

**BAŐKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**ÇOK DEĐİŐKENLİ NORMALLİK TESTLERİNDEN Z_p VE C_p
İÇİN BİR JAVA PROGRAMI VE UYGULAMASI**

BİLGE ÖZGEN TÜRKOĐLU

YÜKSEK LİSANS TEZİ

2009

**ÇOK DEĞİŞKENLİ NORMALLİK TESTLERİNDEN Z_p VE C_p
İÇİN BİR JAVA PROGRAMI VE UYGULAMASI**

**A JAVA PROGRAM FOR THE MULTIVARIATE Z_p AND C_p
TESTS AND ITS APPLICATION**

BİLGE ÖZGEN TÜRKOĞLU

Başkent Üniversitesi
Lisansüstü Eğitim Öğretim ve Sınav Yönetmeliğinin
İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ Anabilim Dalı İçin Öngördüğü
YÜKSEK LİSANS TEZİ
olarak hazırlanmıştır.

2009

Fen Bilimleri Enstitüsü Müdürlüğü'ne,

Bu çalışma, jürimiz tarafından **İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI 'nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Başkan :
Yrd. Doç. Dr. Barış SÜRÜCÜ

Üye (Danışman) :
Yrd. Doç. Dr. İlknur ÖZMEN

Üye :
Yrd. Doç. Dr. Güvenç ARSLAN

ONAY

Bu tez 20/01/2009 tarihinde, yukarıdaki jüri üyeleri tarafından kabul edilmiştir.

/02/2009

Prof. Dr. Emin AKATA

FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRÜ

TEŐEKKÜR

Sayın Yrd. Doç. Dr. İlknur Özmen'e, tez danışmanım olarak tezin oluşmasında, değerlendirilmesinde ve düzenlemesindeki önemli katkı ve destekleri için...

Sayın Yrd. Doç. Dr. Güvenç Arslan'a bu tezin hazırlanış süresinde göstermiş olduđu ilgi ve yardımları için...

Sayın Yrd. Doç. Dr. Barış Sürücü'ye sağladığı tüm olanaklar ve desteđi için...

Desteklerini hiç bir zaman esirgemeyen ve her zaman yanımda olan aileme çok teşekkür ederim.

Bilge Özgen TÜRKOĐLU

ÖZ

ÇOK DEĞİŞKENLİ NORMALLİK TESTLERİNDEN Z_p VE C_p İÇİN BİR JAVA PROGRAMI VE UYGULAMASI

Bilge Özgen TÜRKOĞLU

Başkent Üniversitesi Fen Bilimleri Enstitüsü

İstatistik ve Bilgisayar Bilimleri Ana Bilim Dalı

Bu çalışmada öncelikle tek değişkenli basıklık ve çarpıklık katsayıları, tek değişkenli W , R , Z ve C testleri incelenmiştir. Bir sonraki adım olarak bu tek değişkenli testlerin çok değişkenli uyarlamaları incelenmiştir.

Çok değişkenli normal dağılım testleri içerisinde Z_p ve C_p testlerinin JAVA dilinde uygulaması yapılmıştır. Bu amaçla, bu iki test istatistiğinin dağılım yakınsamalarının doğruluğu simülasyon ile test edilmiştir. Simülasyon sonuçları geliştirilmiş olan yazılımın başarılı olduğu sonucuna ulaşılmasını sağlamıştır. Geliştirilen program ile daha önceden çok değişkenli normal dağılım hakkında bilgi sahibi olunan gerçek veri kümeleri ile karşılaştırıldığında tutarlı sonuçlar vermiştir.

Bu çalışmada geliştirilmiş olan program, paket programlarda bulunmayan çok değişkenli normal dağılım testini başarılı bir şekilde gerçekleştirmektedir. Yapılan çalışmanın, veri kümelerine ilişkin çok değişkenli normal dağılım varsayımının test edilmesinde büyük kolaylık sağlayacağı düşünülmektedir.

ANAHTAR SÖZCÜKLER: Bazı tek değişkenli normallik testleri, çok değişkenli normal dağılım ve testleri, JAVA, Polar Tekniği

Danışman: Yrd. Doç. Dr. İlknur Özmen, Başkent Üniversitesi, İstatistik ve Bilgisayar Bilimleri Bölümü

ABSTRACT

A JAVA PROGRAM FOR THE MULTIVARIATE Z_p AND C_p TESTS AND ITS APPLICATION

Bilge Özgen TÜRKOĞLU

Baskent University Institute of Science

The Department of Statistics and Computer Science

In this comprehensive software study of the multivariate normality tests the first approach is to study the univariate versions of W , R , Z , and C test. Afterwards study is continued with the analysis of these tests multivariate versions.

From these wide alternative multivariate tests, the implementation of the software is processed for the multivariate C_p and Z_p tests in JAVA programming language. In this process the tests approximation to their distributions have been tested with a Monte Carlo Simulation study. The result for these simulations indicates a prominent success. In addition the results which have been accomplished from the studies, the distributions were known from the previous studies, are consistent with the known results.

Knowing the fact that the lack of a multivariate normality test in a package program, the software for multivariate normality which has been developed in this study carries a great innovation to the analysts.

KEYWORDS: Some of univariate normality tests, multivariate normal distribution and tests, JAVA, Polar method

Advisor: Asst. Prof. Dr. İlknur Özmen, Başkent University, Department of Statistics and Computer Science

İÇİNDEKİLER LİSTESİ

	<u>Sayfa</u>
ÖZ	i
ABSTRACT	ii
İÇİNDEKİLER LİSTESİ	iii
ŞEKİLLER LİSTESİ	v
ÇİZELGELER LİSTESİ	vi
EKLER LİSTESİ	vii
SİMGELER VE KISALTMALAR LİSTESİ	v
1. GİRİŞ	1
2. GENEL BİLGİLER	2
2.1 Tek Değişkenli Basıklık ve Çarpıklık Katsayıları	2
2.2 Tek Değişkenli Normallik Testleri	3
2.2.1 Tek değişkenli W testi	4
2.2.2 Tek değişkenli R testi	5
2.2.3 Tek değişkenli Z testi	7
2.2.4 Tek değişkenli C testi	8
2.3 Çok Değişkenli Normal Dağılım	9
2.3.1 Çok değişkenli normal dağılımın özellikleri	10
2.4. Çok Değişkenli Normallik Testleri	11
2.4.1 Çok değişkenli W testi	14
2.4.2 Çok değişkenli Z testi	14
2.4.3 Çok değişkenli R testi	16
2.4.4 Çok değişkenli C istatistiği	16
2.4.5 Çok değişkenli basıklık ve çarpıklık katsayıları	17
3. JAVA PROGRAMLAMA DİLİ İLE Z_p VE C_p TEST İSTATİSTİKLERİNİN KODLANMASI	18
3.1 Programda Kullanılan Testler ve Programın İşleyişi	18
3.2 Sürekli Değişkenlerde Rasgele Sayı Üretme Tekniği	18
3.3 Testlerin Algoritmaları	22
3.3.1 Çok değişkenli Z_p testi algoritması	22
3.3.2 Çok değişkenli C_p testi algoritması	23

4. MONTE CARLO ÇALIŞMASI	25
5. UYGULAMALAR	29
5.1 İris Veri Kümesi Uygulaması	29
5.2 Ter Veri Kümesi Uygulaması	31
5.3 Yüzücü Veri Kümesi Uygulaması	32
5.4 Doğum Veri Kümesi Uygulaması	32
5.5 Vücut Yağ Yüzdeleri Kümesi Uygulaması	33
6. SONUÇ	35
KAYNAKLAR LİSTESİ	38
EKLER	41

ŞEKİLLER LİSTESİ

	<u>Sayfa</u>
Şekil 3.1: V_1 ve V_2 üniform dağılım şekli.....	20

ÇİZELGELER LİSTESİ

	<u>Sayfa</u>
Çizelge 2.1: \sqrt{V} değerleri çizelgesi	15
Çizelge 4.1: $n= 10, p=2$ iken 5 tekrarlanma sayısı için hesaplanan a, b, v değerlerinin sonuçları	27
Çizelge 4.2: Hesaplanan \bar{a}, \bar{b} ve \bar{v} değerleri.....	27
Çizelge 5.1: Setosa veri kümesi için hesaplanan test sonuçları ($p = 4$)	29
Çizelge 5.2: Setosa veri kümesi için hesaplanan test sonuçları ($p = 3$).....	30
Çizelge 5.3: Ter veri kümesi için hesaplanan test sonuçları ($p = 3$).....	31
Çizelge 5.4: Yüzücü veri kümesi için hesaplanan test sonuçları ($p = 3$)	32
Çizelge 5.5: Doğum veri kümesi için hesaplanan test sonuçları ($p = 2$)	33
Çizelge 5.6: Vücut yağ yüzdeleri veri kümesi için hesaplanan test sonuçları ($p = 4$)	34

EKLER LİSTESİ

Sayfa

Ek 1. Shapiro Wilk $a_{i:n}$ katsayıları tablosu.....	41
Ek 2. Fisher setosa grubu için süsen çiçekleri veri kümesi.....	43
Ek 3. Süsen çiçekleri veri kümesi için oluşturulmuş olan Y matrisi ve kısmi regresyon katsayıları (p=4).....	45
Ek 4. Süsen çiçekleri veri kümesi için oluşturulmuş olan Y matrisi ve kısmi regresyon katsayıları (p=3).....	47
Ek 5. Johnson ve Wichern'in ter veri kümesi	49
Ek 6. Ter veri kümesi için oluşturulmuş olan Y matrisi (p = 2)	50
Ek 7. Yüzücü veri kümesi	51
Ek 8. Yüzücü veri kümesi için oluşturulmuş olan Y matrisi (p = 3)	52
Ek 9. Alpar doğum veri kümesi.....	53
Ek 10. Doğum veri kümesi için oluşturulmuş olan Y matrisi (p = 2)	54
Ek 11. Vücut yağ yüzdeleri veri kümesi.....	55
Ek 12. Vücut yağ yüzdeleri için oluşturulmuş olan Y matrisi (p = 4)	56

SİMGELER VE KISALTMALAR LİSTESİ

μ_r'	orijine göre r. moment
μ_r	ortalamaya göre r. moment
γ_1	çarpıklık katsayısı
γ_2	basıklık katsayısı
C	tek değişkenli C istatistiği
W	tek değişkenli W istatistiği
R	tek değişkenli R istatistiği
Z	tek değişkenli Z istatistiği
$a_{i:n}$	Shapiro Wilk' in katsayıları
$x_{(i)}$	i. sıralı istatistik
ρ	kitle korelasyon katsayısı
$Kov(X, Y)$	X ve Y rasgele değişkenleri arasındaki kovaryans
σ_X	X rasgele değişkeninin standart sapması
$\hat{\rho}$	korelasyon katsayısı tahmini
Q_i	i. kitle bölene
G_i	genelleştirilmiş örneklem uzayı
b_1	örneklem çarpıklık katsayısının karesi
b_2	örneklem basıklık katsayısı
MANOVA	çok değişkenli varyans analizi
β_{1j,q_j}	kısmi regresyon katsayısı
C_p	çok değişkenli C istatistiği
W_p	çok değişkenli W istatistiği
R_p	çok değişkenli R istatistiği
Z_p	çok değişkenli Z istatistiği
V	ortak varyans değeri
χ^2	ki-kare rasgele değişkeni
$b_{1,p}$	çok değişkenli çarpıklık katsayısı
$b_{2,p}$	çok değişkenli basıklık katsayısı

1. GİRİŞ

Günümüzde kullanılan tüm parametrik çok değişkenli istatistiksel testler normal dağılım varsayımı sağlandığı durumlarda geçerlidir. Bu testlerdeki parametre tahmin edicilerinin yansız ve seçilen örneklemin de çok değişkenli normal bir dağılımdan geldiği varsayılır. Çok değişkenli normal dağılım varsayımı sağlandığı durumlarda, parametrik olan istatistiksel testlerin geçerliliği de sağlanmış olur. Aksi takdirde birçok yanlışlıkla karşılaşılmaktadır.

Çok değişkenli normallik varsayımını kontrol etmek için birçok test bulunmaktadır. Bu testlerden yaygın olarak kullanılanları Mardia' nın çok değişkenli basıklık ve çarpıklık katsayısı, Shapiro Wilk'in tek değişkenli W testinin çok değişkenliye uyarlanmış hali olan W_p testidir. Bu çalışmada öncelikle tek değişkenli W , R , Z ve C testleri ve basıklık, çarpıklık katsayıları incelenecektir. Daha sonra bu testlerin çok değişkenliye uyarlanmış biçimleri incelenecektir.

İncelenilmiş olan bu testlerden, çok değişkenli Z_p ve C_p testlerinin JAVA programlama dilinde uygulaması yapılacaktır. Yapılan uygulama üzerinde beş farklı gerçek veri kümesi için; Fisher'ın süsen çiçekleri üzerinde toplanmış olduğu çok değişkenli veri kümesi (Setosa veri kümesi), ter veri kümesi, yüzücü veri kümesi, doğum veri kümesi ve vücut yağ yüzdeleri veri kümesi incelenip sonuçlar yorumlanacaktır.

Bu çalışmada amaç paket programların çoğunda bulunmayan çok değişkenli normal dağılım testi için bir JAVA programı geliştirmektir. Bu amaçla uygulamadaki bu eksiklik giderilmeye çalışılacaktır.

2. GENEL BİLGİLER

2.1 Tek Değişkenli Basıklık ve Çarpıklık Katsayıları

İstatistikte bir veri kümesinin normal dağılıma sahip olup olmadığını ölçmek için kullanılan çeşitli ölçütler bulunmaktadır. Belirli bir dağılıma sahip verilerin normallikten sapmalarını tespit etmek için kullanılan ölçütlerden ikisi, basıklık ve çarpıklık katsayılarıdır. Çarpıklık, bir başka deyişle asimetri dağılımın grafiğinde bir kuyruğun diğerinden daha uzun olmasıdır. Bu tür eğrilerde ortalama ve medyan aynı noktada olmamaktadır. Grafikseldan eğriler sağa ya da sola çarpık olarak tanımlanmaktadır. Basıklık ise dağılımda daha karmaşık bir değişimdir. Simetrik bir dağılımın bir orta noktasının var olduğu düşünöldüğünde basıklık orta nokta ve kuyruklardaki veri yoğunluğunu tanımlamaktadır [1].

İstatistiksel ölçümlerde sıklıkla kullanılan basıklık ve çarpıklık ölçütleri, bir rasgele değişkenin momentlerinin özel durumlarıdır [2]. Bir dağılımın momenti ilgili rasgele değişkenin çeşitli kuvvetlerinin beklenen değeridir. İlgilenilen dağılımın birçok özelliği momentler cinsinden ifade edilmektedir. Orijine göre birincil moment ortalama ve ortalamaya göre de ikincil moment varyansı vermektedir [2]. Ortalama ve varyans ölçütlerine ek olarak bir dağılım diğer özellikleri de momentler cinsinden ifade edilebilmektedir. Momentler cinsinden ifade edilebilen tek değişkenli basıklık ve çarpıklık katsayıları normal dağılım için bir test geliştirmede oldukça kullanışlı ölçümlerdir [3].

