ISSN:1306-3111
e-Journal of New World Sciences Academy
2011, Volume: 6, Number: 4, Article Number: 3A0043

PHYSICAL SCIENCES
Received: August 2011
Accepted: October 2011
Series : 3A
ISSN : 1308-7304


Hayrinisa Demirci Biçer
Cemal Atakan
Cenker Biçer
Ankara University
hdbicer@hotmail.com
Ankara-Turkey

## İKİ PARAMETRELİ WEİBULL DAĞILIMINA SAHİP KİTLELERDE DİSKRİMİNANT ANALİZİ

**ÖZET**

Weibull dağılımı, bir çok alanda istatistiksel sonuç çıkarımı yapmak için kullanılan çarpık bir dağılımdır. Bu çalışmada, tek değişkenli iki parametreli Weibull dağılımına sahip kitlelerde sınıflandırma problemi, kitle parametrelerinin bilinip bilinmemesi durumları göz önüne alınarak incelenmiştir. Hata oranları önerilen sınıflandırma kuralları kullanılarak yeniden yerine koyma (resubstitution) tahmin edicisi ile elde edilmiştir.

**Anahtar Kelimeler:** Weibull Dağılımı, Sınıflandırma Kuralı, Diskriminant Analizi, Hata Oranı, Yeniden Yerine Koyma Tahmin Edicisi

## DISCRIMINANT ANALYSIS IN TWO PARAMETER WEIBULL DISTRIBUTION POPULATIONS

**ABSTRACT**

Weibull distribution is skewed distribution which is used to make statistical inferences in many areas. In this study, the classification problem in univariate two parameter Weibull populations is investigated which population parameters are considered known and unknown situation. Error rates are obtained by using the proposed classification rules with the resubstitution estimator.

**Keywords:** Weibull Distribution, Classification Rule, Discriminant Analysis, Error Rate, Resubstitution Estimator

## 1. GİRİŞ (INTRODUCTION)

Diskriminant analizi, araştırmacının bir birim üzerinde ölçümler yaptığında, bu ölçümlere bağlı olarak birimi bilinen gruplardan (kitlelerden) birine atanmasını gerçekleştiren istatistiksel bir yöntemdir. Bu çalışmada, işlem kolaylığı için grup sayısı iki olarak alınmıştır.

$\Pi_1$, $\Pi_2$ iki farklı kitle ve $\underline{X} = \left(X_1, X_2, \cdots, X_p\right)'$ bireyler üzerinde ölçümlere karşılık gelen $p$ boyutlu rasgele vektör olsun. $\underline{X}$, $\Pi_j$ $(j=1,2)$'den bir rasgele vektör ise $\underline{X}$'in olasılık yoğunluk fonksiyonu $f_j\left(\underline{x};\underline{\theta}_j\right)$dir. Burada $\underline{\theta}_j$ parametre vektörüdür.

$\underline{X}$ rasgele vektörünün aldığı değerler, $p$ boyutlu $R^p$ örneklem uzayında olmak üzere sınıflandırma problemi, bu uzayı $B_1$ ve $B_2$ bölgelerine ayırmaktadır; öyle ki, $\underline{X}_0$ gözlem vektörü $B_j$ de ise $\underline{X}_0$'a ait $\underline{x}_0$ birimi $\Pi_j$'ye atanır. Burada $B_1 \cup B_2 = R^p$ ve $B_1 \cap B_2 = \varnothing$ dir.

$q_1$ ve $q_2$ bir birimin sırasıyla $\Pi_1$ ve $\Pi_2$ kitlelerine ait olması olasılıkları (prior olasılıkları) ve $\Pi_j$'ye ait bir birimin $\Pi_i$'denmiş gibi hatalı atanmasının maliyeti, $c_{i|j} = C\left(\underline{x}_0\text{'ın } \Pi_i\text{'ye atanması} \middle| \underline{x}_0 \in \Pi_j\right)$ olsun ($i=1,2$, $j=1,2$, $i \neq j$). Bu çalışmada hatalı sınıflandırma maliyetleri eşit alınarak ihmal edilmiştir ve prior olaslıkları $q_1$ ve $q_2$'nin de eşit olduğu varsayılacaktır.

