Volume: 08 Number: 02 Category: 01 Page: 13-20 ISSN No: 1303-6319 Cilt: 08 Sayı: 02 Kategori: 01 Sayfa: 13-20

AYARLAMANIN DİSKRİMİNANT ANALİZİNDE HATA ORANLARINA ETKİSİ

Hayrinisa DEMİRCİ BİÇER* **Cemal ATAKAN****

ÖZET

Bir çok istatistiki analizde olduğu gibi, Diskriminant Analizi'nde de değişken sayısı p, örneklem hacimleri n_i $(i=1,2,\ldots,g)$ olmak üzere $n_i < p$ olduğunda, örneklem varyans-kovaryans matrisinin tahminleri büyük değerler olmakta ve örneklem varyans-kovaryans matrisi singüler olmaktadır. Bu problemi çözmek için Friedman (1989) Ayarlanmış Diskriminant Analizi'ni (Regularized Discriminant Analysis) önermiştir. Ayarlanmış diskriminant analizinde, örneklem varyans-kovaryans matrisi daraltma parametresi ile tekrar elde edilir. Bu çalışmada, $n_i < p$ olduğu durumda, simülasyon çalışması ile hata oranları bakımından Ayarlanmış Diskriminant Analizi ve Lineer Diskriminant Analizi karşılaştırılmıştır.

Anahtar Kelimeler: Ayarlanmış diskriminant analizi, Diskriminant analizi, Hata oranı.

1. GİRİŞ

Diskriminant analizi, üzerinden ölçüm alınan bir birimin sonlu sayıda bilinen farklı kitlelerden birine atanmasını gerçekleştiren istatistiksel bir teknik olarak tanımlanır. Bu atama işlemi yapılırken birimin aldığı gözlem değerine göre ait olduğu kitleden farklı bir kitleye atandığında, bir hata yapılmış olur. Diskriminant analizinde bu hataya, hata oranı ya da hatalı sınıflandırma olasılığı denmektedir. Diskriminant analizinde amaç, atama işlemini minumum hatayla yapmaktır (Atakan, 2003).

$\Pi_1,\ldots,\Pi_g$ birbirinden farklı g tane kitle olmak üzere, $\underline{X}=\left(X_1,X_2,\cdots,X_p\right)'$ birey üzerinde ölçümlere karşılık gelen p boyutlu rasgele vektörü Π_i $(i=1,\ldots,g)$ kitlesinden ise $\underline{X}$'in ortak olasılık yoğunluk fonksiyonu $f_i\left(\underline{x};\underline{\theta}_i\right)$ biçiminde gösterilir. Burada, $\underline{\theta}_i$ parametre vektörüdür.

$\underline{X}$'in aldığı değerler p boyutlu R^p örneklem uzayında olmak üzere, sınıflandırma işlemi bu uzayı $B_1\cup\cdots\cup B_g=R^p$ ve $B_i\cap B_j=\varnothing$ $i\neq j$ $i,j=1,\ldots,g$ olan g tane bölgeye ayırır. Eğer $\underline{X}$'in gözlem değeri B_k bölgesinde ise, bu gözlemin yapıldığı birim Π_k kitlesine atanır, $(k=1,\ldots,g)$.

*Ar. Gör., Ankara Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Bölümü, 06100, Ankara, e-posta: hdbicer@hotmail.com
**Doç. Dr, Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06100, Ankara, e-posta: Cemal.Atakan@science.ankara.edu.tr

$$\delta_i f_i\left(\underline{x};\underline{\theta}_i\right)=\max_{1\le k\le g}\delta_k f_k\left(\underline{x};\underline{\theta}_k\right) \tag{1}$$

olduğunda $\underline{x}$ gözlemi $i.$ sınıfa atanır. Burada, δ_i birimin $i.$ kitleye ait olması olasılığı (önsel olasılık)dır. $\underline{X}_i$ rasgele vektörü $\underline{\mu}_i$ ortalama vektörü ve Σ varyans-kovaryans matrisine sahip p değişkenli

$$f_i\left(\underline{x}\right)=\frac{1}{\left(2\pi\right)^{p/2}\left|\Sigma\right|^{1/2}}e^{-\frac{1}{2}\left(\underline{x}-\underline{\mu}_i\right)'\Sigma^{-1}\left(\underline{x}-\underline{\mu}_i\right)}$$