X , $f_X(x)$ olasılık yoğunluk fonksiyonuna sahip herhangi bir rasgele değişken olsun. Pozitif olan herhangi bir r değeri için;

- X ' in orijine göre r . momenti μ'_r

$$\mu'_r = E(X^r) \quad (2.1)$$

olarak tanımlanmaktadır. $r = 1$ için $\mu'_1 = \mu'$ dür.

- X ' in ortalamaya göre r . momenti μ_r ,

$$\mu_r = E[(X - \mu)^r] \quad (2.2)$$

olarak tanımlanmaktadır.

μ_r ; $j = 1, 2, \dots, r$ olmak üzere μ_j türünden, $(X - \mu)^r$ ifadesinin binom açılımı ile aşağıdaki eşitlikten elde edilir.

$$\mu_r = E[(X - \mu)^r] = \sum_{j=0}^r \binom{r}{j} E(X^j) (-\mu)^{r-j} \quad (2.3)$$

Bu durumda varyans ortalamaya göre ikincil moment olduğundan;

$$\mu_2 = E[(X - \mu)^2] = \sigma^2 = \mu'_2 - (\mu'_1)^2 = \mu'_2 - \mu^2 \quad (2.4)$$

şeklinde ifade edilir. Ayrıca ortalamaya göre üçüncül ve dördüncül moment sırası ile aşağıdaki gibi tanımlanmaktadır.

$$\mu_3 = E[(X - \mu)^3] = \mu'_3 - 3\mu'_1\mu'_2 + 2(\mu'_1)^3 \quad (2.5)$$

$$\mu_4 = E[(X - \mu)^4] = \mu'_4 - 4\mu'_1\mu'_3 + 6(\mu'_1)^2\mu'_2 - 3(\mu'_1)^4 \quad (2.6)$$

Bir olasılık yoğunluk fonksiyonunun çarpıklık katsayısı ortalamaya göre üçüncül moment türünden ifade edilmektedir. Pratikte çarpıklık katsayısı bir olasılık yoğunluk fonksiyonun simetrik yapısını ölçmek için kullanılmaktadır. Çarpıklık katsayısı γ_1 ;

$$\gamma_1 = \frac{E[(X - \mu)^3]}{\sigma^3} \quad (2.7)$$

olarak ifade edilmektedir.

Basıklık katsayısı ise ortalamaya göre dördüncül momentin geliştirilmesi ile bulunmaktadır. Basıklık katsayısı γ_2 ;

$$\gamma_2 = \frac{E[(X - \mu)^4]}{\sigma^4} - 3 \quad (2.8)$$

ifadesine eşittir [2].

2.2 Tek Değişkenli Normallik Testleri

Dağılım fonksiyonu $F(x)$ olan bir rasgele değişkenin bağımsız gözlemleri x_1, x_2, \dots, x_n olsun. $F_0(x)$ belirli bir dağılım fonksiyonuna sahip, kesikli ya da sürekli olmak üzere, aşağıda tanımlanan hipotezi test etmek isteyelim.

$$H_0: F(x) = F_0(x) \quad (2.9)$$

H_0 hipotezinin test edilmesi problemine uyum iyiliği problemi denir. (2.9)' un herhangi bir testine ise uyum iyiliği testi denmektedir [4].

Bu bölümde verilerin tek değişkenli normal dağılıma uygunluğunu araştıran uyum iyiliği testlerinden W , R , Z ve C testleri incelenecektir. Bu testlerde seçilmiş olan örneklemin rasgele olduğu varsayımı altında H_0 yokluk ve H_1 seçenek hipotezleri;

H_0 : Veriler ortalaması ve varyansı bilinmeyen normal dağılıma sahiptir.

H_1 : Veriler ortalaması ve varyansı bilinmeyen normal dağılıma sahip değildir.

biçimindedir.

2.2.1 Tek değişkenli W testi

Shapiro ve Wilk'in W testi, sıralı istatistiklerin beklenen değerlerinin regresyonlarına dayanan bir testtir [4]. Tek değişkenli rasgele bir örneklemin normal dağılıma sahip olup olmadığını test eden W test istatistiği aşağıdaki gibi tanımlanır.

$$W = \frac{(\sum_{i=1}^n a_{i:n} x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad 0 < W < \infty \quad (2.10)$$

Eşitlik (2.10)'daki \bar{x} örneklem ortalaması ve $x_{(i)}$, i . sıralı istatistiktir. a_1, a_2, \dots, a_k olmak üzere $a_{i:n}$ değerleri n örneklem büyüklüğü olarak tanımlandığında; k yaklaşık olarak $n/2$ olmak üzere Ek 1' de verilmiştir.

(X_1, X_2, \dots, X_n) rasgele değişkenlerinin (x_1, x_2, \dots, x_n) olan her bir gözlem değerlerindeki k . en küçük değer $x_{(k)}$; k . sıradaki sıralı istatistik olarak tanımlanmaktadır. Bu durumda $x_{(1)}$ sırası 1 olan sıralı istatistiktir ve her zaman (x_1, x_2, \dots, x_n) 'deki en küçük eleman, $x_{(n)}$ ise en büyük eleman olmaktadır. Bir rasgele örnekleme $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ ise bu örneklem sıralı rasgele örneklem olarak tanımlanmaktadır [5].

Shapiro ve Stephens'in yapmış olduğu çalışmadaki testlerin güçleri karşılaştırmalarına dayanarak söylenebilir ki W testi çoğu tek değişken normallik testinden daha başarılıdır. Tek değişkenli normal dağılımın testinde genel olarak Kolmogorov-Smirnov testi kullanılsa da bu test ancak ve ancak örneklemden tahmin edilmesi gereken bilinmeyen parametreler olmadığı durumlarda geçerlilik taşımaktadır [5].

W testi, kendisini de içerisinde barındıran birleşik bir test uygulandığında yani içerisinde W testi de dahil olmak üzere birden fazla normalliği test eden yeni bir normallik testi öne sürüldüğünde matematiksel hesaplamaları basit kılmaktadır. Bu nedenle birleşik testlerde kullanılması kolay bir testtir [6].

Malkovich ve Afifi [7], Shapiro ve Wilk' in W testi ile Kolmogorov-Smirnov testi ve Mardia'nın basıklık ve çarpıklık katsayılarını incelemiş ve karşılaştırmıştır. Yapmış oldukları Monte Carlo simülasyon çalışmasına göre bu testler arasında testin gücü (Ek 2) açısından belirgin bir fark bulamamışlardır. Ancak daha kapsamlı çalışmalar ile daha iyi sonuçlara ulaşacaklarına ileri sürmektedir.

2.2.2 Tek değişkenli R testi

X ve Y herhangi iki rasgele değişken olarak tanımlandığında bu iki rasgele değişken arasındaki korelasyon katsayısı ρ ;

$$\rho = \frac{Kov(X, Y)}{\sigma_X \sigma_Y} \quad (2.11)$$

olarak tanımlanır [2]. Eşitlikteki $Kov(X, Y)$, X ve Y arasındaki kovaryans ve σ_X, σ_Y ise X ve Y' nin standart sapmasıdır.

Filliben [3] çalışmasında r; normal olasılık grafiğinin korelasyon katsayısı olmak üzere yeni bir test istatistiği öne sürmektedir. Öne sürülen r test istatistiği, $x_{(i)}$ gözlenen değerler ile $N(0,1)$ olan standart normal dağılımın sıralı istatistik medyanları arasındaki çarpım moment korelasyon katsayısıdır. r korelasyon istatistiği (2.11)' de tanımlanmış olan korelasyon katsayısından yola çıkarak;

$$Kor(X_{(i)}, M_{(i)}) = r = \frac{\sum_{i=1}^n (x_{(i)} - \bar{x})(M_{(i)} - \bar{M})}{\sqrt{\sum_{i=1}^n (x_{(i)} - \bar{x})^2 \sum_{i=1}^n (M_{(i)} - \bar{M})^2}} \quad (2.12)$$

olarak tanımlanmaktadır.

r test istatistiğinde kullanılan normal olasılık grafiği, i. sıralı istatistik $x_{(i)}$ 'nin grafiğine karşılık gelen standartlaştırılmış normal dağılımdan gelen i. sıralı istatistiğin bir konum-ölçek parametresi olan $loc(x_{(i)})$ 'nin grafiği olarak tanımlanmaktadır. i. sıralı istatistiğin en yaygın olarak kullanılan konum-ölçek parametresi sıralı istatistik ortalamasıdır. Bu durumda $loc(x_{(i)}) = E(x_{(i)})$ olmaktadır.

Ancak r test istatistiğinde sıralı istatistik ortalamaları yerine $M_{(i)}$ sıralı istatistik medyanları kullanılmıştır. Sıralı istatistik medyanlarının, ortalamalar yerine kullanılmış olmasındaki temel neden, sıralı istatistik ortalamalarının bazı dezavantajları bulunmasıdır. Bu dezavantajlar şu şekildedir.

- Sıralı istatistik ortalamaları hesaplanırken kullanılan entegrasyon tekniği dağılımdan dağılıma değişim göstermektedir ve sıralı istatistik ortalamaları üretilirken her dağılım için farklı bir teknik kullanılmaktadır. Kısaca sıralı istatistik ortalamalarının hesaplanması dağılımdan dağılıma farklılık göstermektedir.
- Sıralı istatistik ortalamaları işlemsel açıdan karmaşıktır ve zaman kaybına yol açmaktadır. Bu nedenle de çoğu zaman ortalamalara tahmin yolu ile bulunmaktadır.
- Son olarak bazı dağılımlar için (örneğin Cauchy) sıralı istatistik ortalamaları tanımlı değildir.

Bu dezavantajlar yüzünden r istatistiğinde sıralı istatistik ortalamaları yerine sıralı istatistik medyanları kullanılmıştır [3]. Sürücü [8] çalışmasında R korelasyon istatistiğini aşağıdaki gibi tanımlamıştır.

$$R = 1 - \hat{\rho}^2, \quad 0 < R < 1 \quad (2.13)$$

R test istatistiğindeki $\hat{\rho}$, i. sıralı istatistik $x_{(i)}$ ile standartlaştırılmış normal sıralı istatistiklerin beklenen değeri olan $\mu_{i:n}$ arasındaki çarpım moment korelasyon katsayısıdır. Sürücü [9] çalışmasında işlemlerde kolaylık olması için $\mu_{i:n}$, standartlaştırılmış normal sıralı istatistiklerin beklenen değerleri yerine Q_i ile ifade edilen kitle bölenlerini (quantiles) kullanmıştır.

Q_i kitle bölenleri aşağıdaki eşitlikten elde edilmektedir.

$$F(Q_i) = \int_{-\infty}^{Q_i} f(z) dz = \frac{i}{n+1}, \quad 1 \leq i \leq n \quad (2.14)$$

Eşitlik (2.14)'te $f(z)$, $z = (x - \mu)/\sigma$ standart rasgele değişkenin olasılık yoğunluk fonksiyonudur. (2.14) eşitliği Q_i için

$$Q_i = F^{-1}\left(\frac{i}{n+1}\right)$$

olarak ifade edilebilir. Bu durumda $(i/(n+1))$, standart normal dağılım tablosundaki olasılıklara ve Q_i 'ler de bu olasılıklara karşılık gelen z değerleri olacaktır. Bu yöntem ile elde edilen Q_i bölenleri, $\mu_{i:n}$ değerleri yerine kullanılmaktadır. Bunun nedeni hesaplama kolaylığının yanında dağılımın μ ve σ parametrelerinin genelde bilinmemesidir.

2.2.3 Tek değişkenli Z testi

Tiku' nun [10] öne sürdüğü Z test istatistiği bir rasgele örneklemin sıralı istatistiklerine ve örneklem farklarına (sample spacing) dayanan tek değişkenli bir normallik testidir. Z istatistiği; $\mu_{i:n}$ standartlaştırılmış normal sıralı istatistiklerin beklenen değeri ve G_i genelleştirilmiş örneklem farkları olmak üzere;

$$Z = \frac{2 \sum_{i=1}^{n-2} (n-1-i) G_i}{(n-1) \sum_{i=1}^{n-1} G_i}, \quad 0 < Z < \infty \quad (2.15)$$

olarak tanımlanmaktadır. G_i ve $\mu_{i:n}$ değerleri aşağıdaki eşitliklerden bulunur.

$$G_i = \frac{x_{(i+1)} - x_{(i)}}{\mu_{i+1:n} - \mu_{i:n}}, \quad 1 \leq i \leq n-1 \quad (2.16)$$

$$\mu_{i:n} = \frac{E[x_{(i)} - \mu]}{\sigma}, \quad 1 \leq i \leq n \quad (2.17)$$

Z istatistiği de, R istatistiği gibi herhangi bir konum-ölçek dağılımının (location-scale distribution) testi için kullanılabilir. Büyük örneklem için (özellikle $n \geq 10$) Z istatistiğinin dağılımı normaldir [9]. Burada test edilen olasılık dağılımı f olmak üzere $(1/\sigma)f((x - \mu)/\sigma)$ türünde bir konum ölçek-dağılımıdır. f 'nin fonksiyonel yapısı tam olarak belirlidir, ancak konum ve ölçek parametreleri olan μ ve σ bilinmemektedir ([10], [11]).

Olasılık teorisinde, özellikle istatistik alanında; bir konum-ölçek ailesi; $\sigma \geq 0$ ölçek ve μ konum parametreleri ile tek değişkenli olasılık dağılımları ailesi olarak ifade edilir. Eğer X olasılık dağılımı bu aileye ait herhangi bir rasgele değişken ise, $Y = \mu + \sigma X$ olarak tanımlanan her dağılım da bu aileye aittir [12].

2.2.4 Tek değişkenli C testi

Günümüzde tüm alternatif dağılımlara karşı güçlü olan, çok kapsamlı bir normallik testi bulunmamaktadır. Bu nedenle yaygın olarak kullanılan normallik testlerinin düzlemsel kombinasyonları daha güçlü bir test elde etmek için kullanılmaktadır. Düzlemsel kombinasyonlar olarak ifade edilen kavram, birden fazla testin bir arada düzlemsel olarak kullanılması sonucunda yeni ve daha güçlü bir test elde edilmesidir. Ancak elde edilen yeni testler de yalnızca belirli bir alternatif dağılıma karşı güçlü olmaktadır. Sürücü [8] tarafından önerilen tek değişkenli C test istatistiği 2.2.1 ve 2.2.2 bölümlerinde değinilmiş olan W ve R istatistiklerinin doğrusal kombinasyonundan oluşmaktadır. Tüm tek değişkenli normal dağılım testleri içerisinde W testi, çarpık ve kısa kuyruklu simetrik alternatifler için, R ise uzun kuyruklu simetrik alternatifler için en güçlü testtir.

C test istatistiği W ve R istatistiklerinin ağırlıklandırılmış toplamından elde edilmekte olup;

$$C = 1 - \{[1 + a_1(a_2 - 1)]W + a_1(1 - a_2)(1 - R)\} \quad (2.18)$$

olarak tanımlanmaktadır. Burada sözü edilen ağırlıklar örneklem çarpıklık ve basıklık katsayıları ile belirlenmektedir. (2.18) eşitliğinde $0 < a_1$ ve $a_2 < 1$ olmak üzere C konum ve ölçeğe göre değişim göstermemektedir. C ' nin büyük değerleri için normal dağılım red edilmektedir. C test istatistiğindeki a_1 ve a_2 katsayıları aşağıdaki eşitliklerden hesaplanmaktadır.

$$a_1 = \exp\left(-\left(\frac{b_1}{0.6}\right)^5\right), \quad a_2 = \exp\left(-\left(\frac{b_2}{3.5}\right)^5\right) \quad (2.19)$$

Eşitlik (2.19)' daki b_1 , örneklem çarpıklık katsayısının karesi ve b_2 , basıklık katsayısıdır. b_1 ve b_2 değerleri (2.7) ve (2.8) eşitliklerinde tanımlanmış olan kitle çarpıklık ve basıklık katsayılarından yola çıkarak hesaplanmaktadır [8].

