ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.:4 - Sayı/No: 1 : 59-66 (2003)

ARAŞTIRMA MAKALESİ/**RESEARCH ARTICLE**

DİSKRİMİNANT ANALİZİNDE GERÇEK HATA ORANI TAHMİN EDİCİLERİNİN JACKKNIFE VE BOOTSTRAP DEĞERLENDİRMESİ

Cemal ATAKAN[1]

ÖZ

Bu çalışmada diskriminant analizinde iki çok değişkenli normal kitle olması durumunda, gerçek hata oranı için bazı hata oranı tahmin edicilerinin jackknife ve bootstrap değerlendirmesi simülasyon çalışmasıyla verilmektedir.

Anahtar Kelimeler: Diskriminant analizi, Hata oranı, Hata oranı tahmin edicisi, Jackknife ve Bootstrap.

AN EVALUATION OF THE JACKKNIFED AND BOOTSTRAPED ERROR RATE ESTIMATORS OF ACTUAL ERROR RATE IN DISCRIMINANT ANALYSIS

ABSTRACT

In this study the jackknifed and bootstraped version of the some error rate estimators for the actual error rate have been evaluated by simulation in discriminant analysis of two different multivariate normal populations.

Key Words: Discriminant analysis, Error rate, Error rate estimator, Jackknife and Bootstrap.

1. GİRİŞ

Diskriminasyon, üzerinde ölçüm yapılan bir bireyi sonlu sayıda bilinen farklı kitleden birine atanmasını gerçekleştiren istatistiksel bir tekniktir. Birey ait olduğu kitlelerden farklı bir kitleye atandığında bir hata yapılmış olur. Diskriminant analizinde amaç, atama işlemini mümkün olan minimum hatayla yapmaktır. Bu optimizasyon kriterine göre elde edilen diskriminant fonksiyonlarının değerlendirilmesinde hatalı sınıflandırma olasılığının (hata oranı) bilinmesi önemlidir. Hata oranı, dağılımların ve parametrelerin bilinip bilinmemesine, örneklem hacmi ve atama kuralı gibi değişik kriterlere bağlı olarak hesaplanır.

Bu çalışmada sadece iki kitlenin olduğu durum göz önüne alınacaktır. İkiden çok kitle için de benzer ifadeler elde edilebilir.

Π_1 ve Π_2 birbirinden farklı iki kitle olmak üzere, $\underline{X}' = (X_1, X_2,..., X_p)$ birey üzerinde ölçümlere karşılık gelen p boyutlu rasgele değişkeni, Π_i (i=1,2)'den bir rasgele değişken ise $\underline{X}'$ in olasılık yoğunluk fonksiyonu $f_i\,(\underline{x}_i\,;\,\underline{\theta}_i)$ biçiminde gösterilir. Burada $\underline{\theta}_i$ parametre vektörüdür.

$\underline{X}$ vektörünün aldığı değerler p boyutlu $\Re^p$ örneklem uzayında olmak üzere sınıflandırma problemi bu uzayı $B_1 \cup B_2 = \Re^p$ ve $B_1 \cap B_2 = \phi$ olan öyle B_1,B_2 bölgelerine ayırır ki, eğer $\underline{X}$ gözlemi B_1'de ise $\underline{X}$ gözlemine sahip birey Π_1'e; aksi halde Π_2'ye atanır.

Bu çalışmada atama işlemi, olabilirlik oranı kriterine (Welch, 1939) göre elde edilen lineer diskriminant fonksiyonu göz önüne alınarak yapılmaktadır.

Π_1 ve Π_2, $\underline{\mu}_1$ ve $\underline{\mu}_2$ ortalamalı ve ortak Σ varyans - kovaryans matrisine sahip çok değişkenli normal dağılama sahip kitleler olmak üzere, parametreler bilinmediğinde bu kitlelerden alınan n_1 ve n_2 hacimli rasgele örneklemlerden elde edilen $\underline{\overline{X}}_1$, $\underline{\overline{X}}_2$ örneklem ortalamaları ve S birleştirilmiş örneklem varyans - kovaryans matrisi tahmin edicilerine bağlı optimal atama kuralı,

$$\hat{\xi}:\begin{cases} W(\underline{X}) > k \text{ ise} & \underline{x},\ \Pi_1\text{'}e, \\ \textit{diger durumda} & \Pi_2\text{'ye } \textit{ata.} \end{cases} \tag{1}$$

[1] Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Tandoğan, ANKARA/TÜRKİYE
E-posta: atakan@science.ankara.edu.tr

Geliş: 07 Kasım 2002; **Kabul:** 24 Mart 2003

ile verilir;burada,

$$W(\underline{X})=\left[\underline{X}-\frac{1}{2}(\underline{\bar{X}}_1+\underline{\bar{X}}_2)\right]' S^{-1}(\underline{\bar{X}}_1-\underline{\bar{X}}_2) \quad (2)$$

olup, parametreler bilinmediğinde $f_i(\underline{x}_i;\hat{\underline{\theta}}_i)$ ($\hat{\underline{\theta}}_i=(\underline{\bar{X}}_i,S)$, i=1,2) olasılık yoğunluk fonksiyonlarının birbirine oranlanmasıyla elde edilen örneklem doğrusal diskriminant fonksiyonu, $k=\ln(k_1)$, $k_1=\frac{q_2c_2}{q_1c_1}$ ve q_i ile c_i (i = 1,2) sırasıyla önsel (prior) olasılıklar ve hatalı atamanın maliyetleridir (Anderson,1984). Bu çalışmada maliyetler ihmal edilip, önsel olasılıklar eşit alındığından k=0 dır.

