu italik olarak yazılmalıdır.

9. Matematik simge ve formüllerin yazımında aşağıdaki hususlara dikkat edilir:

- Simgelerin ayırt edilmesi önemlidir. Özellikle büyük ve küçük harfler, düz ve koyu harfler, Klasik Yunan ve Latin harfleri, alt ve üst indisler, sıfır (0) rakamı ve O harfi, Bir (1) rakamı ve le (l) harfi ayırt edilebilmelidir. Çoklu indislerden sakınılmalıdır.

- Denklemler word, standart (default) ölçülerde 1 tab (1,27 cm) içerden ve numara vermek gerekliyse bölüm numarasını içermeksizin en sağına parantez içinde yazılmalıdır. Uzun formüller metin içinde yer almamalıdır.

- Kesirler, metin içinde (/) işareti ile gösterilmelidir.

- Karmaşık ifadeler içeren denklemler olabildiğince kısaltma simgeleri kullanılarak yazılmalıdır.

- İç içe çoklu ayraçlar aynı formülde yer aldığında, sıra düzeni örneğin {[(**0**)]} biçiminde olmalıdır.

10. Araştırmanın Türkçe yazım kurallarına uygun olması yazarın sorumluluğu altındadır.

Açıklama: ***Yukarıda verilen açıklamalar, Dergi'nin 2005 yılı Aralık sayısındaki makalelerin yazım biçimine yöneliktir. 2007 yılından itibaren, Dergi'de yayınlanacak makalelerin yazım biçimi ve kurallarında Dergi Editörlüğü tarafından güncelleme çalışması yapılmış ve bu konuda bir Kılavuz hazırlanmıştır. Kılavuz TÜİK web sayfasından indirilebilecek ve bu konudaki duyurular elektronik posta yolu ile yapılacaktır.***

İSTATİSTİK ARAŞTIRMA DERGİSİ
JOURNAL OF STATISTICAL RESEARCH

Cilt 04 *Volume* **No 03** *Number* **Aralık 2005** *December*