normal dağılıma sahip olsun $\left(i=1,\ldots,g\right)$. Bu durumda, (1) eşitliği ile verilen optimal sınıflandırma fonksiyonu

$$D_i\left(\underline{x}\right)=\left(\underline{x}-\underline{\mu}_i\right)'\Sigma^{-1}\left(\underline{x}-\underline{\mu}_i\right)+\ln\left|\Sigma\right|-2\ln\delta_i \tag{2}$$

ile verilir. $\underline{\mu}_i$ ve Σ parametreleri genelde bilinmediği için örneklemden elde edilen tahmin edicileri kullanılır. Bu parametrelerin örneklemden elde edilen tahmin değerleri sırası ile $\underline{\overline{x}}_i$ ortalama vektörü ve S birleştirilmiş örneklem varyans-kovaryans matrisi olmak üzere örneklemden elde edilen optimal sınıflandırma fonksiyonu

$$\hat{D}_i\left(\underline{x}\right)=\left(\underline{x}-\underline{\overline{x}}_i\right)'S^{-1}\left(\underline{x}-\underline{\overline{x}}_i\right)+\ln\left|S\right|-2\ln\hat{\delta}_i \tag{3}$$

olarak elde edilir. Bu ifadeye Fisher'in örneklem lineer diskriminant fonksiyonu denir. (Mkhadri, 1995). $\hat{\delta}_i$, birimin $i.$ kitleye ait olması olasılığının tahmin edicisidir.

Değişken sayısı p, örneklem hacimleri n_i'den $\left(i=1,2,\ldots,g\right)$ büyük olduğunda, örneklem varyans-kovaryans matrisinin elemanlarının tahmin değerleri büyük değerler olmakta ve örneklem varyans-kovaryans matrisleri singüler olmaktadır. Di Pillo (1976), Campbell (1980) ve Peck ve Van Ness (1982) birleştirilmiş örneklem varyans-kovaryans matrisinin yansız tahmin edicisi S'nin yerine, S'nin özdeğerlerinin daraltılması ile elde edilen yanlı bir tahmin ediciyi kullanmayı önermişlerdir. Bu durumda birleştirilmiş örneklem varyans-kovaryans matrisi S'nin elemanlarının tahmin değerlerini daha küçük değerler olarak elde etmişlerdir. Mkhadri (1995) lineer diskriminant fonksiyonunda birleştirilmiş örneklem varyans-kovaryans matrisi S'nin yerine $\left(S+\gamma I\right)$ alındığında lineer diskriminant fonksiyonunun varyansının daha küçük olduğunu göstermiştir. Burada $\gamma\in\left(0,1\right)$ dir. Friedman (1989) bu süreci Ayarlanmış Diskriminant Analizi (Regularization Discriminant Analysis) olarak adlandırmıştır.

Bu çalışmada, $n_i\le p$ olduğunda, Friedman'ın (1982) önerdiği lineer diskriminant fonksiyonu ile hata oranları bakımından Ayarlanmış Diskriminant Analizi ve Lineer Diskriminant Analizi bir simülasyon çalışması ile karşılaştırılmıştır.

2. İKİ KİTLE İÇİN DİSKRİMİNASYON PROBLEMİ

İşlem kolaylığı açısından $g=2$ olarak alınsın. Bu durumda Π_1 ve Π_2, çok değişkenli normal dağılımlı kitleler olsun. Parametreler bilinmediğinde (3) eşitliği ile verilen örneklem diskriminant fonksiyonu

$$W\left(\underline{x}\right)=\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)' S^{-1}\left[\underline{x}-\frac{1}{2}\left(\underline{\bar{x}}_1+\underline{\bar{x}}_2\right)\right] \tag{4}$$

biçiminde tanımlanır. Burada $n=n_1+n_2-2$ olmak üzere $\underline{\bar{x}}_i=\sum_{j=1}^{n_i}\underline{x}_j \Big/ n_i$ ve $nS=\sum_{i=1}^{n_1}\left(\underline{x}_i-\underline{\bar{x}}_1\right)\left(\underline{x}_i-\underline{\bar{x}}_1\right)'+\sum_{i=1}^{n_2}\left(\underline{x}_i-\underline{\bar{x}}_2\right)\left(\underline{x}_i-\underline{\bar{x}}_2\right)'$ yansız örneklem tahminleridir. (4) eşitliğinden atama kuralı