2.3 Çok Değişkenli Normal Dağılım

Çok değişkenli istatistiksel analizlerin çoğunda veri matrisinin çok değişkenli normal dağıldığı varsayılmaktadır (diskriminant analizi, kanonik korelasyon, MANOVA gibi). Bu varsayım çok değişkenli normal dağılımın avantajlarının bu veri matrisinde kullanılabileceği anlamına gelmektedir. Ancak gerçekte veriler hiçbir zaman tam olarak çok değişkenli normal dağılmamaktadır [13].

Çok değişkenli normal dağılım, tek değişkenli normal dağılımın p boyut olmak üzere $p \geq 2$ olduğu durumlara genellemesidir. Ortalaması μ , varyansı σ^2 ve gösterimi $X \sim N(\mu, \sigma^2)$ şeklinde olan tek değişkenli normal dağılımın olasılık yoğunluk fonksiyonu,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[(x-\mu)]^2}{2\sigma^2}}, \quad -\infty < \mu < \infty, \quad \sigma^2 > 0 \quad (2.20)$$

biçiminde olup fonksiyonun üstel kısımda bulunan

$$\left(\frac{(x-\mu)}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu) \quad (2.21)$$

gösterimi standart değişim birimi olarak x değerinin μ 'ye olan karesel uzaklığını ölçmektedir.

Tek değişkenli normal dağılım fonksiyonundan yararlanarak Σ , $(p \times p)$ boyutlu varyans-kovaryans matrisi ve μ , $(p \times 1)$ boyutlu ortalama vektörü olmak üzere X rasgele vektörü için gösterim $X \sim N(\mu, \Sigma)$ biçimindedir. Σ varyans-kovaryans matrisinin simetrik, sonlu ve pozitif tanımlı olduğu varsayılmaktadır. Buna göre çok değişkenli normal dağılımın olasılık yoğunluk fonksiyonu,

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-(x-\mu)'\Sigma^{-1}(x-\mu)/2}, \quad -\infty < x_j < \infty, \quad \mu_j \in \mu, \quad j = 1, 2, \dots, p \quad (2.22)$$

biçimindedir.

Cramér-Wold teoremine göre, bir rasgele p boyutlu X vektörünün çok değişkenli dağılımı tamamen $a'X$ düzlemsel kombinasyonun tüm tek boyutlu dağılımlarının kümesi ile belirlenmektedir [14]. Burada a herhangi rasgele olmayan p boyutlu bir vektördür. Bu teoremden yola çıkarak çok değişkenli normal dağılımı farklı bir bakış açısı ile tanımlayabiliriz. X , ancak ve ancak $a'X$, a' nın tüm alt

bileşenleri tek değişkenli normal olduğu durumlarda p boyutlu normal dağılıma sahiptir.

Çok değişkenli normal dağılımın bu tanımı dağılımın geometrik yapısını yorumlamak için faydalıdır. X p-boyutlu uzayda rasgele bir nokta olduğu durumlarda X' in doğrusal kombinasyonları X' in tek boyutlu uzaya olan izdüşümleri olacaktır. Bu geometrik yorum X' in dönüşümlere uğramasına rağmen normallik özelliğini koruyacağını belirtmektedir.

2.3.1 Çok değişkenli normal dağılımın özellikleri

Genel olarak çok değişkenli normal dağılımın özellikleri aşağıdaki gibidir [13].

1. Eğer $X \sim N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ dağılımına sahipse $\mathbf{a}'X = a_1X_1 + a_2X_2 + \dots + a_pX_p$ doğrusal kombinasyonu $N_p(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ şeklinde bir dağılıma sahiptir. Ayrıca, $\mathbf{a}'X$ her bir \mathbf{a} için $N_p(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ şeklinde bir dağılıma sahip ise, $X \sim N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ dağılımına sahiptir.
2. $X \sim N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ dağılımına sahip olduğunda, q doğrusal kombinasyonları \mathbf{A} ($q \times p$) boyutlu bir matris olmak üzere;

$$\mathbf{A}_{(q \times p)} \mathbf{X}_{(p \times 1)} = \begin{bmatrix} a_{11}X_1 + \dots + a_{1p}X_p \\ a_{21}X_1 + \dots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \dots + a_{qp}X_p \end{bmatrix}$$

$N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ dağılımına sahiptir. Ayrıca \mathbf{d} sabitler vektörü olmak üzere, $(\mathbf{X}_{(p \times 1)} + \mathbf{d}_{(p \times 1)}) \sim N_p(\boldsymbol{\mu} + \mathbf{d}, \boldsymbol{\Sigma})$ dağılımına sahiptir.

3. X çok değişkenli normal dağılıma sahip ise, X' in tüm alt kümeleri de bu dağılıma sahiptir. X' i parçalara ayırdığımızda, ortalama vektörü ve kovaryans matrisi şu şekildedir:

$$\mathbf{X}_{(p \times 1)} = \begin{bmatrix} \mathbf{X}_{1(q \times 1)} \\ \vdots \\ \mathbf{X}_{2((p-q) \times 1)} \end{bmatrix}, \quad \boldsymbol{\mu}_{(p \times 1)} = \begin{bmatrix} \boldsymbol{\mu}_{1(q \times 1)} \\ \vdots \\ \boldsymbol{\mu}_{2((p-q) \times 1)} \end{bmatrix} \text{ ve}$$

$$\boldsymbol{\Sigma}_{(p \times p)} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

$\begin{matrix} (q \times q) & (q \times (p-q)) \\ ((p-q) \times q) & ((p-q) \times (p-q)) \end{matrix}$

O halde X_1 'in dağılımı $N_q(\mu_1, \Sigma_{11})$ şeklinde olmaktadır.

4. a) Eğer $q_1 = q$ ve $q_2 = p - q$ için $X_{1(q_1 \times 1)}$ ve $X_{2(q_2 \times 1)}$ bağımsız ise $Kov(X_1, X_2) = 0$ ' dir.

b) Eğer $\begin{bmatrix} X_1 \\ \vdots \\ X_2 \end{bmatrix}$ ' nin dağılımı $N_{q_1+q_2} \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$ ise, X_1 ve X_2 ancak ve ancak $\Sigma_{12} = 0$ olduğunda bağımsızdır.

c) Eğer X_1 ve X_2 bağımsız ve sırası ile $N_{q_1}(\mu_1, \Sigma_{11})$, $N_{q_2}(\mu_2, \Sigma_{22})$ dağılımlarına sahipse $\begin{bmatrix} X_1 \\ \vdots \\ X_2 \end{bmatrix}$ matrisinin çok değişkenli normal dağılımı

$$N_{q_1+q_2} \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \right)$$

şeklinde dir.

5. Ortalama vektörü $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_2 \end{bmatrix}$ ve kovaryans matrisi $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ ($|\Sigma_{22}| > 0$)

olan $X = \begin{bmatrix} X_1 \\ \vdots \\ X_2 \end{bmatrix}$ ' nin dağılımı $N_p(\mu, \Sigma)$ olsun. $X_2 = x_2$ bilindiğinde X_1 'in koşullu dağılımı ortalaması, $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ ve kovaryans matrisi $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ olan normal dağılımdır.

6. X , $|\Sigma| > 0$ ile $N_p(\mu, \Sigma)$ dağılımına sahip olsun. O halde $(X - \mu)'(\Sigma)^{-1}(X - \mu)$ (karesel istatistiksel uzaklık) değeri p serbestlik derecesi ile ki-kare χ_p^2 dağılımına sahiptir [13].

2.4. Çok Değişkenli Normallik Testleri

Normal ve yaklaşık normal varyasyonlarda, korelasyon parametresi (ρ), iki değişkenin birbirlerine olan bağımlılıklarını ölçmek için kullanılmaktadır. Bu bağımlılık araştırılırken, uygulamada bazı problemler ile karşılaşmaktadır. Oluşan problemler şu şekilde özetlenebilir. İncelenilen değişkenlerden biri bir başka değişken ile ilişkili olabilir. Yani iki değişkenin birbirleri arasındaki ilişki durumu araştırılırken bu iki değişkenden herhangi birinin başka bir değişken ya da değişken kümesi ile bir ilişkisi olması olasıdır. Bu yapı tüm diğer değişkenler sabit tutulduğunda iki değişken arasındaki korelasyonu incelemeye itmektedir. İki

değişken arasındaki ilişki incelenirken diğer değişkenlerin etkisi sabit tutulduğunda kısmi korelasyonlar hesaplanmaktadır [4].

Başka bir değişkeni sabit tuttuğumuzda iki değişken arasındaki korelasyonda bir düşme söz konusu ise, bu iki değişken arasındaki ilişki başka bir değişken yardımı ile güçlenmektedir. Bu diğer değişken sabit tutulduğunda korelasyon 0'a yaklaşıyor ise, bu iki değişken arasındaki ilişki tamamen sabit tutulan değişken ile oluşmaktadır. Buna karşın diğer değişken sabit tutulduğunda iki değişken arasındaki korelasyon artıyor ise, bu diğer değişkenin iki değişken arasındaki korelasyonu maskeleyiği anlamına gelmektedir.

x_1, x_2, x_3 'ün üç değişkenli normal dağılıma sahip değişken olduğunu varsayalım. C korelasyon matrisi olmak üzere x_i ve x_j arasındaki korelasyon ρ_{ij} olarak olmak üzere olasılık yoğunluk fonksiyonu

$$f(x_1, x_2, x_3) = (2\pi)^{-3/2} |C|^{-1/2} \exp \left\{ -\frac{1}{2|C|} \sum_{i,j=1}^3 C_{ij} x_i x_j \right\} \quad (2.26)$$

biçiminde elde edilir. (2.26)'daki C_{ij} değeri simetrik korelasyon matrisindeki (i, j) 'inci elemana karşılık gelen kofaktördür. x_3 sabit tutulduğunda x_1 ve x_2 arasındaki kısmi korelasyon ise,

$$\rho_{12.3} = -\frac{C_{12}}{(C_{11} \cdot C_{22})^{1/2}} = \frac{\rho_{12} - \rho_{13} \cdot \rho_{23}}{\{(1 - \rho_{13}^2) \cdot (1 - \rho_{23}^2)\}^{1/2}} \quad (2.29)$$

olmaktadır.

Sürücü [9] çalışmasında, tek değişkenli verilerin normal dağılıma uygunluğunu test eden tek değişkenli normallik testlerini, çok değişkenli veriler için genellerken; çok değişkenli doğrusal regresyondaki kısmi regresyon katsayılarının en küçük kareler (EKK) tahmin edicileri ile değiştirilip rasgele ve birbirleriyle ilişkili olmayan değişkenlerden yararlanılmıştır.

p boyutlu x_i değişkenleri sıfır ortalamaları ve σ_i^2 varyansları ile çok değişkenli normal dağılıma sahip olmak üzere x_2, \dots, x_p değişkenlerinin etkisi sabit tutulduğunda x_1 'in beklenen değeri $E(x_1 | x_2, \dots, x_p)$ biçimindedir.

Kısmi regresyon katsayılarının elde edilmesinde

$$\frac{E(x_1|x_2, \dots, x_p)}{\sigma_1} = - \sum_{j=2}^p \frac{C_{1j} x_j}{C_{11} \sigma_j} \quad (2.30)$$

eşitliğinden yararlanılır [4].

Genel ifadesi ile x_1 ' in x_j ' deki kısmi regresyon katsayısı diğer $(p - 2)$ tane değişkenin etkisi sabit tutulduğunda $\beta_{1j.2,3,\dots,j-1,j+1,\dots,p}$ biçimindedir. Gösterimlerde kolaylık olması açısından sabit tutulan diğer değişkenler q_j indisi ile gösterildiğinde kısmi regresyon katsayıları $\beta_{1j.q_j}$ ile gösterilebilir. Buna göre $\beta_{1j.q_j}$ cinsinden eşitlik (2.30)'daki koşullu beklenen değer,

$$E(x_1|x_2, \dots, x_p) = \beta_{12.q_2}x_2 + \beta_{13.q_3}x_3 + \dots + \beta_{1p.q_p}x_p \quad (2.31)$$

olarak ifade edilir. Buna göre $\beta_{1j.q_j}$ kısmi regresyon katsayısı,

$$\begin{aligned} & \beta_{1j.q_j} \\ &= - \frac{\sigma_1 C_{1j}}{\sigma_j C_{11}} \end{aligned} \quad (2.32)$$

eşitliğinden elde edilir [4].

Tek değişkenli normallik testlerinde kullanılan test istatistiklerinin (W , Z , R ve C) p değişkenli uyarlamalarını geliştirmek için p değişkenli normal dağılıma sahip X_1, \dots, X_p değişkenlerinin aşağıdaki doğrusal kombinasyonlarından yararlanılmaktadır [9].

$$\begin{aligned} & X_1, \\ & X_2 - \beta_{2.1}X_1, \\ & X_3 - \beta_{32.1}X_1 - \beta_{32.2}X_2, \dots, \\ & X_p - \beta_{p1.q_1}x_1 - \dots - \beta_{p(p-1).q_{(p-1)}}X_{(p-1)} \end{aligned} \quad (2.33)$$

Buna göre $x_{1i}, x_{2i}, \dots, x_{pi}$ ($1 \leq i \leq n$) rasgele örnekleme ve $\hat{\beta}_{p(p-1).q_{(p-1)}}$ kısmi regresyon katsayılarının EKK tahmin edicileri olmak üzere birbirleriyle ilişkisi olmayan yeni değişkenler,

$$\begin{aligned} & y_{1i} = x_{1i} \\ & y_{2i} = x_{2i} - \hat{\beta}_{2.1}x_{1i} \\ & y_{3i} = x_{3i} - \hat{\beta}_{31.2}x_{1i} - \hat{\beta}_{32.1}x_{2i}, \dots, \\ & y_{pi} = x_{pi} - \hat{\beta}_{p1.q_1}x_{1i} - \dots - \hat{\beta}_{p(p-1).q_{(p-1)}}x_{(p-1)i} \end{aligned} \quad (2.34)$$

olarak elde edilir. Bu yeni değişkenlerin büyük örneklem ($n \geq 10$) için çok değişkenli normal dağılım gösterdiği ve $\mu = \mathbf{0}$ ortalama vektörünün testi için sağlam (robust) istatistiklerin geliştirilmesinde kullanıldığı ifade edilmektedir [9].

2.4.1 Çok değişkenli W testi

Çok değişkenli bir veri kümesinin dağılımının normal olup olmadığını test eden çok değişkenli W_p testi, (2.10)' da belirtilmiş olan tek değişkenli W istatistiğinin p değişkene uyarlanmış biçimidir. Malkovich ve Afifi'nin [7] çalışmasında, Shapiro ve Wilk'in W istatistiğinin çok değişkene uyarlanmış biçimine geçiş için aşağıdaki adımları kullanılmıştır.