Atama işlemi yapılırken, gözlem ait olduğu gerçek gruptan farklı bir gruba atanabilir. Bu durumda hatalı atama yapılmış olur. Yani gözlem gerçekte $\Pi_j$'ye aittir ancak $\Pi_i$'ye atanmış olabilir. Belirlenen $B_1$ ve $B_2$ bölgeleri için,

$$P\left(\underline{x}_0\text{'ın } \Pi_i\text{'ye atanması} \middle| \underline{x}_0 \in \Pi_j\right) = \int_{R_i} f_j\left(\underline{x},\underline{\theta}_j\right) d\underline{x} = p_{i|j} \tag{1.1}$$

olasılığı, $\underline{x}_0$ ölçümlü birimin $\Pi_j$'ye ait olduğu bilindiğinde $\Pi_i$'denmiş gibi hatalı atanmasının koşullu olasılığıdır. Bu durumda $\underline{x}_0$ gözleminin $\Pi_i$'ye hatalı atanması olasılığı,

$$P\left(\underline{x}_0\text{'ın } \Pi_i\text{'ye hatalı atanması}\right) = q_j\, p_{i|j} \tag{1.2}$$

dir. Böylece iki grup olduğunda herhangi bir atama kuralı için hatalı sınıflandırmanın beklenen maliyeti $(ECM)$

$$ECM = \sum_{\substack{i=1 \\ i \neq j}}^{2} q_j\, c_{i|j}\, p_{i|j}\,, \quad (j=1,2) \tag{1.3}$$

biçiminde tanımlanır. Diskriminant analizinde atama kurallarından biri, öyle $B_1$ ve $B_2$ bölgeleri seçilimelidir ki ECM minimum olsun.

Diğer bir atama kuralı, ECM kuralında hatalı sınıflandırma maliyetlerinin ihmal edilmesiyle elde edilen ve en iyi sınıflandırma kuralı olan toplam hatalı sınıflandırma olasılığının (TPM) minimizasyonu kuralıdır. Bu kural

$$P(TPM) = \sum_{i=1}^{2} q_j\, p_{i|j} \tag{1.4}$$

biçimindedir [1 ve 2].

$\underline{\theta}_1$ ve $\underline{\theta}_2$ bilindiğinde, $\xi$ optimal sınıflandırma kuralı

$$\xi:\begin{cases} U=\dfrac{f(\underline{x}_0;\underline{\theta}_1)}{f(\underline{x}_0;\underline{\theta}_2)}>k & \text{ise } \underline{x}_0 \quad \Pi_1\text{'e atanır} \\ \text{diğer durumda} & \text{ise } \underline{x}_0 \quad \Pi_2\text{'ye atanır} \end{cases} \quad (1.5)$$

ile verilir ve $k=\dfrac{q_2}{q_1}\dfrac{c_{1|2}}{c_{2|1}}$ dir [3 ve 4].

$\underline{\theta}_1$ ve $\underline{\theta}_2$ parametreleri bilinmediğinde, $\Pi_1$ ve $\Pi_2$ kitlelerinden alınan $n_1$ ve $n_2$ hacimli rasgele örneklemler yardımıyla $\hat{\underline{\theta}}_1$ ve $\hat{\underline{\theta}}_2$ tahminleri elde edilir. Böylece örneklemlerden elde edilen tahminlere bağlı $\hat{\xi}$ optimal sınıflandırma kuralı

$$\hat{\xi}:\begin{cases} W=\dfrac{f\left(\underline{x}_0;\hat{\underline{\theta}}_1\right)}{f\left(\underline{x}_0;\hat{\underline{\theta}}_2\right)}>k & \text{ise } \underline{x}_0 \quad \Pi_1\text{'e atanır} \\ \text{diğer durumda} & \text{ise } \underline{x}_0 \quad \Pi_2\text{'ye atanır} \end{cases} \quad (1.6)$$

ile verilir [3 ve 5].

## 2. ÇALIŞMANIN ÖNEMİ (RESEARCH SIGNIFICANCE)

Literatürde genelde tek ve çok değişkenli normal dağılımlı kitlelerde sınıflandırma problemi ve hata oranları ayrıntılı olarak verilmektedir. Örneklemlerin alındığı kitlelerin dağılımları her zaman normal olamayacağından, normal dağılıma sahip olmayan kitleler arasında da sınıflandırma problemi ele alınmalıdır. Bu nedenle, Weibull dağılımına sahip kitleler arasındaki sınıflandırma problemi ele alınarak, farklı örneklem büyüklükleri ve farklı parametre değerleri için yeniden yerine koyma tahmin edicisi ile hata oranları elde edilmiştir.

## 3. DENEYSEL YÖNTEM (EXPERIMENTAL METHOD)

### 3.1. Diskriminant Analizinde Hata Oranları (Error Rates in Discriminant Analysis)

Diskriminant kurallarına ilişkin optimal hata oranı, koşullu gerçek hata oranı, beklenen gerçek hata oranı gibi farklı hata oranları tanımlanmıştır. Parametrelerin bilindiği durumda (1.5)'deki $\xi$ kuralı kullanıldığında $\Pi_i$'den bir birimin hatalı atanması olasılığı $\alpha_i(\xi)$, $(i=1,2)$ olarak gösterilsin.

$$\alpha_1(\xi)=P(U\le k|\underline{X}\in\Pi_1) \quad (3.1)$$

ve

$$\alpha_2(\xi)=P(U>k|\underline{X}\in\Pi_2) \quad (3.2)$$

olmak üzere bu iki hata oranı optimal hata oranları olarak isimlendirilmektedir.