Çalışmanın ikinci bölümünde diskriminant analizinde tanımlanan hata oranları ve üçüncü bölümde bu hata oranlarından gerçek hata oranına ilişkin literatürde bilinen tahmin edicilerin bazıları detaylarına girilmeden verilmektedir. Dördüncü bölümde jackknife ve bootstrap teknikleri kısaca verildikten sonra, beşinci bölümde hata oranı tahmin edicilerinin jackknife ve bootstrap değerlendirmesi ait simülasyon sonuçları ve son bölümde de elde edilen sonuçların değerlendirilmesi verilmektedir.

2. HATA ORANLARI

Atama işlemi yapılırken, gözlem, ait olduğu kitleye değil de diğer kitleye atanırsa, hata yapılmış olur. Atama kurallarına ilişkin optimal, (koşullu) gerçek ve beklenen gerçek hata oranı gibi farklı hata oranları tanımlanmıştır (Anderson 1984, Hills 1966, Lachenbruch and Mickey 1968, Lachenbruch 1975, Seber 1984, Snapinn 1983).

Optimal hata oranı, parametreler bilindiğinde elde edilen diskriminant fonksiyonu göz önüne alınarak bulunan hata oranıdır. Gerçek hata oranı, parametreler bilinmediğinde örneklemlerden elde edilen tahminlere bağlı örneklem diskriminant fonksiyonuna göre bulunan hata oranıdır ve beklenen gerçek hata oranı, tüm örneklemler üzerinden gerçek hata oranının beklenen değeridir.

Bu çalışmada gerçek hata oranı gözönüne alınmıştır. Böylece (1)'de verilen $\hat{\xi}$ atama kuralına göre bilinmeyen bir $\underline{X}$ gözleminin Π_1'den olduğu bilindiğinde, Π_2'ye hatalı atanması durumunda gerçekleşen gerçek hata oranı,

$$\alpha_1(\hat{\xi})=P(W(\underline{X})\leq 0/\underline{X}\in\Pi_1)$$

$$=\Phi\left\{-\frac{\left[\underline{\mu}_1-\frac{1}{2}(\underline{X}_1+\underline{X}_2)\right]' S^{-1}(\underline{X}_1-\underline{X}_2)}{\left[(\underline{X}_1-\underline{X}_2)'S^{-1}\Sigma S^{-1}(\underline{X}_1-\underline{X}_2)\right]^{1/2}}\right\} \quad (3)$$

dir; burada $\Phi\{.\}$ standart normal dağılım fonksiyonudur. Π_2'den bir $\underline{X}$ gözleminin hatalı atanması durumu için de benzer ifade elde edilebilir.

3. HATA ORANI TAHMİN EDİCİLERİ

Hata oranları diskriminant fonksiyonunun dağılımına bağlı olarak bulunabilir. Ancak parametrelerin bilinmediği durumda elde edilen örneklem doğrusal diskriminant fonksiyonunun dağılımı oldukça karmaşık olduğundan hata oranları için analatik ifadeler bulmak oldukça zordur. Ancak bazı şartlar altında örneklem diskriminant fonksiyonunun dağılımı bulunabilmektedir ve bu şartlara bağlı olarak da hata oranları için karmaşık ifadeler elde edilebilmektedir (Sitgreaves, 1961). Bu zorluktan dolayı, hata oranları değişik tahmin ediciler yardımı ile tahmin edilmeye çalışılmıştır.

Bu çalışmada gerçek hata oranına ilişkin literatürde bilinen hata oranı tahmin edicilerinden D, DS, L, LS, M, MS, O ve OS tahmin edicileri göz önüne alınmıştır. Bu tahmin edicilerin hepsi normallik varsayımı altında elde edilen parametrik tahmin ediciler sınıfına ait tahmin edicilerdir. Parametrik olmayan tahmin ediciler sınıfına ait tahmin ediciler burada verilmeyeceklerdir. Çalışmada, yukarıda verilen tahmin edicilerin birbirlerine göre üstünlüklerindense, bu tahmin edicilerin jackknife ve bootstrap tekniklerine karşı davranışları değerlendirilmeye çalışılmaktadır.