X_m ; aşağıdaki eşitliği sağlayan gözlem vektörü olsun.

$$(x_m - \bar{x})' S^{-1} (x_m - \bar{x}) = \max_{1 \leq i \leq n} \{(x_i - \bar{x})' S^{-1} (x_i - \bar{x})\} \quad (2.35)$$

Burada $S = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$ varyans terimidir. Bu durumda , $x_{(i)}$ sıralı istatistikler olmak üzere W_p ;

$$W_p = \frac{(\sum_{i=1}^n a_{i:n} x_{(i)})^2}{(x_m - \bar{x})' S^{-1} (x_m - \bar{x})}, \quad 0 < W_p < \infty \quad (2.36)$$

olarak tanımlanmaktadır. $a_{i:n}$ (2.10)'daki gibi tablo değeridir ve p 'den bağımsızdır. W_p ' nin küçük değerleri çok değişkenli normal olmayan dağılımı belirtmektedir. $p = 1$ olması durumunda W_p , tek değişkenli W testi haline dönüşmektedir. W_p ' nin dağılımına ilişkin nokta yüzdeliği simülasyon ile elde edilmektedir [9].

2.4.2 Çok değişkenli Z testi

Z_j ($1 \leq j \leq p$), $y_{1i}, y_{2i}, \dots, y_{pi}$ ($1 \leq i \leq n$) örnekleminden hesaplanan Z değerleri olmak üzere örneklemin büyüklüğü n , yeteri kadar büyük olduğu durumlarda tüm Z_j değerlerinin ortalaması 1 ve ortak varyansı V olan normal dağılıma sahiptir. Z istatistiğinin p değişkene uyarlanmış biçimi aşağıdaki gibi tanımlanmaktadır.

$$Z_p = \sum_{j=1}^p \left(\frac{Z_j - 1}{\sqrt{V}} \right)^2, \quad 0 < Z_p < \infty \quad (2.37)$$

(2.37) eşitliğindeki ortak varyans V' nin değerleri simülasyonla elde edilmektedir. Belirli örneklem değerleri için elde edilmiş olan \sqrt{V} değerleri Çizelge 2.1 de verilmiştir.

Çizelge 2.1: \sqrt{V} değerleri çizelgesi

n	10	20	30	40	50	70	100
\sqrt{V}	0.179	0.116	0.104	0.100	0.069	0.064	0.048

n' nin ara değerleri için \sqrt{V} değeri interpolasyon tekniği ile bulunmaktadır. İstatistiğin büyük değerleri için p değişkenli normallik red edilmektedir. Z_p ' nin dağılımı p serbestlik derecesi ile asimptotik olarak ($n \rightarrow \infty$) ki-karedir [9].

Bir tabloda X_1 ve X_2 olarak tanımlanmış iki argumentin arasında bulunan bir başka argument X_i 'nin değerini bulmak için interpolasyon tekniği kullanılmaktadır [15]. Doğrusal, ters ve harmonik olmak üzere üç farklı interpolasyon tekniği bulunmaktadır. Bu teknikler içerisinde en basit olanı doğrusal interpolasyondur.

Bu yöntemde $Z = f(X)$ fonksiyonunun X_1 ve X_2 aralıklarında yaklaşık olarak normal dağıldığı varsayılmaktadır. X_i ise bu iki aralık arasında kalmaktadır. X_i değerine karşılık gelen Z değerini bulmak için aşağıdaki eşitlikten yararlanılmaktadır.

$$Z_i = pZ_2 + (1 - p)Z_1 \quad (2.38)$$

(2.38) eşitliğindeki p değeri ise,

$$p = \frac{(X_i - X_1)(X_2 - X_1)}{(X_2 - X_1)^2} \quad (2.39)$$

olarak tanımlanmıştır.

2.4.3 Çok deęişkenli R testi

$1 \leq j \leq p$ olmak üzere $R_j, y_{1i}, y_{2i}, \dots, y_{pi}$ ($1 \leq i \leq n$) rasgele örneklemeden hesaplanmış olan R deęerleri olduęunda, verilerin çok deęişkenli normal daęılıma uygun olup olmadığını test eden p deęişkenli korelasyon istatistięi R_p ařaęıdaki gibi tanımlanır [1]

$$R_p = \sum_{j=1}^p R_j, \quad 0 < R_p < p \quad (2.40)$$

R_p ' nin büyük deęerleri için çok deęişkenli normal daęılım red edilmektedir. R_p ' nin daęılımına iliřkin yüzdelik noktaları $\rho = 0$ için Simülasyon ile bulunmaktadır.

2.4.4 Çok deęişkenli C istatistięi

$1 \leq j \leq p$ olmak üzere $C_j, y_{1i}, y_{2i}, \dots, y_{pi}$ ($1 \leq i \leq n$) rasgele örneklemeden hesaplanmış olan C deęerleri olduęunda, verilerin çok deęişkenli normal daęılıma uygun olup olmadığını test eden C_p test istatistięi řu řekildedir [9],

$$C_p = \sum_{j=1}^p C_j, \quad 0 < C_p < \infty \quad (2.41)$$

C_p ' nin büyük deęerleri için çok deęişkenli normal daęılım red edilmektedir. C_p 'nin daęılımı v serbestlik derecesi ile üçüncül moment ki-kare (three-moment chi-square) daęılımına yaklařmaktadır [9]. Buna göre ki-kare rasgele deęişkeni

$$\chi_v^2 = \frac{C_p + a}{b} \quad (2.42)$$

biçimindedir. Burada a, b ve v deęerleri daęılımın ilk üç momentinden elde edilmektedir.

μ'_1 ; orijine göre birinci moment, μ_2 ve μ_3 sırasıyla ortalamaya göre (merkezsel) momentler olmak üzere a, b ve v deęerleri ařaęıdaki eřitliklerden bulunmaktadır.

$$v = \frac{8}{\gamma_1^2}, \quad b = \sqrt{\frac{\mu_2}{2v}}, \quad a = bv - \mu'_1 \quad (2.43)$$

(2.43) eřitliğinde $\gamma_1 = \mu_3/\mu_2^{3/2}$ eřitliğinden hesaplanan çarpıklık katsayısıdır.

Hesaplanan C_p istatistięi,

$$P(C_p \geq b\chi_{0.90(v)}^2 - a) \quad (2.44)$$

olasılıęını saęlayan kritik deęerle karşılaştırılmaktadır. C_p ' nin büyük deęerleri için çok deęişkenli normal daęılıma uyum red edilecektir.

2.4.5 Çok deęişkenli basıklık ve arpıklık katsayıları

Mardia [16] çok deęişkenli basıklık ve arpıklık ölçümleri tek deęişkenli basıklık ve arpıklık katsayılarını içeren güçlü alıřmalardaki t istatistiklerinin belirli özelliklerinin genişletilmesi ile geliştirilmiştir. Çok deęişkenli arpıklık ve basıklık katsayıları, $S = (x_i - \bar{x})(x_i - \bar{x})'$ olmak üzere sırasıyla;

$$b_{1,p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \{[(x_i - \bar{x})' S^{-1} (x_j - \bar{x})]^3\}, \quad 0 < b_{1,p} < \infty \quad (2.45)$$

$$b_{2,p} = \frac{1}{n} \sum_{i=1}^n \{[(x_i - \bar{x})' S^{-1} (x_i - \bar{x})]^2\}, \quad 0 < b_{2,p} < \infty \quad (2.46)$$

olarak tanımlanır. $b_{1,p}$ arpıklık katsayısının dağılımı asimptotik olarak ki-kareye ve $b_{2,p}$ basıklık katsayısının dağılımı asimptotik olarak normal dağılıma yakınsamaktadır. Bu yakınsamaların çok yavaş olmasından ötürü istatistiklerin yüzdelik noktaları simülasyon yöntemi ile üretilmektedir [9].

3. JAVA PROGRAMLAMA DİLİ İLE Z_p VE C_p TEST İSTATİSTİKLERİNİN KODLANMASI

Bu bölümde önceki bölümlerde incelenmiş olan çok değişkenli normal dağılım testlerinden Z_p ve C_p test istatistiklerinin JAVA dili ile nasıl hazırlandığı kısaca açıklanmıştır.

3.1 Programda Kullanılan Testler ve Programın İşleyişi

JAVA programlama dili ile yazılmış ve Netbeans IDE'si kullanılmış olan programda daha önceki bölümlerde tanımlanmış olan çok değişkenli normal dağılım testlerinden 2.4.2 bölümünde tanımlanmış olan çok değişkenli Z_p testi ve 2.4.4 bölümünde tanımlanmış olan çok değişkenli C_p testinin uygulaması yapılmıştır. Programın genel işleyişi şu şekildedir.

- Programda öncelikle kullanıcıdan değişken sayısı ve örneklem büyüklüğü alınır. Daha sonra bu örneklem büyüklüğü ve değişken sayısına bağlı olarak kullanıcıdan veriler alınır.
- Bu çok değişkenli veri kümesinden (2.32)' de verilen kısmi regresyon katsayıları hesaplanır ve bu katsayılar kullanılarak (2.34)' de tanımlanmış olan y_{pi} değerleri elde edilir.
- y_{pi} değerleri elde edildikten sonra kullanıcı isteğine bağlı olarak girmiş olduğu veri kümesinin çok değişkenli normal dağılıma sahip olup olmadığını Z_p ya da C_p testlerinden birini seçerek ilgili verilerin çok değişkenli normal dağılıma sahip olup olmadığı sonucuna ulaşır.

3.2 Sürekli Değişkenlerde Rasgele Sayı Üretme Tekniği

Programda rasgele sayıların üretilmesini içeren paket GeneratingRandom paketidir. GeneratePolar sınıfında polar (kutupsal) tekniği ile rasgele sayılar üretilmektedir.

X ve Y birbirinden bağımsız birim normal dağılıma sahip rasgele değişkenler ve R ve θ , (X, Y) vektörünün polar koordinatları olsun [17]. Bu durumda,

$$\begin{aligned} R^2 &= X^2 + Y^2 \\ \tan \theta &= \frac{Y}{X} \end{aligned} \quad (3.1)$$

olmaktadır.

X ve Y birbirinden bağımsız rasgele değişkenler olduklarına göre, bu iki değişkenin bileşik yoğunlukları her birinin marjinal yoğunluklarının çarpımına eşit olmaktadır ve aşağıdaki gibi tanımlanmaktadır.

$$f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} \quad (3.2)$$

Birbirinden bağımsız standart normal rasgele değişkenler olan X ve Y yi öncelikle polar koordinatlarını üretip daha sonra bu koordinatları köşegenel koordinatlara dönüştürerek aşağıdaki gibi hesaplayabiliriz.

- U_1 ve U_2 rasgele sayıları üretilir.
- $R^2 = -2\log U_1$ (R^2 , ortalaması 2 ve üstel olmak üzere) ve $\theta = 2\pi U_2$ (θ , 0 ve 2π arasında üniform olmak üzere) değerleri hesaplanır.
- Son olarak X ve Y değerleri aşağıdaki eşitliklerden elde edilir

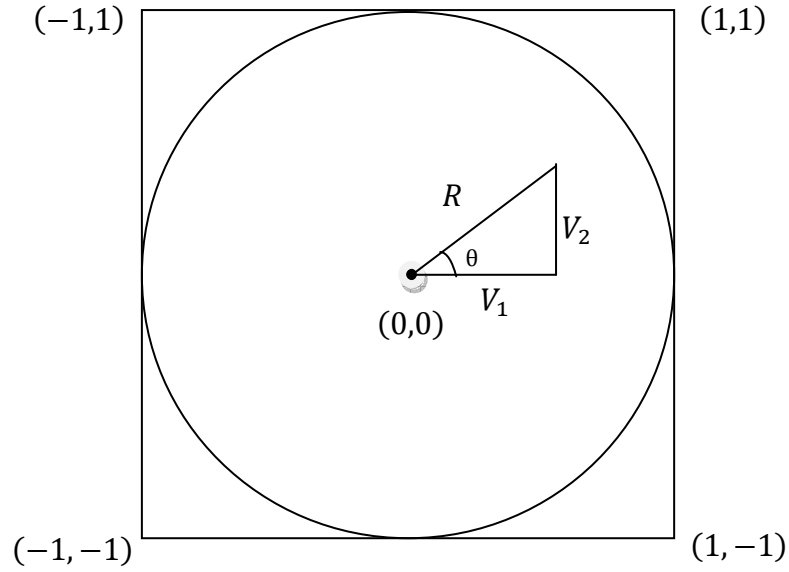
$$\begin{aligned} X &= R \cos \theta = \sqrt{-2\log U_1} \cos(2\pi U_2) \\ Y &= R \sin \theta = \sqrt{-2\log U_1} \sin(2\pi U_2) \end{aligned} \quad (3.3)$$

(3.3)' de tanımlanmış olan dönüşüm, Box-Muller dönüşümü olarak adlandırılmaktadır. Ancak rasgele sayı üretiminde Box-Muller dönüşümünden yararlanmak, hesaplamalar açısından etkin ve kullanışlı değildir. Bunun nedeni sinüs ve kosinüs içeren trigonometrik fonksiyonların hesaplamalarının yapılmasıdır. Bu nedenle sinüs ve kosinüs değerlerinin doğrudan hesaplanmasını gerektirmeyen (3.5) eşitliği kullanılmıştır [17].

Eğer U rasgele değişkeni $(0, 1)$ aralığında üniform ise $2U$ $(0, 2)$ aralığında üniformdur ve $2U - 1$ de $(-1, 1)$ aralığında üniform dağılmaktadır. O halde U_1 ve U_2 rasgele sayıları üretilip

$$\begin{aligned} V_1 &= 2U_1 - 1 \\ V_2 &= 2U_2 - 1 \end{aligned} \quad (3.4)$$

eşitlikleri yazılır ise V_1 ve V_2 $(0, 0)$ merkezli karesel alanda üniform dağılmaktadır.



Şekil 3.1: V_1 ve V_2 uniform dağılım şekli

$V_1^2 + V_2^2 \leq 1$ koşulunu sağlayana kadar (V_1, V_2) değerleri ürettiğimizde bu iki değer çember içerisinde üniform olarak dağılmaktadır. Bu durumda bu çiftin polar koordinatları sırası ile R ve θ olarak tanımlandığında R^2 $(0,1)$ aralığında ve θ $(0, 2\pi)$ aralığında üniform dağıldığına göre, R ve θ birbirinden bağımsızdır. θ rasgele bir açı olduğuna göre θ 'nin sinüs ve kosinüsü; rasgele bir (V_1, V_2) noktası ile aşağıdaki eşitliklerden hesaplanır.

$$\begin{aligned}\sin \theta &= \frac{V_2}{R} = \frac{V_2}{(V_1^2 + V_2^2)^{1/2}} \\ \cos \theta &= \frac{V_1}{R} = \frac{V_1}{(V_1^2 + V_2^2)^{1/2}}\end{aligned}\quad (3.5)$$

(3.3) eşitliğindeki Box-Muller dönüşümünden yola çıkarak birbirinden bağımsız birim normal olan değişkenler U rasgele sayısını üretildikten sonra şu şekilde hesaplanır.

$$\begin{aligned}X &= (-2 \log U)^{1/2} \frac{V_1}{(V_1^2 + V_2^2)^{1/2}} \\ Y &= (-2 \log U)^{1/2} \frac{V_2}{(V_1^2 + V_2^2)^{1/2}}\end{aligned}\quad (3.6)$$

$R^2 = V_1^2 + V_2^2$ olup, $(0,1)$ aralığında üniform ve θ rasgele açılarından bağımsız olduğuna göre U rasgele sayısı yerine kullanılır. $S = R^2$ olarak tanımlanırsa;

$$X = (-2 \log S)^{1/2} \frac{V_1}{S^{1/2}} = V_1 \left(\frac{-2 \log S}{S} \right)^{1/2}$$

$$Y = (-2 \log S)^{1/2} \frac{V_2}{S^{1/2}} = V_2 \left(\frac{-2 \log S}{S} \right)^{1/2} \quad (3.7)$$

değerleri $S = V_1^2 + V_2^2$ ve (V_1, V_2) rasgele seçilmiş çevresi 1 olan orijin merkezli değerler olduğunda bağımsız birim normal olmaktadır.