$$\Lambda=\begin{cases} W\left(\underline{x}\right)\le k \text{ ise } \underline{x},\ \Pi_1\text{'e atanır} \\ W\left(\underline{x}\right)> k \text{ ise } \underline{x},\ \Pi_2\text{'ye atanır}\end{cases} \tag{5}$$

biçimindedir. Burada k bir sabittir. Maliyetler ihmal edilip, önsel olasılıklar eşit alınır ise $k=0$ olacaktır (Anderson, 1984 ve Lachenbruch, 1975).

Varyans-kovaryans matrisinin singüler olmaması için Di Pillo (1976) ve Campbell (1980) $W\left(\underline{x}\right)$ lineer diskriminant fonksiyonundaki birleştirilmiş varyans-kovaryans matrisine bir daraltma parametresi uygulayarak aynı fonksiyonu

$$\overline{W}\left(\underline{x}\right)=\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)' S^{-1}\left(\gamma\right)\left[\underline{x}-\frac{1}{2}\left(\underline{\bar{x}}_1+\underline{\bar{x}}_2\right)\right] \tag{6}$$

biçiminde elde etmişlerdir. Burada $S^{-1}\left(\gamma\right)=\left(S+\gamma I\right)^{-1}$, $\gamma\in\left(0,1\right)$ ve I birim matristir. Mkhadri (1995) iki grup olduğunda $\overline{W}\left(\underline{x}\right)$ fonksiyonunun varyansının $W\left(\underline{x}\right)$ fonksiyonunun varyansından daha küçük olduğunu göstermiştir.

Peck ve Van Ness (1982) ise (4) eşitliği ile verilen $W\left(\underline{x}\right)$'de S^{-1} yerine

$$S^{*-1}\left(\gamma\right)=\left(1-\gamma\right)aS^{-1}+\left[\frac{b\gamma}{tr\left(S\right)}\right]I \tag{7}$$

önermişlerdir. Burada $a=\left(n-p-3\right)/n$ olarak seçildiğinde aS^{-1}, Σ^{-1}'in yansız tahmini olmaktadır ve b pozitif bir sabittir. Eğer $\Sigma=\sigma^2 I$ olarak seçilir ise, $\left[b\gamma/tr\left(S\right)\right]I$ Σ^{-1}'in doğal bir tahmin edicisi olmaktadır. Σ^{-1}'in herhangi bir yansız tahmini için $b=p\left(2n-2\right)-2$ olarak alınabilir. Mkhadri (1995) daraltma parametresinin en iyi seçimi için $b=p/n$ 'yi almayı önermiştir.

Bu durumda sınıflandırma fonksiyonu

$$W^*\left(\underline{x}\right)=\left(\underline{\overline{x}}_1-\underline{\overline{x}}_2\right)' S^{*-1}\left(\gamma\right)\left[\underline{x}-\frac{1}{2}\left(\underline{\overline{x}}_1+\underline{\overline{x}}_2\right)\right] \tag{8}$$

olarak elde edilir.

Friedman (1989) Ayarlanmış Diskriminant Analizi (RDA) olarak adlandırdığı süreçte diskriminant fonksiyonunda grup varyans-kovaryans matrisinin tahmini S_i, $\left(i=1,\ldots,g\right)$, yerine

$$S_i\left(\gamma\right)=\frac{\left(1-\gamma\right)S_i+\gamma S}{\left(1-\gamma\right)n_i+\gamma n} \tag{9}$$

alınmasını önermiştir. Burada $\gamma\in\left(0,1\right)$ dir.

3. DARALTMA PARAMETRESİNİN SEÇİMİ

(8) eşitliği ile verilen diskriminant fonksiyonu γ daraltma parametresine bağlıdır. γ daraltma parametresinin optimal değeri olarak hatalı sınıflandırma olasılığının tahminini minimize edecek şekilde seçilebilir (Lachenbruch, 1975).