$\hat{\alpha}$, gerçek hata oranı için herhangi bir tahmin edici olsun. Π_1'den rasgele bir gözlemin Π_2'ye hatalı atanması olasılığı için bazı hata oranı tahmin edicilerinin analitik ifadeleri aşağıda verilmektedir. Π_2'den rasgele bir gözlemin hatalı atanması için de benzer ifadeler verilebilir. Hata oranı tahmin edicilerinin daha detaylı ifadeleri Lachenbruch and Mickey (1968), Lachenbruch (1975), Snapinn (1983) de verilmiştir.

3.1. D Tahmin Edicisi

Fisher tarafından önerilen bu tahmin ediciye, Plug-in Tahmin edicisi de denmektadir (Fisher, 1936). Burada, bilinmeyen parametreler yerine örneklemlerden elde edilen tahminlerinin gerçek hata oranında doğrudan kullanılmasıyla

$$\hat{\alpha}_1^D=\Phi\left\{-\frac{\left[\underline{\bar{X}}_1-\frac{1}{2}(\underline{\bar{X}}_1+\underline{\bar{X}}_2)\right]' S^{-1}(\underline{\bar{X}}_1-\underline{\bar{X}}_2)}{\left[(\underline{\bar{X}}_1-\underline{\bar{X}}_2)'S^{-1}SS^{-1}(\underline{\bar{X}}_1-\underline{\bar{X}}_2)\right]^{1/2}}\right\}$$

$$=\Phi\left\{-\frac{D}{2}\right\} \quad (4)$$

biçiminde elde edilir, burada $D^2=(\underline{\bar{X}}_1-\underline{\bar{X}}_1)'S^{-1}(\underline{\bar{X}}_1-\underline{\bar{X}}_1)$, iki kitle arasındaki Mahalanobis kare uzaklığı $\Delta^2=(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1-\underline{\mu}_2)$'nin tahmin edicisidir.

3.2. DS Tahmin Edicisi

D^2, Δ^2 için yanlı bir tahmin edici olduğundan, Lachenbruch ve Mickey (1968), Δ^2'nin yansız bir tahmin edicisi olarak

$$(DS)^2 = c_1 D^2 \tag{5}$$

eşitliğinden DS tahmin edicisini elde etmişlerdir. DS tahmin edicisi, D tahmin edicisinde D yerine DS alınmasıyla

$$\hat{\alpha}_1^{DS}=\Phi\left\{-\sqrt{c_1}\frac{D}{2}\right\} \tag{6}$$

biçiminde elde edilir; burada,

$$c_1=\frac{n_1+n_2-p-3}{n_1+n_2-2}$$

dir.

3.3. L Tahmin Edicisi

Hata oaranları için Lachenbruch (1968) de yaklaşık ifadeler elde etmiştir. Bu ifadelerde bilinmeyen Δ^2 parametresi yerine D^2 tahmin edicisi alınarak L tahmin edicisi,

$$\hat{\alpha}_1^{L}=\Phi\left\{-\frac{\frac{c_2}{2}\left(D^2-\frac{p(n_2-n_1)}{n_1n_2}\right)}{\sqrt{c_3\left(D^2+\frac{p(n_1+n_2)}{n_1n_2}\right)}}\right\} \tag{7}$$

olarak elde edilir; burada,

$$\begin{aligned} c_2&=\frac{n_1+n_2-2}{n_1+n_2-p-3}\\ c_3&=\frac{(n_1+n_2-3)(n_1+n_2-2)^2}{c_4c_5c_6}\\ c_4&=(n_1+n_2-p-2)\\ c_5&=(n_1+n_2-p-3)\\ c_6&=(n_1+n_2-p-5)\end{aligned} \tag{8}$$

dir.

3.4. LS Tahmin Edicisi

L tahmin edicisinde D^2 yerine (5)'de verilen $(DS)^2$ ifadesi alınarak LS tahmin edicisi,

$$\hat{\alpha}_1^{LS}=\Phi\left\{-\frac{\frac{1}{2}\left(D^2-c_2\frac{p(n_2-n_1)}{n_1n_2}\right)}{\sqrt{\frac{c_3}{c_2}\left(D^2+\frac{p(n_1+n_2)}{n_1n_2}\right)}}\right\} \tag{9}$$

elde edilmiş olur, burada

$$c_2\,,\,c_3 \tag{8}$$

ifadesinde verilen sabitlerdir (Lachenbruch, 1968).

3.5. M Tahmin Edicisi

McLachlan (1974), gerçek hata oranı için asimptotik yansız bir tahmin edici olan M tahmin edicisini aşağıdaki biçimde ifade etmiştir

$$\begin{aligned}\hat{\alpha}_1^{M}=&\Phi\left\{-\frac{D}{2}\right\}+\\ &\phi\left\{-\frac{D}{2}\right\}\{A_1+A_2+A_3+A_4+A_5\}\end{aligned} \tag{10}$$

burada, $\phi\{.\}$ standart normal dağılımın olasılık yoğunluk fonksiyonu ve

$$A_1=\frac{p-1}{n_1D},\quad A_2=\frac{D[4(4p-1)-D^2]}{32(n_1+n_2-2)},$$

$$A_3=\frac{D(p-1)(p-2)}{4n_1},$$

$$A_4=\frac{(p-1)[-D^3+8D(2p+1)+16/D]}{64n_1(n_1+n_2-2)},$$

$$A_5=\frac{D[a_1+a_2+a_3]}{12288(n_1+n_2-2)},$$

$$a_1=3D^6-4(24p+7)D^4,$$

$$a_2=16(48p^2-48p-53)D^2,$$

$$a_3=192(-8p+15),$$

dır.