$\Pi_1$ ve $\Pi_2$ kitleleri gelişigüzel seçildiğinden, sadece $\alpha_1(\xi)$'nin düşünülmesi yeterlidir. $\alpha_2(\xi)$ değeri $\alpha_1(\xi)$'ye benzer biçimde elde edilebilir. Böylece, bilinmeyen bir $\underline{X}_0$ gözleminin $\Pi_1$'den geldiği bilindiğinde, $\Pi_2$'ye hatalı sınıflandırılması olasılığı $\alpha$'nın indisi göz önüne alınmayıp,

$$\alpha(\xi)=P(U\le k) \quad (3.3)$$

ile verilir.

Koşullu gerçek hata oranı (1.6)'daki $\hat{\xi}$ kuralı kullanıldığında $\Pi_1$'den rasgele bir gözlemin hatalı atanması olasılığı olan,

$$\alpha\left(\hat{\xi}\right)=P(W\le k) \quad (3.4)$$

değeridir. Bu hata oranı örneklemlerden elde edilen parametre tahminleri üzerine koşullandırılmıştır. $\alpha\left(\hat{\xi}\right)$ koşullu gerçek hata oranına kısaca koşullu hata oranı da denmektedir.

Beklenen gerçek hata oranı olarak tanımlanan hata oranı, (3.4)'deki koşullu gerçek hata oranının tüm örneklemler üzerinden,

$$E\left(\alpha\left(\hat{\xi}\right)\right) = E(P(W \le k)) \tag{3.5}$$

beklenen değeridir. $E\left(\alpha\left(\hat{\xi}\right)\right)$ beklenen gerçek hata oranına koşulsuz hata oranı da denmektedir [2, 3, 5, 6, 7 ve 8].

Parametreler bilinmediğinde, örneklemlerden elde edilen tahminler kullanılarak oluşturulan örneklem diskriminant fonksiyonunun dağılımından yararlanarak hata oranları bulunabilir. Ancak örneklem diskriminant fonksiyonunun dağılımını elde etmek her zaman kolay değildir. Bu nedenle örneklem diskriminant fonksiyonunun dağılımı bilinmediğinde ya da analitik olarak elde edilemediğinde hata oranları için birçok tahmin edici söz konusudur.

Bu çalışmada, parametrik olmayan ve Smith (1947) tarafından önerilen yeniden yerine koyma (resubstitution) tahmin edicisi veya görünüşte (apparent) hata oranı olarak isimlendirilen R tahmin edicisi ile hata oranı tahmin edilecektir ve $\hat{\alpha}^R$ ile gösterilecektir. $\hat{\alpha}^R$ tahmin edicisi, $\Pi_1$'den $n_1$ gözlem ve $\Pi_2$'den $n_2$ gözlem alınarak oluşturulan diskriminant fonksiyonu kullanılarak $n_1+n_2$ gözlemin tekrar sınıflandırılmasıyla, $\Pi_1$'den $\Pi_2$'ye atananların sayısının $n_1$'e ve $\Pi_2$'den $\Pi_1$'e atananların sayısının $n_2$'ye oranlanması ile elde edilir [5 ve 9].

## 4. WEIBULL DAĞILIMINA SAHİP KİTLELERDE DİSKRİMİNANT ANALİZİ (DISCRIMINANT ANALYSIS IN WEIBULL POPULATIONS)

$\Pi_1$ ve $\Pi_2$, $\mu_1$ ve $\mu_2$ $(\mu_1 < \mu_2)$ konum parametreli iki tek değişkenli kitle olmak üzere $x_0$ ölçümlü birim için $R_N$ kuralı,

$$R_N : \begin{cases} x_0 < \frac{1}{2}(\mu_1+\mu_2) & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \text{diğer durumda} & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \tag{4.1}$$

ile verilir. Parametrelerin bilinmediği durumda ise $\Pi_1$ ve $\Pi_2$'den alınan $n_1$ ve $n_2$ hacimli örneklemlerden elde edilen örneklem atama kuralı $R_N(s)$,

$$R_N(s) : \begin{cases} \bar{x}_1 < \bar{x}_2 \ \text{ve} \ \underline{x}_0 < \frac{1}{2}(\bar{x}_1+\bar{x}_2) & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \bar{x}_1 < \bar{x}_2 \ \text{ve} \ \underline{x}_0 \ge \frac{1}{2}(\bar{x}_1+\bar{x}_2) & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \tag{4.2}$$