$\varepsilon\left(\underline{x}\right)=\left(\underline{\overline{x}}_1-\underline{\overline{x}}_2\right)'\left[\underline{x}-\left(\underline{\overline{x}}_1+\underline{\overline{x}}_2\right)/2\right]$ olmak üzere (8) eşitliğinde $W\left(\underline{x}\right)$ fonksiyonu yerine

$$W^*\left(\underline{x},\gamma\right)=\left(1-\gamma\right)aW\left(\underline{x}\right)+\left[\frac{b\gamma}{tr\left(S\right)}\right]\varepsilon\left(\underline{x}\right) \tag{11}$$

fonksiyonu alınırsa optimal daraltma parametresi

$$\gamma_{opt}=\frac{aW^{(j)}\left(x_j\right)}{aW^{(j)}\left(x_j\right)+\left(b/tr\left(S_{/j}\right)\right)\varepsilon^{(j)}\left(x_j\right)} \quad , \quad \forall x_j,\ 1\le j\le n \tag{12}$$

olarak elde edilir. Burada $W^{(j)}\left(x_j\right)$, $\varepsilon^{(j)}\left(x_j\right)$ ve $S_{/j}$ örneklemden x_j gözlemi çıkarıldığında geriye kalan n_i-1 gözlemden elde edilen ifadelerdir. $\gamma_{opt}<0$ ise $\gamma_{opt}=0$ ve $\gamma_{opt}>1$ ise $\gamma_{opt}=1$ olarak alınır (Mkhadri, 1995).

4. SİMÜLASYON ÇALIŞMASI

$$n_1 = n_2 = 20\,,\quad \underline{\mu}_1\,,\quad \underline{\mu}_2\,,\quad \Sigma = \Sigma_1 = \Sigma_2 = I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad \text{ve} \quad \Delta^2 = 1,2,3,4,5 \quad \text{olacak}$$

şekilde çok değişkenli normal dağılımdan sayı üretilmiştir. $\Delta^2 = \left(\underline{\mu}_1 - \underline{\mu}_2\right)' \Sigma^{-1} \left(\underline{\mu}_1 - \underline{\mu}_2\right)$ Mahalanobis uzaklığının değişik değerleri için, Matlab paket programı ile elde edilen 1000 tekrarlı simülasyon sonuçları Tablo 1 – Tablo 5 ile verilmiştir. Bununla birlikte, değişken sayısına göre hata oranlarının değişimi Şekil 1 – Şekil 6'da verildiği gibi elde edilmiştir.

Tablo 1. $\Delta^2 = \left(\underline{\mu}_1 - \underline{\mu}_2\right)' \Sigma^{-1} \left(\underline{\mu}_1 - \underline{\mu}_2\right) = 1$ **olduğunda iki kitle için lineer diskriminant analizi ve ayarlanmış diskriminant analizinde hata oranları**

	$p=5$	$p=10$	$p=15$	$p=20$	$p=30$	$p=40$
LDA	0.3183	0.2818	0.2449	0.2058	0.1436	0.4828
RDA	0.2465	0.1559	0.1247	0.1213	0.1235	0.1295
$\hat{\gamma}$	0.1304	0.0696	0.0527	0.0853	0.0589	0.0750

Tablo 2. $\Delta^2 = \left(\underline{\mu}_1 - \underline{\mu}_2\right)' \Sigma^{-1} \left(\underline{\mu}_1 - \underline{\mu}_2\right) = 2$ **olduğunda iki kitle için lineer diskriminant analizi ve ayarlanmış diskriminant analizinde hata oranları**

	$p=5$	$p=10$	$p=15$	$p=20$	$p=30$	$p=40$
LDA	0.2794	0.2450	0.2125	0.1865	0.1354	0.4880
RDA	0.2219	0.1489	0.1234	0.1224	0.1238	0.1283
$\hat{\gamma}$	0.1122	0.0713	0.0606	0.0658	0.0468	0.0731

Tablo 3. $\Delta^2 = \left(\underline{\mu}_1 - \underline{\mu}_2\right)' \Sigma^{-1} \left(\underline{\mu}_1 - \underline{\mu}_2\right) = 3$ **olduğunda iki kitle için lineer diskriminant analizi ve ayarlanmış diskriminant analizinde hata oranları**