Özetle polar yöntemi ile birbirinden bağımsız birim normal rasgele değişkenler aşağıdaki adımlar sonucunda bulunur.

- U_1 ve U_2 rasgele sayıları üretilir. Bu rasgele sayılar JAVA'nın Random sınıfı ile üretilmiştir.
- İkinci adım olarak $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$, $S = V_1^2 + V_2^2$ eşitlikleri bulunur.
- Eğer $S > 1$ 'den ilk adıma geri dönülür.
- Son adım olarak X ve Y bağımsız rasgele değişkenlikleri aşağıdaki eşitliklerden bulunur [17].

$$X = \sqrt{\frac{-2 \log S}{S}} V_1, \quad Y = \sqrt{\frac{-2 \log S}{S}} V_2 \quad (3.8)$$

Programda polar tekniği ile X ve Y değerleri üretildikten sonra; GenerateVektorX sınıfında p değişkenli normalliği sağlamak için üretilmiş olan X değerleri belirli bir dönüşüm ile X vektörü haline getirilir. Bu dönüşümde öncelikle birbirinden bağımsız p tane standart normal değişken üretilir. Bu üretilmiş olan değişkenlerin gözlem değerlerinin birbirinden bağımsız olması sağlanır. Daha sonra birbirleriyle ilişkili değişkenler yaratmak için aşağıdaki eşitlikten dönüşümü tamamlanmış değerler üretilir [18].

$$\begin{pmatrix} X_1 \\ X_p \end{pmatrix} \rightarrow \begin{cases} X_1 = W_2 \\ X_2 = \rho X_{j-1} + \sqrt{1 - \rho^2} W_j \end{cases} \quad (3.9)$$

(3.9) eşitliğinde tanımlanmış olan dönüşüm ile üretilen değişkenler bir ArrayList' e atıldıktan sonra GenerateX sınıfı bu üretilmiş olan ArrayList içerisinde bulunan her bir değişkeni örneklem sayısına bağlı olarak alır ve sonuç olarak bir X matrisi oluşturulmuş olur.

Programın tamamında matrisler kullanılırken JAMA paketinden yararlanılmıştır. JAMA; JAVA programlama dili için tanımlanmış standart doğrusal cebir paketidir. JAMA matrislerin oluşturulmasında kullanıcı seviyesinde sınıflar sağlamaktadır. JAMA paketinin kullanılması kolay ve yeterli fonksiyon kapasitesi olmasından ötürü oldukça başarılı bir pakettir [19].

3.3 Testlerin Algoritmaları

Bu bölümde programda kullanılan çok değişkenli Z_p ve C_p test istatistiklerinin algoritmaları kısaca anlatılmaktadır.

3.3.1 Çok değişkenli Z_p testi algoritması

Çok değişkenli Z_p testi kullanıcıdan alınmış olan X veri matrisinin Y matrisine dönüştürülmesinden sonra MultivariateTests paketinde bulunan MultiZ sınıfı içerisinde ana işlemlerini gerçekleştirir.

MultiZ sınıfında yapıcı öncelikle Y matrisini, örneklem büyüklüğünü ve değişken sayısını alır. Daha sonra bu bilgiler doğrultusunda çok değişkenli veri kümesini tek değişkenli veriler haline dönüştürmek için sütun sütun bu verilerin bölünmesi işlemi yapılır. Bölünen her bir veri kümesi ArrayList'e atılır. Bu Y ArrayListindeki tek değişkenli verilerden tek değişkenli Z testi hesaplamaları yapılır.

Tek değişkenli Z testi hesaplamaları UnivariateTests paketinde bulunan CalZ sınıfında yapılmaktadır. CalZ sınıfı yapıcısında MultiZ sınıfında yaratılmış olan içerisinde tek değişken veri kümesini barındıran Y_i ArrayListini alır. Alınmış olan Y_i listesi öncelikle sıralı istatistikleri elde etmek için küçükten büyüğe doğru sıralanır. Daha sonra ise (2.16)' da tanımlanmış olan G_i değerleri GiCalculate fonksiyonu ile hesaplanarak tek değişkenli Z değeri elde edilir. G_i değerleri hesaplanırken (2.17) eşitliğindeki $\mu_{i:n}$ değerleri Harter [24]' in normal sıralı istatistiklerin beklenen değerleri tablosundan alınmıştır.

Her bir Z değerinden (2.37)'de tanımlanmış olan çok değişkenli Z_p değeri bulunur. Bu hesap değerindeki ortak varyans değeri örneklem büyüklüğü 10, 20, 30, 40 ve 50 olduğu durumlarda doğrudan bu örneklem büyüklüklerine ait değerlerin CommonV adındaki listeden (ArrayList) alınmasıyla elde edilir. Eğer örneklem büyüklüğü bu beş değere eşit değil ise, doğrusal interpolasyon yöntemi ile ortak varyans elde edilir.

Bir istatistiksel tabloda iki değer arasında kalan bir başka değeri bulmak için (2.4.2) bölümünde ifade edilen interpolasyon tekniği kullanılmaktadır.

3.3.2 Çok değişkenli C_p testi algoritması

Çok değişkenli C_p testi kullanıcıdan alınmış olan X veri matrisinin, Y matrisine dönüştürülmesinden sonra MultivariateTests paketinde bulunan MultiC sınıfı içerisinde ana işlemlerini gerçekleştirir.

MultiC sınıfında yapıcı; örneklem büyüklüğünü, değişken sayısını ve dönüştürülmüş Y matrisini alır. Bir sonraki adım olarak UnivariateTests paketinde bulunan tek değişkenli C değerlerini hesaplayan CalC sınıfının bir nesnesi oluşturularak C değerleri her bir ArrayList için hesaplanarak toplanır ve C_p değeri elde edilir.

Tek değişkenli C değerlerinin hesaplama işlemlerinin yapıldığı CalC sınıfının yapıcısında öncelikle Y ArrayList'i alınır. Daha sonra (2.18) eşitliğinden tek değişkenli C eşitliğini elde edebilmek için ihtiyacımız olan W değerini elde etmek için CalW sınıfının bir nesnesi tanımlanır.

CalW sınıfında yapıcı gönderilmiş olan listeyi ve örneklem büyüklüğünü alır. Bu değerler alındıktan sonra (2.10) eşitliğindeki tek değişkenli W değeri hesaplanır. W eşitliğindeki $a_{i:n}$ değerlerini hesaplamak için WCoefficients isimli sınıftan yararlanılmaktadır. CalW sınıfında yaratılan WCoefficients nesnesi ile çağırılan sınıf örneklem büyüklüğüne göre Coefficient.txt isimli dosyadan o örneklem büyüklüğüne ait olan $a_{i:n}$ değerini bulur ve geri gönderir.

CalW sınıfı ile hesaplanmış olan W değerinin ardından CalC sınıfında tek değişkenli C eşitliğindeki, (2.18) R istatistiğinin hesaplanması için CorrelationCal sınıfının bir nesnesi tanımlanır.

Tanımlanmış olan CorrelationCal nesnesi ile bu sınıftaki yapıcı gönderilmiş olan listeyi alır. Alınan liste Descriptives sınıfındaki Ordersta yöntemi ile küçükten büyüğe doğru sıralanır. Sıralı istatistikler elde edildikten sonra (2.13) eşitliğinden tek değişkenli R istatistiği değeri elde edilir. R istatistiği hesaplanırken (2.14) eşitliğindeki kitle bölenleri yerine Harter [24]' in normal sıralı istatistiklerin beklenen değerlerini veren tablosundan yararlanılmıştır. R istatistiği değeri elde edildikten sonra tek değişkenli C istatistiği eşitlik (2.18)'de a_1 , a_2 , W ve R değerleri yerine konularak elde edilir.

4. MONTE CARLO ÇALIŞMASI

Monte Carlo çalışması rasgele örnekleme tekniğinin kullanımı ile genellikle bir bilgisayar simülasyonunun yardımıyla, özellikle çözümün olasılıksal hesaplamalarının değerlerini veren matematiksel ya da fiziksel sorunlara yaklaşımsal çözümler getirmeye yarayan bir yöntemdir [20].

Bazı durumlarda teorideki yaklaşımın sonucunu bulmak neredeyse imkansızdır. İşte bu durumlarda Monte Carlo simülasyonu araştırmacılara teorik yaklaşımın bir alternatifini sunar. Deneysel bir alternatif sonucun bulunması günümüzde bilgisayar alanındaki teknolojinin gelişmesi sonucunda olasıdır.

Her ne kadar istatistiksel teoriler yeterli olsa da bir istatistiksel teorinin geçerliliği belirli varsayımlara dayanmaktadır. Kullandığımız veriler eğer bu varsayımları sağlıyor ise istatistiksel teoriler sayesinde örneklem dağılımlarının karakteristik özelliklerini geçerli ve yeterli tahminlerle elde edebiliriz. Öte yandan eğer sahip olduğumuz veriler bu varsayımların ihlaline neden oluyor ise, teoriye dayalı olarak bulunan belli örneklem dağılımlarının tahminlerinin geçerliliği belirsizleşmektedir; bu nedenle de doğrudan teorilere bağlı kalırsak bulacağımız sonuçların hatalı olması kaçınılmazdır. İşte bu durumlarda Monte Carlo Simülasyon (MCS) çalışması araştırmacılar için yararlı bir hale gelmektedir. Çünkü MCS örneklem dağılımlarının karakteristiklerini teorik beklentilerden çok deneysel yaklaşımlarla bulmaya dayanan bir yaklaşımdır. Yeterli sayıda üretilen örneklem ile deneysel sonuçlar asimptotik olarak teorik sonuçlara ulaşacaktır ve bu teorik sonuçlara ulaşılabilindiğinde kanıtlanabilmektedir.

Özet olarak çeşitli dallardaki nicel araştırmalarda, Monte Carlo çalışması bir istatistiksel teorinin yetersizliği ya da geçerliliği olmaması durumunda kullanılmaktadır.

Bazı durumlarda; bir istatistiğin çok karmaşık olmasından kaynaklı olarak, istatistiğin teorik örneklem dağılımı var olmayabilir. Böyle durumlarda Monte Carlo çalışmaları teorik örneklemin dağılımını deneysel olarak bulmaya yaramaktadır. Örneğin diskriminant analizi ve kanonik korelasyon bir çok alanda sıklıkla kullanılan çok değişkenli istatistiksel tekniklerdir. Her iki teknikte de regresyon

analizindeki regresyon katsayılarına benzer fonksiyon katsayıları bulunmaktadır. Buna ek olarak her iki istatistiksel yöntemde de ölçümsel değişkenler ve fonksiyonlar arasındaki korelasyonu veren yapısal katsayılar bulunmaktadır. Bu iki istatistiğin de çok karmaşık olmasından kaynaklı olarak; her iki istatistik içinde bu katsayıların teorik dağılımları bulunmamaktadır. Bu nedenle de Monte Carlo çalışması kullanılarak örneklemin deneysel dağılımları bulunmakta ve sonuçlara buradan ulaşılmaktadır.

Sürücü' nün [9] çalışmasındaki çok değişkenli Z_p , R_p ve C_p istatistiklerinin sıfır hipotezi altında yaklaştığı teorik örneklem dağılımlarının karmaşık yapısı nedeni ile bu istatistiklere bağlı hipotezler test edilirken teorik örneklem dağılımlarına simülasyon yardımı ile ulaşılır.

Bu çalışmada çok değişkenli Z_p ve C_p testlerinin dağılım yakınsamalarının doğruluğunu ölçme amacı ile Simülasyon çalışması yapılmıştır. 2.4.2 bölümünde tanımlanmış olan çok değişkenli Z_p istatistiğinin dağılımı p serbestlik derecesi ile asimtotik olarak ki-kareye yakınsaması aşağıdaki olasılık hesaplanarak test edilmektedir.

$$P[Z_p \geq \chi_{0.9(p)}^2]$$

2.4.4 bölümünde tanımlanmış olan C_p test istatistiğinin dağılımı v serbestlik derecesi ile üçüncül moment ki-kareye asimtotik olarak yakınsamaktadır. Bu yakınsama aşağıda tanımlanmış olan olasılığın hesaplanması sonucunda elde edilmektedir.

$$P[C_p \geq b\chi_{0.9(v)}^2 - a]$$

Yukarıdaki olasılığa karşılık gelen değerler Simülasyon yardımı ile elde edilmektedir. Simülasyon yapılırken öncelikle ilk adımda 10 000 tane C_p değeri üretilmektedir. Daha sonra üretilmiş olan C_p ' değerlerinden a , b ve v değerleri elde edilir. Bu işlem tekrarlanarak \bar{a} , \bar{b} ve \bar{v} ortalama değerleri elde edilir. Bu değerlerden yararlanarak

$$A = \bar{b}\chi_{0.9(\bar{v})}^2 - \bar{a}$$

değeri hesaplanır. Elde edilmiş olan A değeri ile karşılaştırmak üzere 10 000 tane başlangıçta üretilenden farklı C_p değeri üretilir. $C_p \geq A$ koşulunu sağlayan C_p

değerlerinin sayısı toplam deneme sayısına bölünerek (bu durumda 10 000) bir olasılık değerine ulaşılmış olunur.

Bu çalışmada Monte Carlo simülasyon yöntemi ile yapılmış olan işlemler geliştirilmiş programdaki Simülasyon adlı paket içerisinde gerçekleştirilmiştir. C_p testi için yapılmış olan simülasyonda değişken sayısı 2 ($p = 2$) ve örneklem büyüklüğü 10 ($n = 10$) olarak alındığında 5 tekrarlanma sayısı için hesaplanan a , b ve v değerleri Çizelge 4.1' de verilmiştir.