ya da

$$R_N(s) : \begin{cases} \bar{x}_1 \ge \bar{x}_2 \ \text{ve} \ \underline{x}_0 \ge \frac{1}{2}(\bar{x}_1+\bar{x}_2) & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \bar{x}_1 \ge \bar{x}_2 \ \text{ve} \ \underline{x}_0 < \frac{1}{2}(\bar{x}_1+\bar{x}_2) & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \tag{4.3}$$

dir. $R_N$ kuralı Fisher'in Diskriminant Fonksiyonu ile elde edilir [2 ve 10]. $\Pi_1$ ve $\Pi_2$ sırasıyla $\mu_1$ ve $\mu_2$ ortalamalı normal kitleler olduğunda $R_N$ optimal kuraldır. Kitleler normal olmadığında $R_N$ optimal bir kural olmayabilir. Bu çalışmada, Weibull kitleleri için optimal bir kural elde edilmeye çalışılmıştır.

### 4.1. Parametreler Bilindiğinde Weibull Kitleleri için Diskriminasyon (Discrimination in Weibull Populations with Known Parameters)

$\Pi_1$ ve $\Pi_2$ sırasıyla ortalamaları $\Gamma(1+1/\beta_1)/\theta_1$ ve $\Gamma(1+1/\beta_2)/\theta_2$ olan ve

$$f_{WE}(x;\beta_i,\theta_i)=\beta_i\theta_i x^{\beta_i-1}e^{-\theta_i x^{\beta_i}} \quad , \quad i=1,2 \tag{4.4}$$

olasılık yoğunluk fonksiyonlarına sahip Weibull dağılımlı iki kitle olmak üzere, bu iki yoğunluk fonksiyonunun oranı

$$\frac{f(x;\theta_1,\beta_1)}{f(x;\theta_2,\beta_2)}=\frac{\beta_1\theta_1 x^{\beta_1-1}e^{-\theta_1 x^{\beta_1}}}{\beta_2\theta_2 x^{\beta_2-1}e^{-\theta_2 x^{\beta_2}}} \tag{4.5}$$

olarak elde edilir. Elde edilen bu ifade, parametrelerin bilindiği durum için diskriminant fonksiyonudur. Böylece sınıflandırma kuralı,

$$R_{WE}:\begin{cases} x_0^{\beta_1-\beta_2}e^{-\left[\theta_1 x_0^{\beta_1}-\theta_2 x_0^{\beta_2}\right]} > \dfrac{\beta_2\theta_2}{\beta_1\theta_1}\dfrac{q_2}{q_1}\dfrac{c_{1/2}}{c_{2/1}} & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \text{diğer durumda} & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \tag{4.6}$$

olarak elde edilir. İki kitle için önsel olasılıkların ve hatalı sınıflandırma maliyetlerinin eşit olduğu varsayıldığında $(q_1=q_2=1/2$ ve $c_{1/2}=c_{2/1})$ $R_{WE}$ optimal kuralı,

$$R_{WE}:\begin{cases} x_0^{\beta_1-\beta_2}e^{-\left[\theta_1 x_0^{\beta_1}-\theta_2 x_0^{\beta_2}\right]} > \dfrac{\beta_2\theta_2}{\beta_1\theta_1} & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \text{diğer durumda} & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \tag{4.7}$$

ile verilir.

Sınıflandırma kuralı, $\beta_1$, $\beta_2$, $\theta_1$ ve $\theta_2$ parametrelerinin farklı durumlarına göre aşağıdaki biçimlerde elde edilir.

**Durum 1 (Case 1):** $\beta_1=\beta_2=\beta$ ve $\theta_2>\theta_1$ ise $R_{WE}$ sınıflandırma kuralı,

$$R_{WE}:\begin{cases} x_0>c_1 & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \text{diğer durumda} & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \tag{4.8}$$

olarak elde edilir ve burada $c_1=\left[\dfrac{\log\theta_2-\log\theta_1}{\theta_2-\theta_1}\right]^{1/\beta}$ dir.

$R_{WE}$ kuralı kullanıldığında toplam hatalı sınıflandırma olasılığı

$$P(TPM\,|\,R_{WE})=\frac{1}{2}\left[1-\exp\left(-\theta_1 c_1^{\beta}\right)+\exp\left(-\theta_2 c_1^{\beta}\right)\right] \tag{4.9}$$

dır. $R_{WE}$ deki $WE$ harfi Weibull dağılımını göstermektedir.

**Durum 2 (Case 2):** $\theta_1=\theta_2=\theta$ ise $R_{WE}$ sınıflandırma kuralı,

$$R_{WE}:\begin{cases} (\beta_1-\beta_2)\log x_0-\theta\left(x_0^{\beta_1}-x_0^{\beta_2}\right)>c_2 & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \text{diğer durumda} & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \tag{4.10}$$

olarak elde edilir ve burada $c_2=(\log\beta_2-\log\beta_1)$ dir.