3.6. MS Tahmin Edicisi

M Tahmin edicisinde D^2 yerine (5)'de verilen $(DS)^2$ ifadesinin alınmasıyla MS tahmin edicisi

$$\begin{aligned}\hat{\alpha}_1^{MS}=&\Phi\left\{-\sqrt{c_1}\frac{D}{2}\right\}+\\ &B_0\{B_1+B_2+B_3+B_4+B_5\}\end{aligned} \tag{11}$$

biçiminde elde edilir. Burada,

$$B_0=\phi\left\{-\frac{\sqrt{c_1}D}{2}\right\},\quad B_1=\frac{p-1}{n_1\sqrt{c_1}D},$$

$$B_2=\frac{\sqrt{c_1}D[4(4p-1)-c_1D^2]}{32(n_1+n_2-2)},$$

$$B_3=\frac{\sqrt{c_1}D(p-1)(p-2)}{4n_1},$$

$$B_5=\frac{\sqrt{c_1}D[b_2+b_3+192(-8p+15)]}{12288(n_1+n_2-2)},$$

$$b_1 = 8\sqrt{c_1}D(2p+1) + 16/\sqrt{c_1}D,$$

$$b_2 = 3(\sqrt{c_1}D)^6 - 4(24p+7)(\sqrt{c_1}D)^4,$$

$$b_3 = 16(48p^2 - 48p - 53)c_1 D^2,$$

dir (McLachlan, 1974).

3.7. O Tahmin Edicisi

Okamoto (1963) tarafından verilen hata oranı açılımlarında bilinmeyen Δ^2 parametresi yerine D^2 tahmin edicisinin konulmasıyla O tahmin edicisi,

$$\hat{\alpha}_1^O = \Phi\left\{-\frac{D}{2}\right\} + \phi\left\{-\frac{D}{2}\right\}\{E_1 + E_2 + E_3\} \tag{12}$$

biçiminde elde edilir; burada,

$$E_1 = \frac{\left[\frac{D}{16} + \frac{3(p-1)}{4D}\right]}{n_1}, E_2 = \frac{\left[\frac{D}{16} - \frac{(p-1)}{4D}\right]}{n_2},$$

$$E_3 = \left[\frac{D(p-1)}{4(n_1+n_2)}\right]$$

dir.

3.5. OS Tahmin Edicisi

O tahmin edicisinde D^2 yerine (5)'de verilen $(D)^2$ ifadesinin yazılmasıyla OS tahmin edicisi

$$\hat{\alpha}_1^{OS} = \Phi\left\{-\sqrt{c_1}\frac{D}{2}\right\} + \phi\left\{-\sqrt{c_1}\frac{D}{2}\right\} \{F_1 + F_2 + F_3\} \tag{13}$$

elde edilmiş olur (Anderson, 1984). Burada,

$$F_1 = \frac{\left[\frac{\sqrt{c_1}D}{16} + \frac{3(p-1)}{4\sqrt{c_1}D}\right]}{n_1},$$

$$F_2 = \frac{\left[\frac{\sqrt{c_1}D}{16} - \frac{(p-1)}{4\sqrt{c_1}D}\right]}{n_2},$$

$$F_3 = \left[\frac{\sqrt{c_1}D(p-1)}{4(n_1+n_2)}\right]$$

dır (Okamoto,1963).

4. JACKKNIFE VE BOOTSTRAP YÖNTEMLERİ

Bu bölümde yukarıda verilen hata oranı tahmin edicilerinin, Jackknife ve Bootstrap değerleri için ifadeler verilmektedir.

Jackknife yönteminde, Π_1 ve Π_2 kitlelerinden alınan n_1 ve n_2 hacimli rasgele örneklemlerdeki gözlemlerden her defasında sırasıyla bir gözlemin çıkarılmasıyla geriye kalan $n_1 + n_2 - 1$ tane gözlem kullanılarak üçüncü bölümde verilen hata oranı tahmin edicilerinden herhangi biri göz önüne alınarak hata oranı tahmin edilir. Bu işlem $n_1 + n_2$ tane gözlemin herbiri için tekrarlanarak $n_1 + n_2$ tane hata oranı tahmin değeri elde edilmiş olur. Bu tahmin değerlerinin toplamının $n_1 + n_2$'ye oranlanmasıyla, tahmin edicinin jackknife tahmin değeri bulunur.