Çizelge 4.1: $n = 10$, $p = 2$ iken 5 tekrarlanma sayısı için hesaplanan a , b , v değerlerinin sonuçları

Tekrarlanma Sayısı	v	b	a
1	4.866610588	0.05048803946	0.24568284173
2	5.012108755	0.04918097533	0.24647029230
3	4.881897367	0.04990694445	0.24360844549
4	4.963217535	0.04984881364	0.24738440340
5	5.035963481	0.04926908985	0.24808953526

Elde edilen a , b ve v değerlerinin ortalamaları alınarak \bar{a} , \bar{b} ve \bar{v} değerleri elde edilmiştir. Bu ortalama değerleri Çizelge 4.2'de verilmiştir. $\bar{v} = 5$ serbestlik dereceli ki-kare tablo değeri 9.236' dır. A değeri ise 0.2096822331 olarak hesaplanmıştır. a , b ve v değerleri bulunurken üretilmiş olan C_p değerinden bağımsız olarak üretilen 10 000 tane C_p değeri ile A değeri karşılaştırıldığında $C_p \geq A$ koşulunu sağlayan C_p değerlerinin olasılığı 0.1308 olarak hesaplanmıştır. Üretilmiş olan değerlerin sonucuna bakıldığında C_p test istatistiğinin dağılımı, \bar{v} serbestlik dereceli üçüncül moment ki-kareye asimptotik olarak yakınsamaktadır.

Çizelge 4.2: Hesaplanan \bar{a} , \bar{b} ve \bar{v} değerleri

\bar{v}	\bar{b}	\bar{a}
4.941364031	0.049595977	0.245010963

Z_p testi için yapılmış olan simülasyonda örneklem büyüklüğü 20 ve değişken sayısı 2 olarak alındığında üretilen 10 000 tane Z_p değeri ile A değeri karşılaştırıldığında $Z_p \geq A$ koşulunu sağlayan Z_p değerlerinin olasılığı 0.0849 olarak hesaplanmıştır. Bu değer 0.1 değerine çok yakın çıkması Z_p testinin asimtotik olarak ki-kare dağılımına yakınsadığını göstermektedir..

5. UYGULAMALAR

Bu bölümde çeşitli kaynaklardan elde edilen veri kümelerinin çok değişkenli normal dağılıma sahip olup olmadığı incelenmiştir. Bu veri kümelerinin daha önce çok değişkenli normal dağılımlı olup olmadığı farklı yöntemler ile araştırılmıştır. Bu çalışmada ise geliştirilmiş olan programdan uygulama sonuçları verilmiştir.

5.1 İris Veri Kümesi Uygulaması

Bu çalışmada Fisher'in süsen (iris germanica) çiçekleri üzerinden toplamış olduğu çok değişkenli veri kümesinin (Ek 2) uygulaması gerçekleştirilmiştir. Fisher'in süsen çiçeği veri kümesi Setosa, Versicolour ve Virginica olmak üzere üç farklı çiçek türünden (grup) oluşmaktadır. Her bir grupta 4 değişken bulunmaktadır.

Bu değişkenler;

- X_1 : çanak yaprak uzunluğu
- X_2 : çanak yaprak genişliği
- X_3 : taç yaprak uzunluğu
- X_4 : taç yaprak genişliği

olarak tanımlanmıştır [21, 22].

Setosa veri kümesi için yapılmış olan uygulamada $p = 4$, $n = 50$ olduğunda kısmi regresyon katsayıları ile X veri matrisinden hesaplanan Y matrisi Ek 3'te verilmiştir. Y matrisi üzerinden yapılan işlemler sonucunda hesaplanan Z_4 ve C_4 değerleri bu değerlere ait tablo kritik değerleri ve testlerin sonuçları (5.1) Çizelgesinde özetlenmektedir.

Çizelge 5.1: Setosa veri kümesi için hesaplanan test sonuçları ($p = 4$)

Test	Hesaplanan değer	Kritik değer	Sonuç
Z_4	11.4197	7.7779	$Z_4 > Z_k$
C_4	0.16475	0.1510	$C_4 > C_k$

C_4 değerinin kritik değeri olan C_k , $A = b\chi_{0.9(v)}^2 - a$ değerinin hesaplanması ile bulunmaktadır. Bu değerler daha önceden simulasyon yolu ile hesaplanmıştır. a, b ve v değerleri örneklem büyüklüğü ve değişken sayısına göre dosyalardan alınmakta ve işlemler buna göre yapılmaktadır. Yapılmış olan uygulamaların tümünde C_k kritik değeri bulunurken bu yöntem kullanılmıştır. Z_k kritik değeri 0.90 güven düzeyindeki 4 serbestlik derecesi ile ki-kare tablo değeridir.

Çizelge sonuçları yorumlanacak olursa Z_4 hesap değeri 11.4197 bulunmuştur. Bu değer kritik değerden büyük olmasından dolayı çok değişkenli Z testine göre veriler %90 güvenilirlik düzeyinde çok değişkenli normal dağılımdan sapmaktadır.

C_4 testinin sonuçlarına göre, C_4 değeri 0.16475 olarak bulunmuştur. Bu değer karşılaştırılan kritik değerden (0.1510) büyük olduğu görülmektedir. Sonuç olarak Setosa veri kümesi, çok değişkenli C testine göre %90 güvenilirlik düzeyinde 4 değişkenli normal dağılıma sahip değildir.

Setosa veri kümesi için taç yaprak genişliği (X_4) değişkenin normal dağılıma sahip olmadığı bilinmektedir [20]. Bu bilgi doğrultusunda X_4 değişkeni veri kümesinden çıkarıldığında testlerdeki hesaplanan değerler ile kritik değerlerin birbirine daha yakın çıkması beklenmektedir.

X_4 değişkeni çıkarıldıktan sonra $p = 3$, $n = 50$ değerleri için kısmi regresyon katsayıları ile X veri matrisinden hesaplanan Y matrisi Ek 4' te verilmiştir. Y matrisi üzerinden yapılan işlemler sonucunda hesaplanan Z_3 ve C_3 değerleri ve bu değerlere ait tablo kritik değerleri ve testlerin sonuçları Çizelge 5.2' de özetlenmektedir.

Çizelge 5.2: Setosa veri kümesi için hesaplanan test sonuçları ($p = 3$)

Test	Hesaplanan değer	Kritik değer	Sonuç
Z_3	0.460138	6.251	$Z_3 < Z_k$
C_3	0.06359	0.115836100	$C_3 < C_k$

Çizelge 5.2' deki sonuçlara göre X_4 değişkeni veri kümesinden çıkarıldığında her iki test için de üç değişkenli normallik sağlanmaktadır.

5.2 Ter Veri Kümesi Uygulaması

Johnson ve Wichern [13]' in üç değişkenli normalliğini test ettikleri veri kümesinde 20 sağlıklı kadının terlemeleri ile ilgili veriler toplanmıştır. Veri kümesinde incelenmiş olan üç değişken bulunmaktadır. Bu değişkenler;

- X_1 : Terleme oranı
- X_2 : Sodyum içeriği
- X_3 : Potasyum içeriği

olarak tanımlanmaktadır (Ek 5).

Johnson ve Wichern çok değişkenli normallik varsayımını $Q - Q$ grafiği ile test etmiştir. Her bir değişken için marjinal grafiklere bakıldığında değerlerin normal dağılıma yakın olduğu sonucuna ulaşılmıştır.

Bu çalışmada geliştirilmiş olan program ile hesaplanan test sonuçları Çizelge 5.3' te özetlenmiştir. Ayrıca hesaplanmış olan Y matrisi ve veri kümesi Ek 6'da verilmiştir.

Çizelge 5.3: Ter veri kümesi için hesaplanan test sonuçları ($p = 3$)

Test	Hesaplanan değer	Kritik değer	Sonuç
Z_3	2.535109	6.251	$Z_3 < Z_k$
C_3	0.0829	0.200254	$C_3 > C_k$

Çizelge 5.3'e göre, üç değişkenli Z_3 ve C_3 testleri sonuçlarına bakıldığında Z_3 testine göre %90 güvenilirlik düzeyinde veri kümesinin normal dağılıma sahip olduğu sonucuna ulaşılmaktadır. Buna karşın C testinin sonucuna bakıldığında %90 güvenilirlik düzeyinde veri kümesinin normal dağılıma sahip olmadığı sonucuna ulaşılmaktadır. Sürücü [9]'nün yapmış olduğu çalışma sonucuna göre çok değişkenli C_p testinin daha güçlü bir test olduğu bilinmektedir. Bu durumda çok değişkenli C_p testinin sonucu dikkate alınabilir.

5.3 Yüzücü Veri Kümesi Uygulaması

Alpar [23]' in incelemiş olduğu veri kümesi en az 20 kez milli olmuş, 18 yaşın üzerindeki erkek orta mesafe yüzücülerinin oksijen tüketim kapasitelerinin beç presteki maksimum kaldırma kuvvetlerinin ve genel spor bilgisi puanlarının belirli rakamlara eşit olup olmadığının incelenmesi üzerine toplanmıştır (Ek 7). İstenilen test uygulanmadan önce çok değişkenli verilerin normal dağılıma sahip olup olmadığının araştırılması yapılmış ve sonuç olarak normallikten aşırı bir ayrılış olmadığı saptanmıştır. Hesaplanmış olan Y matrisi ve veri kümesi Ek 8'de verilmiştir.

Çok değişkenli C ve Z testlerine bu veri kümesi uygulanmış ve elde edilen sonuçlar 5.4 çizelgesinde özetlenmiştir. Bu sonuçlara bakıldığında Z_3 değerinin Z kritik değerinden küçük olduğu gözlemlenmektedir ($1.12984 < 6.251$). Z testi sonucuna göre yüzücü verileri %90 güvenilirlik düzeyinde üç değişkenli normallik varsayımını Alpar [23]'ün de belirttiği gibi sağlamaktadır. C_3 testine göre de hesaplanmış olan değer, kritik değerden küçük çıkması bu verilerin üç değişkenli normal dağılıma %90 güvenilirlik düzeyinde sahip olduğunu göstermiştir.

Çizelge 5.4: Yüzücü veri kümesi için hesaplanan test sonuçları ($p = 3$)

Test	Hesaplanan değer	Kritik değer	Sonuç
Z_3	1.12984	6.251	$Z_3 < Z_k$
C_3	0.09639	0.206	$C_3 < C_k$

5.4 Doğum Veri Kümesi Uygulaması

Testlerin uygulandığı bu veri kümesi en az bir canlı doğum yapmış 15 kadına ilişkin;

- X_1 : Hemoglobin düzeyi
- X_2 : Canlı doğum sayısı
- X_3 : Kadının yaşı
- X_4 : Çocuk bakım bilgi puanı

değişkenlerini içermektedir. Bu dört değişken içerisinde hemoglobin düzeyi ile kadının yaşı değişkenlerinin iki değişkenli normal dağılıma sahip olup olmadıkları

incelenmiştir (Ek 9). C_2 ve Z_2 testlerinden elde edilen sonuçlar Çizelge 5.5'te verilmiştir.

Çizelge 5.5: Doğum veri kümesi için hesaplanan test sonuçları ($p = 2$)

Test	Hesaplanan değer	Kritik değer	Sonuç
Z_2	0.1064	4.605	$Z_2 < Z_k$
C_2	0.062	0.1730	$C_2 < C_k$

Çok değişkenli Z istatistiğine göre hesaplanmış olan Z_2 değeri 0.1064 olarak bulunmuştur. ($p = 2$) serbestlik dereceli χ^2 tablo kritik değeri ise 4.605' dir. $Z_2 < Z_k$ olduğundan yaş ve hemoglobin düzeyini içeren iki değişkenli veri kümesi %90 güvenilirlik düzeyinde iki değişkenli normal dağılıma sahiptir.

Çok $C_2 < C_k$ (0.062<0.1730) olduğundan en az bir kere doğum yapmış kadınların yaş ve hemoglobin düzeyi bilgilerinden oluşan veri kümesinin %90 güvenilirlik düzeyinde iki değişkenli normal dağıldığı sonucuna ulaşılmaktadır.

Alpar [23] grafik yöntemi ile yapmış olduğu testin sonucunda X_1 ve X_3 değişkenlerinin içerdiği veri kümesinin iki değişkenli normal dağılmadığı sonucuna ulaşmıştır. Bu farklılığın nedeni, grafik yöntemlerinin testler kadar hassas olmamasından kaynaklanabilir.

5.5 Vücut Yağ Yüzdeleri Kümesi Uygulaması

30 sporcunun 4 farklı zaman dilimindeki vücut yağ yüzdeleri bilgilerinin oluşturduğu veri kümesi Ek 11'de verilmiştir. Alpar [23] bu veri kümesinde bulunan uç değerleri çıkartarak veri kümesinin dört değişkenli normal dağılım gösterip göstermediğini incelemiştir. Bu veri kümesi üzerinde yapılmış olan dört değişkenli C ve Z testlerinin sonuçları Çizelge 5.6' da verilmiştir.

Çizelge 5.6: Vücut yağ yüzdeleri veri kümesi için hesaplanan test sonuçları ($p = 4$)

Test	Hesaplanan değer	Kritik değer	Sonuç
Z_4	1.1830	7.779	$Z_4 < Z_k$
C_4	0.2011	0.2203	$C_4 < C_k$

Bu tablo sonuçlarına göre C_4 değerinin 0.2011 olarak bulunduğu ve kritik değerden (0.2203) küçük olduğu görülmektedir. Sonuç olarak C_4 istatistiğine göre verilerin %90 güvenilirlik düzeyinde dört değişkenli normal dağıldığı söylenebilir. Benzer sonuç Z_4 istatistiği için de söylenebilmektedir ($1.1830 < 7.779$).

6. SONUÇ

Günümüzde çok deęişkenli bir veri kümesinin normal dağılıma sahip olup olmadığını inceleyen bir çok test bulunmaktadır. Ancak bu testlerin hiç biri tüm alternatiflere karşı güçlü deęildir.

Tek deęişkenli W testi, kendisini de içerisinde barındıran birleşik bir test uygulandığında yani içerisinde W testi de dahil olmak üzere birden fazla normalliği test eden yeni bir normallik testi öne sürüldüğünde matematiksel hesaplamaları basit kılmaktadır. Bu nedenle birleşik testlerde kullanılması kolay bir testtir [6].

Tek deęişkenli R test istatistięi olasılık grafięi ve korelasyon katsayısının birleşiminden oluştuęu için kavramsal olarak anlaşılması kolaydır. Hesaplamalar açısından ise basittir çünkü istatistikte hesaplanan katsayılar herhangi bir yerde tutulmadan doğrudan hesaplanabilmektedir. Buna ek olarak R istatistięi normal olmayan hipotezler için; dağılım test istatistięi olarak genişletilebilecek bir yapıya sahiptir [3].

Sürücü [8]' nün öne sürmüş olduęu C test istatistięi W ve R istatistiklerinin ağırlıklandırılmış toplamından elde edilmektedir. Tüm tek deęişkenli normal dağılım testleri içerisinde W testi, çarpık ve kısa kuyruklu simetrik alternatifler için, R ise uzun kuyruklu simetrik alternatifler için en güçlü test olduğuna göre ve C testi bu testlerin ağırlaklandırılmış toplamından elde edildięine göre tek deęişkenli C istatistięi tüm alternatifleri kapsayan çok güçlü bir normallik testidir.

Buna ek olarak deneysel dağılım fonksiyonlarına baęlı Kolmogorov-Smirnov, Anderson-Darling ve bunun gibi tek deęişkenli normallik testlerinin W ve C istatistikleri kadar güçlü olmadığı ifade edilmektedir [8].

Sürücü [8] tarafından yapılan simülasyon sonuçlarına göre tek değişkenli C istatistiğinin çok değişkenli genellemesi ile oluşturulmuş olan C_p testinin çok değişkenli basıklık, çarpıklık katsayılarına, çok değişkenli R , W ve Z testlerine göre en güçlü test olduğu görülmektedir.