$R_{WE}$ kuralı kullanıldığında toplam hatalı sınıflandırma olasılığının hesaplanabilmesi için $(\beta_1-\beta_2)\ln X-\theta\left(X^{\beta_1}-X^{\beta_2}\right)$ rasgele değişkeninin dağılımının elde edilmesi gerekmektedir.

**Durum 3 (Case 3):** $\beta_1=\beta_2=1$ ve $\theta_1<\theta_2$ ise Weibull dağılımları, konum parametreleri $\theta_1$ ve $\theta_2$ olan üstel dağılımlara dönüşürler. Bu durumda $\theta_1$ ve $\theta_2$ parametreli üstel dağılıma sahip $\Pi_1$ ve $\Pi_2$ kitleleri için $R_E$ optimal sınıflandırma kuralı

$$R_E:\begin{cases} x_0 > \left[\dfrac{(\log\theta_2 - \log\theta_1)}{\theta_2-\theta_1}\right] & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \text{diğer durumda} & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \qquad (4.11)$$

olarak elde edilir. $R_E$ deki $E$ harfi üstel dağılımı göstermektedir. Üstel dağılımlı kitlelerde diskriminasyon problemini Adegboye (1993) ayrıntılı olarak çalışmıştır.

### 4.2. Parametreler Bilinmediğinde Weibull Kitleleri için Diskriminasyon (Discrimination in Weibull Populations with Unknown Parameters)

Parametreler bilinmediğinde $\Pi_1$ ve $\Pi_2$ kitlelerinden alınan $n_1$ ve $n_2$ hacimli $X_{11}, X_{12}, \ldots, X_{1n_1}$ ve $X_{21}, X_{22}, \ldots, X_{2n_2}$ bağımsız rasgele örneklemeler yardımıyla bilinmeyen parametreler tahmin edilebilir. Parametrelerin bilindiği durumda elde edilen diskriminant kurallarında, bilinmeyen parametreler yerine, örneklemlerden elde edilen tahminlerinin direkt olarak kullanılmasıyla (plug-in), örneklemlere bağlı diskriminant kuralları elde edilmiş olur. Böylece örneklemlere bağlı $R_{WE}(s)$ sınıflandırma kuralı

$$R_{WE}(s):\begin{cases} x_0^{\hat\beta_1-\hat\beta_2} e^{-\left[\hat\theta_1 x_0^{\hat\beta_1} - \hat\theta_2 x_0^{\hat\beta_2}\right]} > \dfrac{\hat\beta_2\hat\theta_2}{\hat\beta_1\hat\theta_1} & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \text{diğer durumda} & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \qquad (4.12)$$

ile verilir. Burada $\hat\beta_1$, $\hat\beta_2$, $\hat\theta_1$ ve $\hat\theta_2$ parametrelerin örneklemlerden elde edilen en çok olabilirlik tahminleridir.

Parametreler bilinmediğinde, sırasıyla şekil parametrelerinin eşit $(\beta_1=\beta_2=\beta)$ olduğu ve ölçek parametrelerinin eşit $(\theta_1=\theta_2=\theta)$ olduğu durumlar gözönüne alınarak aşağıdaki biçimlerde örneklemlere bağlı diskriminant kuralları elde edilir.

**Durum 4 (Case 4):** $R_{WE}(s)$ örneklemlere bağlı sınıflandırma kuralı

$$R_{WE}(s):\begin{cases} \bar x_1 < \bar x_2 \text{ ve } \underline{x}_0 < c_1' & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \bar x_1 < \bar x_2 \text{ ve } \underline{x}_0 \ge c_1' & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \qquad (4.13)$$

veya

$$R_{WE}(s):\begin{cases} \bar x_1 \ge \bar x_2 \text{ ve } \underline{x}_0 \ge c_1' & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \bar x_1 \ge \bar x_2 \text{ ve } \underline{x}_0 < c_1' & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \qquad (4.14)$$

olarak elde edilir ve burada $c_1' = \left[\dfrac{\log\hat\theta_2 - \log\hat\theta_1}{\hat\theta_2 - \hat\theta_1}\right]^{1/\hat\beta}$ ve $\hat\beta = \left(\hat\beta_1 + \hat\beta_2\right)/2$ dir.