$\hat{\alpha}$ herhangi bir hata oranı tahmin edicisi olmak üzere, bu tahmin edicinin jackknife tahmini

$$\hat{\alpha}_{Jackk} = \frac{1}{n_1+n_2}\sum_{i=1}^{n_1+n_2}\hat{\alpha}_{(i)} \tag{14}$$

dir. Burada $\hat{\alpha}_{(i)}$ i'inci gözlem çıkartıldıktan sonra geriye kalan $n_1 + n_2 - 1$ tane gözlemden elde edilen $\hat{\alpha}$ hata oranı tahmin edicisi için bir tahmin değeridir.

Bootstrap yönteminde ise, Π_1 ve Π_2 kitlelerinden alınan n_1 ve n_2 hacimli rasgele örneklemlerden iadeli çekiliş yapılarak oluşturulan yeni n_1 ve n_2 hacimli örneklemler göz önüne alınarak, üçüncü bölümde verilen herhangi bir hata oranı tahmin edicisi için tahmin değeri bulunur. Bu işlem bootstrap tekrar sayısı olan B defa tekrarlanır. Bu B tane hata oranı tahmin değerinin toplamının B'ye oranlanmasıyla hata oranı tahmin edicisinin, bootstrap tahmin değeri elde edilmiş olur.

$\hat{\alpha}$ hata oranı tahmin edicisinin bootstrap tahmini

$$\hat{\alpha}_{Boot} = \frac{1}{B}\sum_{b=1}^{B}\hat{\alpha}_b \tag{15}$$

dir, burada $\hat{\alpha}_b$, $\hat{\alpha}$ hata oranı tahmin edicisi için her bir bootstrap tekrarından elde edilen tahmin değeridir (Efron 1979, 1982, Efron and Gong 1983).

5. SİMÜLASYON ÇALIŞMASI

Bu bölümde simülasyonla gerçek hata oranına ilişkin üçüncü bölümde verilen tahmin edicilerin jackknife ve bootstrap değerlendirmesi, değişken sayısı $p = 2,3,5$ olan çok değişkenli eşit (ortak) kovaryanslı normal dağılımlı kitleler için $\Delta = 1,2,3$ değerlerine karşılık alınan $n_1 = n_2 = 15, 30$ ve 100 birimlik örneklemler göz önüne alınarak yapılmaktadır. Simülasyon çalışması Matlab paket programında 1000 tekrarlı ve bootsrap tekrar sayı $B = 200$ alınarak yapılmıştır. Sonuçlar Tablo 1,2 ve 3 de verilmektedir.

Tablo 1. $p = 2$ **İçin Hata Oranları Değeri.**

	$\Delta = 1$			$\Delta = 2$			$\Delta = 3$		
$n_1 = n_2 = 15$		**Jackk.**	**Boots.**		**Jackk.**	**Boots.**		**Jackk.**	**Boots.**
Gerçek Hata Oranı	0.3283	0.3297	0,3469	0.1693	0.1701	0.1811	0.0746	0.0753	0.0830
D Tahmin Edicisi	0.2916	0.2910	0.2679	0.1521	0.1518	0.1373	0.0648	0.0647	0.0568
DS Tahmin Edicisi	0.3017	0.3015	0.2784	0.1651	0.1653	0.1495	0.0753	0.0755	0.0661
L Tahmin Edicisi	0.3201	0.3206	0.2950	0.1736	0.1742	0.1573	0.0790	0.0795	0.0694
LS Tahmin Edicisi	0.3246	0.3253	0.3291	0.1801	0.1810	0.1865	0.0847	0.0855	0.0904
M Tahmin Edicisi	0.3280	0.3298	0.3041	0.1726	0.1734	0.1560	0.0751	0.0754	0.0656
MS Tahmin Edicisi	0.3393	0.3419	0.3162	0.1869	0.1882	0.1695	0.0872	0.0880	0.0762
O Tahmin Edicisi	0.3101	0.3191	0.2865	0.1637	0.1636	0.1481	0.0722	0.0705	0.0634
OS Tahmin Edicisi	0.2991	0.3066	0.2900	0.1503	0.1493	0.1503	0.0617	0.0597	0.0632
$n_1 = n_2 = 30$									
Gerçek Hata Oranı	0.3157	0.3160	0.3249	0.1633	0.1634	0.1693	0.0703	0.0705	0.0744
D Tahmin Edicisi	0.2985	0.2983	0.2876	0.1531	0.1530	0.1454	0.0648	0.0648	0.0606
DS Tahmin Edicisi	0.3032	0.3032	0.2925	0.1594	0.1594	0.1515	0.0698	0.0699	0.0652
L Tahmin Edicisi	0.3137	0.3138	0.3025	0.1634	0.1636	0.1554	0.0714	0.0715	0.0668
LS Tahmin Edicisi	0.3159	0.3161	0.3181	0.1665	0.1668	0.1697	0.0740	0.0742	0.0766
M Tahmin Edicisi	0.3157	0.3160	0.3051	0.1632	0.1634	0.1550	0.0700	0.0701	0.0653
MS Tahmin Edicisi	0.3207	0.3112	0.3103	0.1697	0.1700	0.1614	0.0753	0.0755	0.0703
O Tahmin Edicisi	0.3074	0.3117	0.2967	0.1590	0.1589	0.1511	0.0686	0.0677	0.0641
OS Tahmin Edicisi	0.3024	0.3064	0.2984	0.1526	0.1524	0.1526	0.0636	0.0627	0.0644
$n_1 = n_2 = 100$									
Gerçek Hata Oranı	0.3095	0.3096	0.3122	0.1593	0.1593	0.1611	0.0674	0.0674	0.0686
D Tahmin Edicisi	0.3056	0.3056	0.3023	0.1574	0.1574	0.1550	0.0669	0.0669	0.0655
DS Tahmin Edicisi	0.3069	0.3069	0.3037	0.1593	0.1593	0.1569	0.0683	0.0683	0.0669
L Tahmin Edicisi	0.3104	0.3104	0.3071	0.1605	0.1605	0.1581	0.0688	0.0688	0.0674
LS Tahmin Edicisi	0.3110	0.3110	0.3117	0.1614	0.1614	0.1623	0.0695	0.0695	0.0703
M Tahmin Edicisi	0.3106	0.3107	0.3074	0.1605	0.1605	0.1580	0.0684	0.0684	0.0670
MS Tahmin Edicisi	0.3120	0.3121	0.3088	0.1623	0.1624	0.1599	0.0699	0.0700	0.0685
O Tahmin Edicisi	0.3082	0.3096	0.3050	0.1592	0.1592	0.1568	0.0680	0.0678	0.0666
OS Tahmin Edicisi	0.3069	0.3082	0.3056	0.1574	0.1574	0.1574	0.0666	0.0663	0.0668