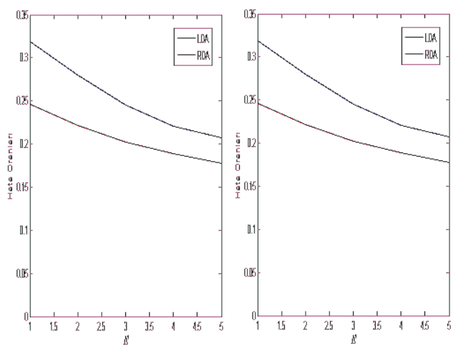
	$p=5$	$p=10$	$p=15$	$p=20$	$p=30$	$p=40$
LDA	0.2447	0.2213	0.1927	0.1669	0.1325	0.4880
RDA	0.2025	0.1449	0.1231	0.1222	0.1243	0.1283
$\hat{\gamma}$	0.1167	0.0636	0.0598	0.0728	0.0314	0.0823

Tablo 4. $\Delta^2 = \left(\underline{\mu}_1 - \underline{\mu}_2\right)' \Sigma^{-1} \left(\underline{\mu}_1 - \underline{\mu}_2\right) = 4$ **olduğunda iki kitle için lineer diskriminant analizi ve ayarlanmış diskriminant analizinde hata oranları**

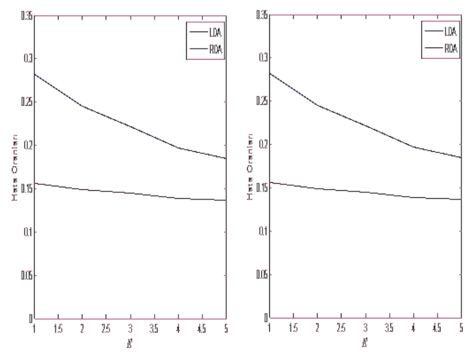
	$p=5$	$p=10$	$p=15$	$p=20$	$p=30$	$p=40$
LDA	0.2208	0.1975	0.1765	0.1565	0.1289	0.4993
RDA	0.1891	0.1390	0.1234	0.1222	0.1244	0.1288
$\hat{\gamma}$	0.1082	0.0619	0.0543	0.0798	0.0300	0.0758

Tablo 5. $\Delta^2 = \left(\underline{\mu}_1 - \underline{\mu}_2\right)' \Sigma^{-1} \left(\underline{\mu}_1 - \underline{\mu}_2\right) = 5$ **olduğunda iki kitle için lineer diskriminant analizi ve ayarlanmış diskriminant analizinde hata oranları**

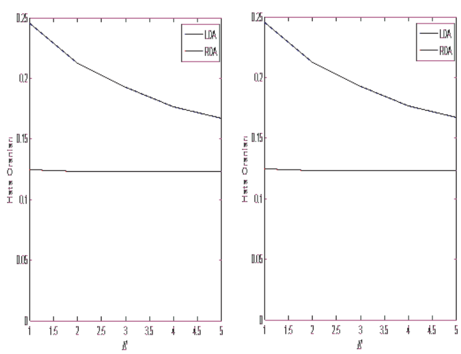
	$p=5$	$p=10$	$p=15$	$p=20$	$p=30$	$p=40$
LDA	0.2074	0.1843	0.1668	0.1498	0.1272	0.5000
RDA	0.1777	0.1371	0.1231	0.1229	0.1246	0.1305
$\hat{\gamma}$	0.1024	0.0715	0.0539	0.0600	0.0223	0.0774