**Durum 5 (Case 5):** $R_{WE}(s)$ sınıflandırma kuralı,

$$R_{WE}(s):\begin{cases} \bar x_1 < \bar x_2 \text{ ve } \underline{x}_0 < c_2' & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \bar x_1 < \bar x_2 \text{ ve } \underline{x}_0 \ge c_2' & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \qquad (4.15)$$

veya

$$R_{WE}(s):\begin{cases} \bar x_1 \ge \bar x_2 \text{ ve } \underline{x}_0 \ge c_2' & \text{ise } \underline{x}_0 \ \Pi_1\text{'e atanır} \\ \bar x_1 \ge \bar x_2 \text{ ve } \underline{x}_0 < c_2' & \text{ise } \underline{x}_0 \ \Pi_2\text{'ye atanır} \end{cases} \qquad (4.16)$$

olarak elde edilir ve burada $c_2' = \left(\log\hat\beta_2 - \log\hat\beta_1\right)$ ve $\hat\theta = \left(\hat\theta_1 + \hat\theta_2\right)/2$ dir.

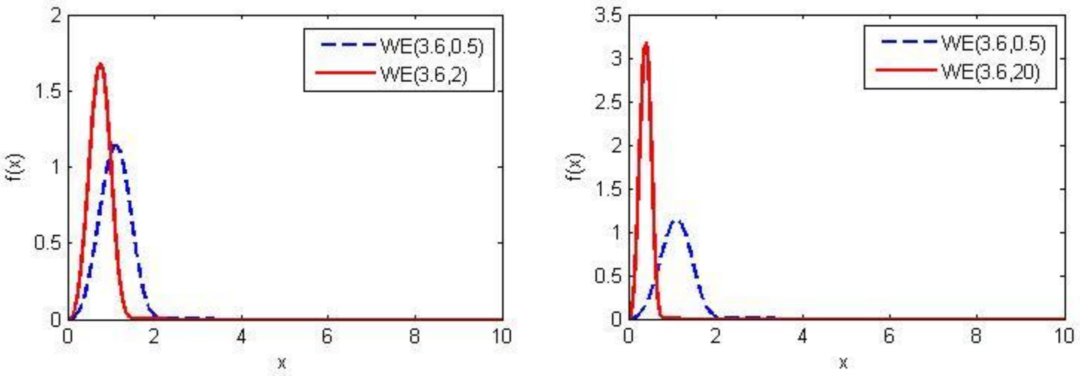
## 5. SİMÜLASYON ÇALIŞMASI (SIMULATION STUDY)

Tek değişkenli Weibull dağılımına sahip kitlelerde, farklı örneklem büyüklükleri $(n_1 = n_2 = n)$ için $\beta_1$, $\beta_2$, $\theta_1$ ve $\theta_2$ parametrelerinin farklı değerleri ve sınıflandırma kurallarına göre hata oranları 1000 tekrar ile elde edilmiştir ve sonuçlar Tablo 1 ve Tablo 2' de verilmiştir.

Tablo 1. Durum 1 ve Durum 4
(Table 1. Case 1 and Case 4)

| | | | | $n=20$ | $n=40$ | $n=60$ | $n=100$ |
|---|---|---|---|---|---|---|---|
| $\theta_1$ | $\theta_2$ | $\beta$ | $\alpha$ | $\hat{\alpha}^R$ | $\hat{\alpha}^R$ | $\hat{\alpha}^R$ | $\hat{\alpha}^R$ |
| 0.5 | 1 | 3.6 | 0.3750 | 0.3722 | 0.3740 | 0.3745 | 0.3749 |
| 0.5 | 2 | 3.6 | 0.2637 | 0.2616 | 0.2591 | 0.2625 | 0.2613 |
| 0.5 | 4 | 3.6 | 0.1749 | 0.1724 | 0.1735 | 0.1751 | 0.1747 |
| 0.5 | 6 | 3.6 | 0.1343 | 0.1331 | 0.1329 | 0.1319 | 0.1334 |
| 0.5 | 8 | 3.6 | 0.1104 | 0.1094 | 0.1088 | 0.1087 | 0.1106 |
| 0.5 | 10 | 3.6 | 0.0943 | 0.0936 | 0.0940 | 0.0944 | 0.0941 |
| 0.5 | 20 | 3.6 | 0.0565 | 0.0565 | 0.0565 | 0.0567 | 0.0564 |
| 2 | 4 | 3.6 | 0.3750 | 0.3751 | 0.3741 | 0.3760 | 0.3749 |
| 2 | 6 | 3.6 | 0.3075 | 0.3138 | 0.3098 | 0.3076 | 0.3078 |
| 2 | 8 | 3.6 | 0.2638 | 0.2654 | 0.2688 | 0.2646 | 0.2638 |
| 2 | 10 | 3.6 | 0.2325 | 0.2361 | 0.2362 | 0.2335 | 0.2322 |
| 2 | 20 | 3.6 | 0.1516 | 0.1540 | 0.1534 | 0.1506 | 0.1521 |
| 4 | 10 | 3.6 | 0.3371 | 0.3383 | 0.3378 | 0.3375 | 0.3372 |
| 4 | 20 | 3.6 | 0.2325 | 0.2350 | 0.2373 | 0.2343 | 0.2335 |
| 4 | 40 | 3.6 | 0.1516 | 0.1564 | 0.1525 | 0.1518 | 0.1524 |