Tablo 2. $p = 3$ İçin Hata Oranları Değeri.

	$\Delta = 1$			$\Delta = 2$			$\Delta = 3$		
$n_1 = n_2 = 15$		**Jackk.**	**Boots.**		**Jackk.**	**Boots.**		**Jackk.**	**Boots.**
Gerçek Hata Oranı	0.3402	0.3419	0.3626	0.1780	0.1792	0.1969	0.0805	0.0817	0.0937
D Tahmin Edicisi	0.2760	0.2748	0.2406	0.1424	0.1418	0.1209	0.0599	0.0596	0.0489
DS Tahmin Edicisi	0.2904	0.2898	0.2558	0.1601	0.1601	0.1372	0.0738	0.0741	0.0607
L Tahmin Edicisi	0.3138	0.3142	0.2760	0.1716	0.1723	0.1471	0.0789	0.0795	0.0649
LS Tahmin Edicisi	0.3200	0.3207	0.3262	0.1804	0.1814	0.1891	0.0867	0.0876	0.0945
M Tahmin Edicisi	0.3401	0.3429	0.2986	0.1803	0.1818	0.1541	0.0804	0.0812	0.0655
MS Tahmin Edicisi	0.3579	0.3620	0.3375	0.2011	0.2034	0.1736	0.0982	0.0996	0.0806
O Tahmin Edicisi	0.3072	0.3227	0.2689	0.1612	0.1623	0.1377	0.0711	0.0691	0.0582
OS Tahmin Edicisi	0.2905	0.3029	0.2175	0.1425	0.1419	0.1392	0.0570	0.0545	0.0573
$n_1 = n_2 = 30$									
Gerçek Hata Oranı	0.3226	0.3231	0.3369	0.1675	0.1677	0.1771	0.0729	0.0732	0.0793
D Tahmin Edicisi	0.2943	0.2941	0.2765	0.1506	0.1505	0.1393	0.0617	0.0617	0.0557
DS Tahmin Edicisi	0.3008	0.3006	0.2832	0.1590	0.1590	0.1474	0.0684	0.0684	0.0618,
L Tahmin Edicisi	0.3152	0.3154	0.2966	0.1649	0.1651	0.1529	0.0706	0.0708	0.0639
LS Tahmin Edicisi	0.3182	0.3184	0.3212	0.1691	0.1694	0.1733	0.0741	0.0743	0.0776
M Tahmin Edicisi	0.3262	0.3269	0.3074	0.1692	0.1695	0.1568	0.0717	0.0719	0.0647
MS Tahmin Edicisi	0.3335	0.3344	0.3150	0.1783	0.1788	0.1656	0.0792	0.0794	0.0716
O Tahmin Edicisi	0.3103	0.3191	0.2920	0.1604	0.1611	0.1485	0.0676	0.0666	0.0611
OS Tahmin Edicisi	0.3033	0.3114	0.2938	0.1517	0.1520	0.1498	0.0609	0.0598	0.0611
$n_1 = n_2 = 100$									
Gerçek Hata Oranı	0.3114	0.3115	0.3162	0.1606	0.1606	0.1635	0.0681	0.0681	0.0700
D Tahmin Edicisi	0.3043	0.3043	0.2989	0.1559	0.1559	0.1523	0.0660	0.0660	0.0640
DS Tahmin Edicisi	0.3061	0.3061	0.3007	0.1584	0.1584	0.1547	0.0680	0.0680	0.0659
L Tahmin Edicisi	0.3111	0.3111	0.3056	0.1602	0.1602	0.1565	0.0686	0.0686	0.0665
LS Tahmin Edicisi	0.3120	0.3120	0.3129	0.1614	0.1614	0.1626	0.0696	0.0697	0.0706
M Tahmin Edicisi	0.3138	0.3138	0.3083	0.1614	0.1615	0.1577	0.0690	0.0690	0.0669
MS Tahmin Edicisi	0.3156	0.3157	0.3102	0.1639	0.1640	0.1602	0.0711	0.0711	0.0689
O Tahmin Edicisi	0.3091	0.3119	0.3037	0.1589	0.1592	0.1553	0.0679	0.0676	0.0658
OS Tahmin Edicisi	0.3073	0.3100	0.3043	0.1565	0.1567	0.1558	0.0659	0.0656	0.0660