Çok değişkenli normal dağılımın test edilmesinde kullanılan yöntemlerden bir tanesi de Q-Q grafikleridir. Liang ve Yang [22]'in Fisher' in veri kümesi üzerinde yapmış oldukları çalışma sonuçlarına göre Q-Q grafikleri her ne kadar çok değişkenli bir veri kümesinin dağılımı hakkında görsel bilgi verse bile değişken sayısı örneklem sayısına yakın olduğu durumlarda tutarlı sonuçlar vermemektedir. Çok değişkenli C_p testi için böyle bir kısıt bulunmamaktadır.

Basıklık ve çarpıklık katsayıları çok değişkenli normalliğin testinde her ne kadar yaygın olarak kullanılsa da Sürücü [9]' nün Fisher' in verisi üzerinde yapmış olduğu Simülasyon sonuçlarına göre normal dağılmadığı bilinen bu veri kümesi için çarpıklık testi çok değişkenli normal dağılımdan herhangi bir sapma olmadığı sonucuna ulaşmıştır. Aynı şekilde verilerin çok değişkenli normal dağılıma sahip olduğu sonucu W_p testi için de geçerlidir. Bu bilgiler doğrultusunda normal dağılmayan bir veri kümesinin normal dağılıma sahip olduğu sonuçlarını veren bu iki testin (çok değişkenli W_p testi ve çok değişkenli çarpıklık katsayısı) güçlü testler olmadığı sonucuna ulaşılmaktadır.

Fisher' in süsen çiçekleri üzerinden toplamış olduğu verilerin Setosa grubu için X_3 ve X_4 değişkenlerinin çok değişkenli normal dağılmadığı bilinmektedir. Bu çalışmada JAVA dilinde geliştirilmiş olan programın sonuçları ile Sürücü [9]' nün çalışmasındaki bu sonuçlar tutarlıdır.

(5.4) bölümünde incelenmiş olan doğum veri kümesinden alınan sonuçlara bakıldığında bu verinin çok değişkenli normal dağılıma sahip olduğu görülmektedir. Alpar [23] grafik yöntemi ile yapmış olduğu testin bu veri kümesinin normal dağılıma sahip olmadığı sonucuna ulaşırken, grafiksel yöntemin C_p yöntemine göre daha az güvenilir olması bilindiğinden çalışmada bulunmuş olan sonuçların doğruluğu düşünülmektedir.

Her ne kadar R paket programı Shapiro-Wilk'in çok deęişkenli W_p testini barındıran bir program olsa da yapılmış olan çalışmalar sonucunda çok deęişkenli W_p testinin güçlü bir test olmadığı bilinmektedir. Bu çalışma, çok deęişkenli C_p ve Z_p testlerinin JAVA programlama dilinde uygulamasını sunarken zaman ve pratiklik açısından kullanıcılara bir çok kolaylık sağlamaktadır.

KAYNAKLAR LİSTESİ

- [1] SOKAL, Robert R. and ROHLF, F. James, Biometry, Third Edition, W. H Freeman and Company: New York, 2003.
- [2] LARSEN, Richard J. and MARX L. Morris, An introduction to mathematical statistics and its applications, Third Edition, Prentice Hall: New Jersey, 2001.
- [3] FILLIBEN, J. J., The probability plot correlation coefficient test for normality, Technometrics 17:111-117, 1975.
- [4] KENDALL, M. and STUART, A., The advanced theory of statistics, Vol. 2., London: Charles Griffin, 1979.
- [5] CONOVER, W. J., Practical nonparametric statistics, Second Edition, Toronto Singapore: John Wiley and Sons, 1980.
- [6] ROYSTON, J. P., Some techniques for assessing multivariate normality based on the Shapiro-Wilk W, Appl. Statist., 32:2, 121-133, 1983.
- [7] MALKOVICH, J. F., On tests for multivariate normality, Journal of the American Statistical Association, 68: 176-179, 1973.
- [8] SÜRÜCÜ, Barış, A power comparison and simulation study for goodness-of-fit test, Computers and Mathematics with Applications, 56, 1617-1625.
- [9] SÜRÜCÜ, Barış, Goodness-of-Fit tests for multivariate distributions, Communications in Statistics – Theory and Methods, 35: 1319-1331, 2006.
- [10] TIKU, M. L., Goodness-of-fit statistics based on the spacing of complete or censored samples, Australian Journal of Statistics, 22:260-275, 1980.
- [11] TIKU, M. L., Order statistics in goodness-of-fit tests, Commun. Statist. – Theory and Methods, 17:2369-2387, 1988.

- [12] Internet: Wikipedia The Free Encyclopedia, 2008,
[http://en.wikipedia.org/wiki/ Location-scale_family](http://en.wikipedia.org/wiki/Location-scale_family)
- [13] JOHNSON, R.A. and WICHERN, D.W., Applied multivariate statistical analysis, Prentice- Hall, Upper Saddle River, New Jersey, 149–194, 2002.
- [14] MARDIA, Kanti V., KENT, J. T., and BIBBY J. M., Multivariate analysis, Academic Press, London, 30–60, 1989.
- [15] ROHLF, F. James and SOKAL, Robert R., Statistical tables, Third Edition, W. H Freeman and Company: New York, 1995.
- [16] MARDIA, Kanti V., Measures of multivariate skewness and kurtosis with applications, *Biometrika*, 57, 3, 519.
- [17] ROSS, Sheldon, M., Simulation, Second Edition, Academic Press: USA California, 1997.
- [18] GNANADESIKAN, R., Methods for statistical data analysis of multivariate observations, Second Edition, Wiley Inter-Science: New Jersey, 1997.
- [19] Internet: Package JAMA, 2008,
<http://math.nist.gov/JAVAnumerics/jama/doc/>
- [20] FAN, X., FELSÖVÁLYI, Á. SIVO, A. S., Keenan, S. C., SAS for monte carlo studies a guide for quantitative researchers, SAS Institute Inc., NC, USA, 2001.
- [21] MARDIA Kanti V. and KENT, J. T., Rao score tests for goodness of fit and independence, *Biometrika*, 78, 2, 355-363, 1991
- [22] LIANG, Jiajuan, PAN, William S. Y. and YANG, Zhen-Hai, Characterization-based Q-Q plots for testing multinormality, *Statistics and Probability Letters*, 70, 183-190, 2004.

- [23] ALPAR, Reha, Uygulamalı çok deęişkenli istatistiksel yöntemlere giriş 1, 2. Baskı, Nobel Yayın Dağıtım Ltd. Şti., 2003.
- [24] HARTER, H. Leon, Expected values of normal order statistics, Biometrika, 48, No.1/2, 151-165, 1961.

EKLER

Ek 1. Shapiro Wilk $a_{i:n}$ katsayıları tablosu

$i \setminus n$	2	3	4	5	6	7	8	9	10
1	0,7071	0,7071	0,6872	0,6646	0,6431	0,6233	0,6052	0,5888	0,5739
2		0	0,1667	0,2413	0,2806	0,3031	0,3164	0,3244	0,3291
3				0	0,0875	0,1401	0,1743	0,1976	0,2141
4						0	0,0561	0,0947	0,1224
5								0	0,0399

$i \setminus n$	11	12	13	14	15	16	17	18	19	20
1	0,5601	0,5475	0,5359	0,5251	0,515	0,5056	0,4968	0,4886	0,4808	0,4734
2	0,3315	0,3325	0,3325	0,3318	0,3306	0,329	0,3273	0,3253	0,3232	0,3211
3	0,226	0,2347	0,2412	0,246	0,2495	0,2521	0,254	0,2553	0,2561	0,2565
4	0,1429	0,1586	0,1707	0,1802	0,1878	0,1939	0,1988	0,2027	0,2059	0,2085
5	0,0695	0,0922	0,1099	0,124	0,1353	0,1447	0,1524	0,1587	0,1641	0,1686
6	0	0,0303	0,0539	0,0727	0,088	0,1005	0,1109	0,1197	0,1271	0,1334
7			0	0,024	0,0433	0,0593	0,0725	0,0837	0,0932	0,1013
8					0	0,0196	0,0359	0,0496	0,0612	0,0711
9							0	0,0163	0,0303	0,0422
10									0	0,014

$i \setminus n$	21	22	23	24	25	26	27	28	29	30
1	0,4643	0,459	0,4542	0,4493	0,445	0,4407	0,4366	0,4328	0,4291	0,4254
2	0,3185	0,3156	0,3126	0,3098	0,3069	0,3043	0,3018	0,2992	0,2968	0,2944
3	0,2578	0,2571	0,2563	0,2554	0,2543	0,2533	0,2522	0,251	0,2499	0,2487
4	0,2119	0,2131	0,2139	0,2145	0,2148	0,2151	0,2152	0,2151	0,215	0,2148
5	0,1736	0,1764	0,1787	0,1807	0,1822	0,1836	0,1848	0,1857	0,1864	0,187
6	0,1399	0,1443	0,148	0,1512	0,1539	0,1563	0,1584	0,1601	0,1616	0,163
7	0,1092	0,115	0,1201	0,1245	0,1283	0,1316	0,1346	0,1372	0,1395	0,1415
8	0,0804	0,0878	0,0941	0,0997	0,1046	0,1089	0,1128	0,1162	0,1192	0,1219
9	0,053	0,0618	0,0696	0,0764	0,0823	0,0876	0,0923	0,0965	0,1002	0,1036
10	0,0263	0,0368	0,0459	0,0539	0,061	0,0672	0,0728	0,0778	0,0822	0,0862
11	0	0,0122	0,0228	0,0321	0,0403	0,0476	0,054	0,0598	0,065	0,0697
12			0	0,0107	0,02	0,0284	0,0358	0,0424	0,0483	0,0537
13					0	0,0094	0,0178	0,0253	0,032	0,0381
14							0	0,0084	0,0159	0,0227
15									0	0,0076

$i \setminus n$	31	32	33	34	35	36	37	38	39	40
1	0,422	0,4188	0,4156	0,4127	0,4096	0,4068	0,404	0,4015	0,3989	0,3964
2	0,2921	0,2898	0,2876	0,2854	0,2834	0,2813	0,2794	0,2774	0,2755	0,2737
3	0,2475	0,2462	0,2451	0,2439	0,2427	0,2415	0,2403	0,2391	0,238	0,2368
4	0,2145	0,2141	0,2137	0,2132	0,2127	0,2121	0,2116	0,211	0,2104	0,2098
5	0,1874	0,1878	0,188	0,1882	0,1883	0,1883	0,1883	0,1881	0,188	0,1878
6	0,1641	0,1651	0,166	0,1667	0,1673	0,1678	0,1683	0,1686	0,1689	0,1691
7	0,1433	0,1449	0,1463	0,1475	0,1487	0,1496	0,1505	0,1513	0,152	0,1526
8	0,1243	0,1265	0,1284	0,1301	0,1317	0,1331	0,1344	0,1356	0,1336	0,1376
9	0,1066	0,1093	0,1118	0,114	0,116	0,1179	0,1196	0,1211	0,1225	0,1237
10	0,0899	0,0931	0,0961	0,0988	0,1013	0,1036	0,1056	0,1075	0,1092	0,1108

11	0,0739	0,0777	0,0812	0,0844	0,0873	0,09	0,0924	0,0947	0,0967	0,0986
12	0,0585	0,0629	0,0669	0,0706	0,0739	0,077	0,0798	0,0824	0,0848	0,087
13	0,0435	0,0485	0,053	0,0572	0,061	0,0645	0,0677	0,0706	0,0733	0,0759
14	0,0289	0,0344	0,0395	0,0441	0,0484	0,0523	0,0559	0,0592	0,0622	0,0651
15	0,0144	0,0206	0,0262	0,0314	0,0361	0,0404	0,0444	0,0481	0,0515	0,0546
16	0	0,0068	0,0131	0,0187	0,0239	0,0287	0,0331	0,0372	0,0409	0,0444
17			0	0,0062	0,0119	0,0172	0,022	0,0264	0,0305	0,0343
18					0	0,0057	0,011	0,0158	0,0203	0,0244
19							0	0,0053	0,0101	0,0146
20									0	0,0049

i \ n	41	42	43	44	45	46	47	48	49	50
1	0,394	0,3917	0,3894	0,3872	0,385	0,383	0,3808	0,3789	0,377	0,3751
2	0,2719	0,2701	0,2684	0,2667	0,2651	0,2635	0,262	0,2604	0,2589	0,2574
3	0,2357	0,2345	0,2334	0,2323	0,2313	0,2302	0,2291	0,2281	0,2271	0,226
4	0,2091	0,2085	0,2078	0,2072	0,2065	0,2058	0,2052	0,2045	0,2038	0,2032
5	0,1876	0,1874	0,1871	0,1868	0,1865	0,1862	0,1859	0,1855	0,1851	0,1847
6	0,1693	0,1694	0,1695	0,1695	0,1695	0,1695	0,1695	0,1693	0,1692	0,1691
7	0,1531	0,1535	0,1539	0,1542	0,1545	0,1548	0,155	0,1551	0,1553	0,1554
8	0,1384	0,1392	0,1398	0,1405	0,141	0,1415	0,142	0,1423	0,1427	0,143
9	0,1249	0,1259	0,1269	0,1278	0,1286	0,1293	0,13	0,1306	0,1312	0,1317
10	0,1123	0,1136	0,1149	0,116	0,117	0,118	0,1189	0,1197	0,1205	0,1212
11	0,1004	0,102	0,1035	0,1049	0,1062	0,1073	0,1085	0,1095	0,1105	0,1113
12	0,0891	0,0909	0,0927	0,0943	0,0959	0,0972	0,0986	0,0998	0,101	0,102
13	0,0782	0,0804	0,0824	0,0842	0,086	0,0876	0,0892	0,0906	0,0919	0,0932
14	0,0677	0,0701	0,0724	0,0745	0,0765	0,0783	0,0801	0,0817	0,0832	0,0846
15	0,0575	0,0602	0,0628	0,0651	0,0673	0,0694	0,0713	0,0731	0,0748	0,0764
16	0,0476	0,0506	0,0534	0,056	0,0584	0,0607	0,0628	0,0648	0,0667	0,0685
17	0,0379	0,0411	0,0442	0,0471	0,0497	0,0522	0,0546	0,0568	0,0588	0,0608
18	0,0283	0,0318	0,0352	0,0383	0,0412	0,0439	0,0465	0,0489	0,0511	0,0532
19	0,0188	0,0227	0,0263	0,0296	0,0328	0,0357	0,0385	0,0411	0,0436	0,0459
20	0,0094	0,0136	0,0175	0,0211	0,0245	0,0277	0,0307	0,0335	0,0361	0,0386
21	0	0,0045	0,0087	0,0126	0,0163	0,0197	0,0229	0,0259	0,0288	0,0314
22			0	0,0042	0,0081	0,0118	0,0153	0,0185	0,0215	0,0244
23					0	0,0039	0,0076	0,0111	0,0143	0,0174
24							0	0,0037	0,0071	0,0104
25									0	0,0035