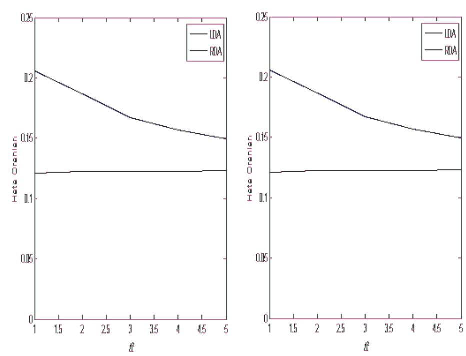
Şekil 1. $p=5$ **için hata oranları**



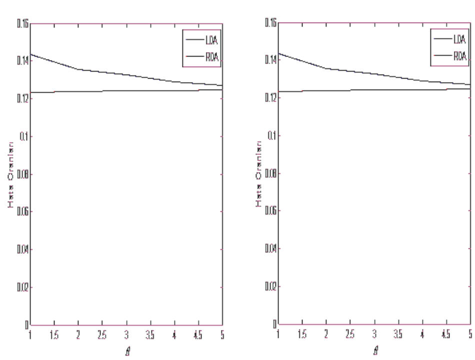
Şekil 2. $p=10$ **için hata oranları**



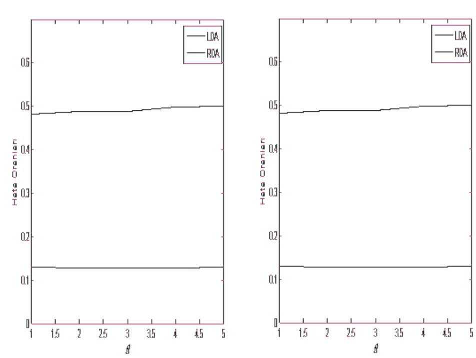
Şekil 3. $p=15$ **için hata oranları**



Şekil 4. $p=20$ **için hata oranları**



Şekil 5. $p=30$ **için hata oranları**



Şekil 6. $p=40$ **için hata oranları**

Mahalanobis uzaklığı arttıkça LDA ve RDA dan elde edilen hata oranlarının azaldığı görülmektedir.

5. TARTIŞMA VE SONUÇ

Küçük hacimli örneklemlerde değişken sayısının fazla olduğu durumda örneklem varyans-kovaryans matrisinin tahmin değerleri olması gereken değerlerden daha büyük değerler olarak elde edilmektedir. Bu problemi çözmek için Friedman (1989) Ayarlanmış Diskriminant Analizi'ni önermiştir.

Bu çalışmada, değişken sayısının ve Mahalanobis uzaklığının farklı değerleri için Lineer Diskriminant Analizi ve Ayarlanmış Diskriminant Analizi hata oranları bakımından karşılaştırılmıştır. Yapılan simülasyon çalışmalarından Ayarlanmış Diskriminant Analizi'nden elde edilen hata oranlarının Lineer Diskriminant Analizi'nden elde edilen hata oranlarına göre daha küçük olduğu gözlemlenmiştir. Bununla birlikte, $p = 40$ olması durumları için hata oranlarına göre Ayarlanmış Diskriminant Analizi'nin Lineer Diskriminant Analizi'nden daha iyi sonuçlar verdiği gözlemlenmiştir.

Benzer olarak $g > 2$ olduğu durumlar içinde hata oranlarını ve $\hat{\gamma}$ değerlerini elde etmek mümkündür.

6. KAYNAKLAR

Anderson, T. W., 1984. An Introduction to Multivariate Satistical Analysis. Second Edition. John Wiley& Sons. Inc., New York.

Atakan, C., 2003. Diskriminant Analizinde Gerçek Hata Oranına İlişkin Güven Aralıkları için Simülasyon Çalışması. Selçuk Üniversitesi Fen-Edebiyat Fakültesi Fen Dergisi, Sayı: 22, 93–100.

Campbell, N. A., 1980. Shrunken Estimator in Discriminant and Canonical Variate Analysis. Applied Statistics, 29, 5–14.

Di Pillo, P. J., 1976. The Application of Bias to Discriminant Analysis. Comm. Stat. Theory Method, A5, 843–854.

Friedman, J. H., 1989. Regularized Discriminant Analysis. JASA, 84, 165–175.

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

Mkhadri, A., 1995. Shrinkage Parameter for Modified Linear Discriminant Analysis. Pattern Recognition Letters, 16, 267–275.

Peck, R., Van Ness, J., 1982. The Use of Shrinkage Estimators in Linear Discriminant Analysis. IEEE Trans. Patt. Anal. Mach. Intel., 5, 530–537.

EFFECT OF REGULARIZATION ON ERROR RATES IN DISCRIMINANT ANALYSIS

ABSTRACT

As in many statistical analyses, when the number of variables (p) is higher than the number of objects in groups (n_i) *, the sample variance- covariance matrix estimates become high and the sample variance-covariance matrix becomes singular in discriminant anaylsis as well. To solve this problem, Friedman (1989) proposed Regularized Discriminant Analysis. The sample variance-covariance matrix is re-obtained by employing a shrinkage parameter in regularized discriminant analysis. In this paper Linear and Regularized Discriminant Analysis are compared with respect to their error rates with a simulation study when the number of variables is bigger than the sample size.*

Keywords: Regularized discriminant analysis, Discriminant analysis, Error rate.