Tablo 1'den, $\theta_1$ ve $\theta_2$ parametreleri arasındaki fark arttıkça iki kitlenin birbirinden uzaklaştığı ve hata oranlarının da azaldığı gözlenmektedir. Bu sonucu Şekil 1'de desteklemektedir. $\theta_1$, $\theta_2$'nin belli bir katı olduğunda (4.9) eşitliği ile elde edilen optimal hata oranı $\alpha$'nın aynı olduğu görülmektedir. Örneğin, $\theta_1 = 0.5$ ve $\theta_2 = 1$ ile $\theta_1 = 2$ ve $\theta_2 = 4$ olarak seçildiğinde optimal hata otanı $\alpha = 0.3750$ olarak elde edilmektedir. Örneklem büyüklüğü $n$ arttıkça, yeniden yerine koyma tahmin edicisi ile elde edilen hata oranı $\hat{\alpha}^R$'nin optimal hata oranı $\alpha$'ya yaklaştığı görülmektedir.



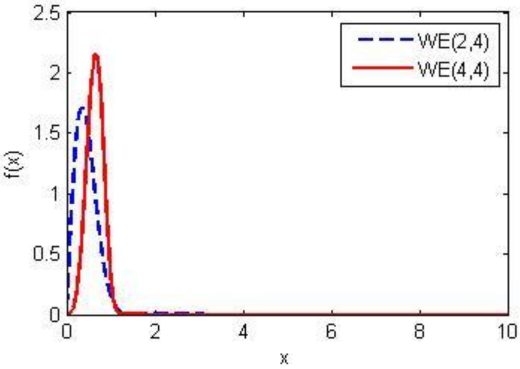

(a) (b)

Şekil 1. (a) $WE(3.6,0.5)$ ve $WE(3.6,2)$ ile (b) $WE(3.6,0.5)$ ve $WE(3.6,20)$ dağılımlarının olasılık yoğunluk fonksiyonları
(Figure 1. The probability density function of Weibull distributions with different parameter values (a) $WE(3.6,0.5)$ and $WE(3.6,2)$ with (b) $WE(3.6,0.5)$ and $WE(3.6,20)$)

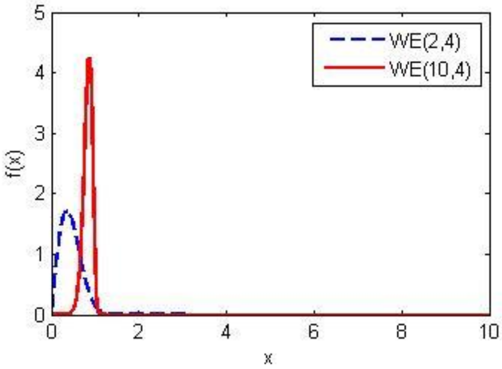
Tablo 2. Durum 5
(Table 2. Case 5)

| | | | $n=20$ | $n=40$ | $n=60$ | $n=100$ |
|---|---|---|---|---|---|---|
| $\beta_1$ | $\beta_2$ | $\theta$ | $\hat{\alpha}^R$ | $\hat{\alpha}^R$ | $\hat{\alpha}^R$ | $\hat{\alpha}^R$ |
| 4 | 2 | 4 | 0.2901 | 0.2903 | 0.2896 | 0.2899 |
| 6 | 2 | 4 | 0.1971 | 0.1963 | 0.1971 | 0.1982 |
| 8 | 2 | 4 | 0.1415 | 0.1418 | 0.1417 | 0.1419 |
| 10 | 2 | 4 | 0.1095 | 0.1111 | 0.1117 | 0.1120 |
| 20 | 2 | 4 | 0.0478 | 0.0482 | 0.0484 | 0.0488 |
| 8 | 4 | 4 | 0.2917 | 0.2914 | 0.2910 | 0.2903 |
| 10 | 4 | 4 | 0.2350 | 0.2355 | 0.2354 | 0.2350 |
| 20 | 4 | 4 | 0.1105 | 0.1115 | 0.1114 | 0.1115 |
| 12 | 6 | 4 | 0.2900 | 0.2907 | 0.2897 | 0.2899 |
| 20 | 6 | 4 | 0.1727 | 0.1735 | 0.1737 | 0.1743 |
| 16 | 8 | 4 | 0.2909 | 0.2908 | 0.2912 | 0.2910 |
| 24 | 8 | 4 | 0.1923 | 0.1923 | 0.1921 | 0.1926 |