Tablo 3. $p = 5$ İçin Hata Oranları Değeri.

	$\Delta = 1$			$\Delta = 2$			$\Delta = 3$		
$n_1 = n_2 = 15$		**Jackk.**	**Boots.**		**Jackk.**	**Boots.**		**Jackk.**	**Boots.**
Gerçek Hata Oranı	0.3582	0.3609	0.3850	0.1964	0.1984	0.2269	0.0918	0.0936	0.1167
D Tahmin Edicisi	0.2485	0.2463	0.1920	0.1271	0.1259	0.0935	0.0492	0.0486	0.0336
DS Tahmin Edicisi	0.2726	0.2715	0.2176	0.1544	0.1543	0.1176	0.0696	0.0699	0.0492
L Tahmin Edicisi	0.3031	0.3031	0.2415	0.1708	0.1716	0.1301	0.0770	0.0777	0.0544
LS Tahmin Edicisi	0.3131	0.3137	0.3231	0.1845	0.1860	0.1983	0.0892	0.0905	0.1014
M Tahmin Edicisi	0.3635	0.3683	0.2880	0.2020	0.2052	0.1538	0.0907	0.0924	0.0633
MS Tahmin Edicisi	0.3972	0.4044	0.3228	0.2392	0.2436	0.1875	0.1236	0.1258	0.0886
O Tahmin Edicisi	0.3000	0.3228	0.2339	0.1585	0.1605	0.1186	0.0663	0.0634	0.0460
OS Tahmin Edicisi	0.2699	0.2853	0.2385	0.1281	0.1265	0.1195	0.0453	0.0418	0.0440
$n_1 = n_2 = 30$									
Gerçek Hata Oranı	0.3356	0.3364	0.3555	0.1760	0.1765	0.1923	0.0777	0.0782	0.0894
D Tahmin Edicisi	0.2803	0.2797	0.2497	0.1406	0.1403	0.1223	0.0590	0.0589	0.0492
DS Tahmin Edicisi	0.2906	0.2903	0.2608	0.1534	0.1534	0.1345	0.0691	0.0691	0.0581
L Tahmin Edicisi	0.3110	0.3112	0.2785	0.1623	0.1625	0.1422	0.0727	0.0728	0.0612
LS Tahmin Edicisi	0.3156	0.3159	0.3202	0.1687	0.1690	0.1751	0.0782	0.0784	0.0836
M Tahmin Edicisi	0.3393	0.3406	0.3035	0.1763	0.1770	0.1547	0.0797	0.0801	0.0670
MS Tahmin Edicisi	0.3520	0.3536	0.3169	0.1912	0.1921	0.1690	0.0922	0.0927	0.0782
O Tahmin Edicisi	0.3087	0.3245	0.2755	0.1574	0.1593	0.1376	0.0689	0.0676	0.0579
OS Tahmin Edicisi	0.2968	0.3108	0.2777	0.1438	0.1447	0.1388	0.0586	0.0570	0.0576
$n_1 = n_2 = 100$									
Gerçek Hata Oranı	0.3157	0.3158	0.3239	0.1627	0.1628	0.1680	0.0695	0.0696	0.0728
D Tahmin Edicisi	0.3003	0.3002	0.2906	0.1537	0.1537	0.1478	0.0641	0.0641	0.0608
DS Tahmin Edicisi	0.3030	0.3030	0.2938	0.1574	0.1574	0.1515	0.0670	0.0670	0.0637
L Tahmin Edicisi	0.3109	0.3109	0.3010	0.1603	0.1603	0.1543	0.0681	0.0681	0.0647
LS Tahmin Edicisi	0.3122	0.3122	0.3135	0.1622	0.1622	0.1640	0.0696	0.0696	0.0711
M Tahmin Edicisi	0.3183	0.3184	0.3082	0.1641	0.1642	0.1580	0.0700	0.0700	0.0665
MS Tahmin Edicisi	0.3213	0.3215	0.3113	0.1680	0.1681	0.1618	0.0731	0.0732	0.0695
O Tahmin Edicisi	0.3093	0.3149	0.2994	0.1590	0.1599	0.1530	0.0673	0.0669	0.0639
OS Tahmin Edicisi	0.3064	0.3118	0.3000	0.1553	0.1560	0.1536	0.0643	0.0639	0.0640

6. SONUÇ

Tablolardaki sonuçlar incelendiğinde, hem gerçek hata oranı için ve hem de bu hata oranına ilişkin tahmin ediciler için tüm durumlarda genelde jackknife tekniğine göre elde edilen sonuçlar, bootstrap tekniğine göre elde edilen sonuçlara göre gerçek değerlere (analitik ifadelerden elde edilen değerler) daha yakın olduğu görülmektedir. Buradan jackknife yönteminin, bootstrap yöntemine tercih edilebileceği söylenebilir. Jackknife yönteminin bu üstünlüğü örneklem hacmi artıkça daha da belirginleşmektedir.

Bununla birlikte Δ değeri artıkça, jackknife ve bootstrap değerlerinin gerçek değerlere daha yakın olduğu ancak bu durumda da yine jackknife yönteminin bootstrap yöntemine tercih edilebileceği görülmektedir.

Örneklem hacminin ve değişken sayısının küçük olduğu durumlarda, genelde D, DS, L, LS ve M tahmin edicilerinin jackknife değerlendirmesi daha iyi sonuç vermektedir. Ancak örneklem hacmi artıkça diğer tahmin edicilerin de jackknife değerleri gerçek değerlerine daha yakın olduğu görülmektedir.

Bu sonuçlardan gerçek hata oranı tahmin edicileri için jackknife tekniğinin uygulanabileceği söylenebilir. Bununla birlikte hata oranı tahmin edicilerinin jackknife ve bootstrap değerleri analitik olarak da elde edilebilir. Bu durum başka bir çalışma konusu olarak düşünülebilir.

KAYNAKÇA

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis.* Second Edition, New York, John Wiley & Sonsc Inc.

Efron, B. (1979). Bootstrap Methods: Another look at the Jackknife. *Annals of Stat.* 7, 1-26.

Efron, B. (1982). *The Jackknife, the Bootstrap and other resampling Plans.* J. W. Arrowsmith ltd., Bristol, England.

Efron, B. ve Gong. (1983). A Leisurely look at the bootstrap, the Jackknife and Cross validation. *The American Statistician* 37, 36-48.

Fisher, R.A. (1936). The use of the multiple measurements in taxonomic problems. *Annals of Eugenics* 7, 179-188.

Hills, M. (1966). Allocation rules and their error rates. *Journal of the Royal Statistical Society* Ser. B, 28, 1-20.

Lachenbruch, P.A. (1968). On expected probability of misclassification in discriminant analysis, necessary sample size and a relation with the multiple correlation coefficient. *Biometrics* 24, 823-834.

Lachenbruch, P.A. ve Mickey, M.R. (1968). Estimation of error rates in discriminant analysis. *Technometrics* 10, 1-11.

Lachenbruch, P.A. (1975). *Discriminant Analysis.* Hafner Press., New York.

McLachlan, G.J. (1974). An Asimptotic unbiased technique for estimation the error rates in discriminant analysis. *Biometrics* 30, 239-249.

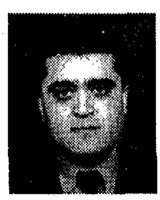
Okamoto, M. (1963). An Asymptotic Expansion for the Distribution of the Linear Discriminant Function. *Annals of Math. Statistics,* 34, 1286-1301.

Seber, G.A.F. (1984). *Multivariate Observations.* John Wiley & Sons. Inc.

Sitgreaves, R. (1961). Some results on the distribution of the W classification statistics. *Item Analysis and Pridiction,* Ed. H. Solomon Stanford, Calif. Stanford University Press. 241-251.

Snapinn, S.M. (1983). An Evaluation of Smoothed Error Rate Estimators in Discriminant Analysis, *Institute of Statistics Mimeo Series* No:1483, University North Carolina at Chapel Hill.

Welch. B.L. (1939). Note on the discriminant functions. *Biometrica* 31, 218-220.



Cemal Atakan, 1965 yılında Sivas Zara'da doğdu. 1986 yılında Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü'nden lisans öğrenimi tamamlayarak istatistikçi unvanını aldı. 1990 yılında Orta Doğu Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı'nda Yüksek Lisans öğrenimini tamamladı. 1997 yılında, Ankara Üniversitesi Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı'nda Doktora öğrenimini tamamlayarak İstatistik Doktoru unvanını aldı. Halen Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü'nde Yardımcı Doçent Dr. olarak görev yapmaktadır. Çok Değişkenli İstatistik (Diskriminant Analizi) ve Uygulamalı İstatistik çalışmalarını yürüttüğü alanlardır. Evli ve bir çocuğu vardır.