Tablo 2'den, $\beta_1$ ve $\beta_2$ parametreleri arasındaki fark arttıkça hata oranının azaldığı gözlenmektedir. Bu azalma Şekil 2'de de görülmektedir. Burada da, $\beta_1$, $\beta_2$'nin belli bir katı olduğunda $\hat{\alpha}^R$ hata oranından optimal hata oranının aynı olacağı söylenebilir. Örneğin, $\beta_1=4$ ve $\beta_2=2$, $\beta_1=12$ ve $\beta_2=6$ ile $\beta_1=16$ ve $\beta_2=8$ olarak seçildiğinde $\hat{\alpha}^R$ yaklaşık olarak 0.29 olarak elde edilmektedir.




(c)




(d)

Şekil 2. (c) $WE(2,4)$ ve $WE(4,4)$ ile (d) $WE(2,4)$ve $WE(10,4)$ dağılımlarının olasılık yoğunluk fonksiyonları
(Figure 2. The probability density function of Weibull distributions with different parameter values (c) $WE(2,4)$ and $WE(4,4)$ with (d) $WE(2,4)$ and $WE(10,4)$)

**6. SONUÇ (CONCLUSION)**

Bu çalışmada, tek değişkenli Weibull dağılımına sahip kitleler için kitle parametrelerinin bilinip bilinmemesi durumları göz önüne alınarak diskriminasyon problemi ele alınmıştır.

Simülasyon çalışmaları sonucunda, her iki kitlenin şekil parametresinin birbirine eşit olduğu durumda $(\beta_1 = \beta_2 = \beta)$, kitlelerin ölçek parametreleri $\theta_1$ ve $\theta_2$ değerlerinin birbirine yaklaştıkça $R_{WE}(.)$ optimal sınıflandırma kuralından elde edilen hata oranlarının daha büyük olduğu Tablo 1'den gözlenmektedir. İki kitlenin ölçek parametreleri arasındaki fark arttığında ise, $R_{WE}(.)$ sınıflandırma kuralı ile elde edilen hata oranlarının daha küçük olduğu gözlenmiştir.

Her iki kitlenin ölçek parametresinin birbirine eşit olduğu durumda $(\theta_1 = \theta_2 = \theta)$ ise, parametrelerin bilindiği durum için $R_{WE}(.)$ optimal sınıflandırma kuralında elde edilen rasgele değişkenin dağılımı analitik olarak elde edilemediğinden optimal hata oranı hesaplanamamıştır. Ancak yeniden yerine koyma tahmin edicisi $\hat{\alpha}^R$ ile hata oranları Tablo 2' de verildiği gibi elde edilmiştir.

Her iki durumda parametrelerin farklı değerlerinin tümü için iki kitle birbirine yakın olduğunda Şekil 1 (a) ve Şekil 2 (c) de olduğu gibi hata oranının diğer durumlara göre daha büyük olduğu gözlenmektedir. Şekil 1 (b) ve Şekil 2 (d) deki gibi iki kitle birbirinden uzaklaştıkça hata oranının azaldığı gözlenmektedir.

**KAYNAKLAR (REFERENCES)**

1. Lachenburch, P.A., (1975), Discriminant Analysis, Hafner: New York.
2. Seber, G.A.F., (1984), Multivariate Observations, John Wiley & Sons. Inc.
3. Anderson, T.W., (1984), Introduction to Multivariate Statistical Analysis,2nd. Ed. Wiley: New York.
4. Welch, B.L., (1939), Note on Discriminant Functions, Biometrika, 31, 218-220.

**5.** Atakan, C. and Öztürk, F., (1997), The Optimal Classification Rule for Gamma Populations, Jour. of Inst. of Math&Comp. Sci., 10, 2, 131-140.
**6.** Hills, M., (1966), Allocation Rules and Their Error Rates, J. Roy. Stat. Soc. B, 28, 1-31.
**7.** Lachenburch, P.A. and Mickey, R., (1968), Estimation of Error Rates in Discriminant Analysis, Technometrics, 10, 1–11.
**8.** Snapin, S.M., (1983), An Evaluation of Smooted Error Rate Estimators in Discriminant Analysis, Institute of Statistics Miemeo Series, 1438.
**9.** Smith, C.A.B., (1978), Some Examples of Discrimination, Annals of Eugenics, 18, 272-282.
**10.** Fisher, R.A., (1936), The use of the multiple measurements in taxonomic problems, Annals of Eugenics, 7, 179-188.
**11.** Adegboye, O.S., (1993). The Optimal Classification Rule for Exponential Populations. Austral. J. Statist., Volume:35, pp.185-194.