Ek 2. Fisher setosa grubu için süsen çiçekleri veri kümesi

Çanak Yaprak Uzunluğu (X_1)	Çanak Yaprak Genişliği (X_2)	Taç Yaprak Uzunluğu (X_3)	Taç Yaprak Genişliği (X_4)	Tür
5.1	3.5	1.4	0.2	<i>setosa</i>
4.9	3.0	1.4	0.2	<i>setosa</i>
4.7	3.2	1.3	0.2	<i>setosa</i>
4.6	3.1	1.5	0.2	<i>setosa</i>
5.0	3.6	1.4	0.2	<i>setosa</i>
5.4	3.9	1.7	0.4	<i>setosa</i>
4.6	3.4	1.4	0.3	<i>setosa</i>
5.0	3.4	1.5	0.2	<i>setosa</i>
4.4	2.9	1.4	0.2	<i>setosa</i>
4.9	3.1	1.5	0.1	<i>setosa</i>
5.4	3.7	1.5	0.2	<i>setosa</i>
4.8	3.4	1.6	0.2	<i>setosa</i>
4.8	3.0	1.4	0.1	<i>setosa</i>
4.3	3.0	1.1	0.1	<i>setosa</i>
5.8	4.0	1.2	0.2	<i>setosa</i>
5.7	4.4	1.5	0.4	<i>setosa</i>
5.4	3.9	1.3	0.4	<i>setosa</i>
5.1	3.5	1.4	0.3	<i>setosa</i>
5.7	3.8	1.7	0.3	<i>setosa</i>
5.1	3.8	1.5	0.3	<i>setosa</i>
5.4	3.4	1.7	0.2	<i>setosa</i>
5.1	3.7	1.5	0.4	<i>setosa</i>
4.6	3.6	1.0	0.2	<i>setosa</i>
5.1	3.3	1.7	0.5	<i>setosa</i>
4.8	3.4	1.9	0.2	<i>setosa</i>
5.0	3.0	1.6	0.2	<i>setosa</i>
5.0	3.4	1.6	0.4	<i>setosa</i>
5.2	3.5	1.5	0.2	<i>setosa</i>
5.2	3.4	1.4	0.2	<i>setosa</i>
4.7	3.2	1.6	0.2	<i>setosa</i>

4.8	3.1	1.6	0.2	<i>setosa</i>
5.4	3.4	1.5	0.4	<i>setosa</i>
5.2	4.1	1.5	0.1	<i>setosa</i>
5.5	4.2	1.4	0.2	<i>setosa</i>
4.9	3.1	1.5	0.2	<i>setosa</i>
5.0	3.2	1.2	0.2	<i>setosa</i>
5.5	3.5	1.3	0.2	<i>setosa</i>
4.9	3.6	1.4	0.1	<i>setosa</i>
4.4	3.0	1.3	0.2	<i>setosa</i>
5.1	3.4	1.5	0.2	<i>setosa</i>
5.0	3.5	1.3	0.3	<i>setosa</i>
4.5	2.3	1.3	0.3	<i>setosa</i>
4.4	3.2	1.3	0.2	<i>setosa</i>
5.0	3.5	1.6	0.6	<i>setosa</i>
5.1	3.8	1.9	0.4	<i>setosa</i>
4.8	3.0	1.4	0.3	<i>setosa</i>
5.1	3.8	1.6	0.2	<i>setosa</i>
4.6	3.2	1.4	0.2	<i>setosa</i>
5.3	3.7	1.5	0.2	<i>setosa</i>
5.0	3.3	1.4	0.2	<i>setosa</i>

Ek 3. Süsen çiçekleri veri kümesi için oluşturulmuş olan Y matrisi ve kısmi regresyon katsayıları (p=4)

Y_1	Y_2	Y_3	Y_4
5.1	7.563096892875645	1.8915972771732736	0.7359154658730326
4.9	6.9037597598217	1.882712917627565	0.71698711255481
4.7	6.944422626767753	1.753771014716409	0.6950352096941148
4.6	6.7647540602407785	1.9450307899366728	0.7223722814450324
5.0	7.583428326348672	1.8771263257176958	0.7333957745453918
5.4	8.202102592456567	2.2149525881745613	1.0081010711351368
4.6	7.064754060240778	1.8364346999229095	0.811412177238022
5.0	7.383428326348673	1.9828570523935376	0.7463400174185366
4.4	6.405416927186832	1.8275503403772007	0.6924838239197995
4.9	7.003759759821699	1.979847554289644	0.6358837714263581
5.4	8.002102592456566	2.020683314850403	0.7703077533920408
4.8	7.224091193294726	2.059645876158224	0.7542448776363997
4.8	6.824091193294725	1.8711073295099079	0.6124832825610347
4.3	6.425748360659858	1.5130793889216225	0.5392265719759177
5.8	8.62077685856446	1.758509577307268	0.743537928749304
5.7	8.941108292037487	2.035442535837927	0.9977082140363075
5.4	8.202102592456567	1.8149525881745612	0.9404509903134823
5.1	7.563096892875645	1.8915972771732736	0.8359154658730326
5.7	8.341108292037486	2.252634715865453	0.9196284224503282
5.1	7.863096892875645	1.9830011871595108	0.8587804020768495
5.4	7.702102592456567	2.229279404864166	0.7981803778044647
5.1	7.763096892875645	1.9858665504974318	0.9567962634107151
4.6	7.2647540602407785	1.4307039732470677	0.6477303737486364
5.1	7.363096892875645	2.197328003849116	1.0826847491570044
4.8	7.224091193294726	2.3596458761582237	0.8049824382526406
5.0	6.9834283263486725	2.0943185057452225	0.7553159829594125
5.0	7.383428326348673	2.082857052393538	0.9632525376239502
5.2	7.64276545940262	2.003202865290931	0.7573318160722214
5.2	7.54276545940262	1.9060682286288517	0.7384351572006734
4.7	6.944422626767753	2.0537710147164088	0.7457727703103556

4.8	6.924091193294725	2.0682419661719873	0.7482924616379965
5.4	7.702102592456567	2.029279404864166	0.9643553373936375
5.2	8.24276545940262	1.9860106852634047	0.669236648069028
5.5	8.581771158983539	1.9179620862784548	0.7678197565110745
4.9	7.003759759821699	1.979847554289644	0.7358837714263581
5.0	7.183428326348673	1.68858777906938	0.6916341794700268
5.5	7.88177115898354	1.8380196296439022	0.73701826564272
4.9	7.503759759821699	1.865520737600039	0.6288919445516166
4.4	6.5054169271868325	1.7246849770392798	0.6775554423805202
5.1	7.463096892875646	1.9944626405111947	0.7508438474123118
5.0	7.4834283263486725	1.7799916890556169	0.8144991156738437
4.5	5.885085493713805	1.756348108522384	0.7681703017113544
4.4	6.705416927186832	1.7189542503634379	0.681523719712789
5.0	7.4834283263486725	2.079991689055617	1.1652366762900845
5.1	7.863096892875645	2.3830011871595107	1.0264304828985038
4.8	6.824091193294725	1.8711073295099079	0.8124832825610347
5.1	7.863096892875645	2.083001187159511	0.775692922282263
4.6	6.864754060240779	1.8421654265987517	0.7074438999057531
5.3	7.922434025929593	2.009077726732746	0.7658039233982656
5.0	7.283428326348672	1.885722415731459	0.7274433585469885

Kısmi Regresyon Katsayıları

$\beta_{p(p-1).q(p-1)}$	Hesaplanan değer
$\beta_{41.23}$	0.0450382999377523
$\beta_{42.13}$	0.0198413866613442
$\beta_{43.12}$	0.1691252020541363
$\beta_{32.1}$	0.0286536333792102
$\beta_{31.2}$	0.1160558811765705
$\beta_{2.1}$	0.7966856652697345

Ek 4. Süsen çiçekleri veri kümesi için oluşturulmuş olan Y matrisi ve kısmi regresyon katsayıları ($p = 3$)

Y_1	Y_2	Y_3
5.1	7.60514320937414	2.08341797601786
4.9	6.9441572011633905	2.0642818202032136
4.7	6.983171192952639	1.9303551560727767
4.6	6.802678188847263	2.117617683526355
5.0	7.624650205268765	2.0664546439526417
5.4	8.246622221690266	2.4195174638977255
4.6	7.102678188847263	2.0112788942481594
5.0	7.424650205268765	2.170680503471439
4.4	6.4416921806365135	1.9921427384335129
4.9	7.04415720116339	2.162168890443815
5.4	8.046622221690267	2.2237433234165227
4.8	7.263664197058015	2.240979698859799
4.8	6.863664197058014	2.0494314178973934
4.3	6.461199176531138	1.6751794063682945
5.8	8.668594238111767	1.976806143361606
5.7	8.988101234006393	2.2535040220181917
5.4	8.246622221690266	2.0195174638977256
5.1	7.60514320937414	2.08341797601786
5.7	8.388101234006392	2.4661816005745836
5.1	7.90514320937414	2.1770791867396646
5.4	7.7466222216902665	2.4300821126947185
5.1	7.80514320937414	2.1791921164990633
4.6	7.302678188847263	1.6070530347293623
5.1	7.40514320937414	2.3876438355366574
4.8	7.263664197058015	2.540979698859799
5.0	7.024650205268765	2.2791322225090336
5.0	7.424650205268765	2.2706805034714392
5.2	7.6856362134795155	2.19826837832368

5.2	7.585636213479516	2.1003813080830787
4.7	6.983171192952639	2.230355156072777
4.8	6.963664197058014	2.247318488137995
5.4	7.7466222216902665	2.2300821126947183
5.2	8.285636213479515	2.1855907997672883
5.5	8.627115225795642	2.1280290769253494
4.9	7.04415720116339	2.162168890443815
5.0	7.224650205268765	1.874906362990236
5.5	7.927115225795641	2.0428195852411393
4.9	7.54415720116339	2.051604241646822
4.4	6.541692180636513	1.8900298086741145
5.1	7.50514320937414	2.185530905777259
5.0	7.524650205268765	1.9685675737120403
4.5	5.922185184741888	1.9196707192957245
4.4	6.741692180636514	1.8858039491553173
5.0	7.524650205268765	2.2685675737120405
5.1	7.90514320937414	2.5770791867396645
4.8	6.863664197058014	2.0494314178973934
5.1	7.90514320937414	2.2770791867396647
4.6	6.902678188847263	2.015504753766957
5.3	7.966129217584891	2.208892921110703
5.0	7.324650205268765	2.0727934332308378

Kısmi Regresyon Katsayıları

$\beta_{p(p-1).q(p-1)}$	Hesaplanan değer
$\beta_{32.1}$	-0.021129297593986
$\beta_{31.2}$	0.148504023058198
$\beta_{2.1}$	0.804930041053753

Ek 5. Johnson ve Wichern'in ter veri kümesi

X_1	X_2	X_3
3.7	48.5	9.3
5.7	65.1	8.0
3.8	47.2	10.9
3.2	53.2	12.0
3.1	55.5	9.7
4.6	36.1	7.9
2.4	24.8	14.0
7.2	33.1	7.6
6.7	47.4	8.5
5.4	54.1	11.3
3.9	36.9	12.7
4.5	58.8	12.3
3.5	27.8	9.8
4.5	40.2	8.4
1.5	13.5	10.1
8.5	56.4	7.1
4.5	71.6	8.2
6.5	52.8	10.9
4.1	44.1	11.2
5.5	40.9	9.4

Ek 6. Ter veri kümesi için oluşturulmuş olan Y matrisi ($p = 3$)

Y_1	Y_2	Y_3
3.7	61.968618611763304	7.115562163035438
5.7	85.84895299650022	4.596953226675181
3.8	61.03263533100015	8.6462519567462
3.2	64.84853501557907	10.15503170760453
3.1	66.78451829634223	7.92827607326993
4.6	52.84476908489492	5.089016394868993
2.4	33.536401261684304	12.556867974123358
7.2	59.309203785052915	3.1081231401346074
6.7	71.78912018886868	4.385360614714853
5.4	73.75690283878969	8.046264870838085
3.9	51.096652050237	10.34153431607151
4.5	75.18075236565807	9.642517611808087
3.5	40.540585173289614	7.662516662149349
4.5	56.58075236565807	5.669342247411483
1.5	18.960250788552692	9.190174165074776
8.5	87.34142113513191	1.8652436652962217
4.5	87.98075236565806	5.592874851822954
6.5	76.46108675039498	6.934996673546579
4.1	59.02468548871069	8.74146866537942
5.5	60.920919558026526	6.030138167972526

Ek 7. Yüzücü veri kümesi

X_1	X_2	X_3
62.0	48.0	68.0
68.0	42.0	75.0
57.0	38.0	79.0
56.0	40.0	78.0
74.0	44.0	83.0
62.0	48.0	66.0
62.0	41.0	80.0
52.0	33.0	75.0
63.0	36.0	75.0
73.0	43.0	73.0
65.0	43.0	74.0
55.0	37.0	65.0
57.0	47.0	62.0
60.0	32.0	64.0
63.0	45.0	70.0
61.0	48.0	74.0
67.0	53.0	76.0
67.0	46.0	82.0
62.0	42.0	76.0
63.0	43.0	84.0
55.0	37.0	74.0
58.0	42.0	73.0
66.0	43.0	74.0
58.0	36.0	67.0
64.0	42.0	75.0

Ek 8. Yüzücü veri kümesi için oluşturulmuş olan Y matrisi ($p = 3$)

Y_1	Y_2	Y_3
62.0	78.20448632653466	90.08883388325458
68.0	75.12750113232835	100.6895239392031
57.0	65.76864065503995	100.14984474956901
56.0	67.281471520741	98.41229067938171
74.0	80.05051593812203	111.19070150825124
62.0	78.20448632653466	88.08883388325458
62.0	71.20448632653466	103.0509073092924
52.0	58.332794983545206	94.52366031157071
63.0	66.69165546083362	99.20077856206734
73.0	78.56334680382308	100.8654646206516
65.0	74.6659937294315	98.16405703295393
55.0	63.79430238644204	85.36193191350715
57.0	74.76864065503995	81.9128932018061
60.0	61.23014805793677	87.3625069601309
62.0	75.20448632653466	92.50115106584222
61.0	77.71731719223573	95.62615793479237
67.0	85.64033199802941	99.71501832125288
67.0	78.64033199802941	106.6770917472907
62.0	72.20448632653466	98.91346824842987
63.0	73.69165546083362	107.23870513602952
55.0	63.79430238644204	94.36193191350715
58.0	70.25580978933888	94.06276445458104
66.0	75.15316286373044	98.62673298141614
58.0	64.25580978933888	88.88739881975631
64.0	73.17882459513257	98.83882014535428

Ek 9. Alpar doğum veri kümesi

X_1	X_3
13.0	20.0
14.0	25.0
12.5	40.0
12.0	22.0
12.5	33.0
12.0	35.0
11.0	21.0
10.0	25.0
10.0	42.0
12.0	30.0
10.5	35.0
10.0	28.0
11.0	25.0
9.0	40.0
11.5	33.0

Ek 10. Doğum veri kümesi için oluşturulmuş olan Y matrisi ($p = 2$)

Y_1	Y_2
13.0	-3.4101562499999996
14.0	-0.21093749999999996
12.5	17.490234375000000
12.0	0.39062500000000035
12.5	10.4902343750000004
12.0	13.3906250000000004
11.0	1.1914062500000003
10.0	6.9921875000000003
10.0	23.9921875000000002
12.0	8.3906250000000004
10.5	16.091796875000004
10.0	9.9921875000000004
11.0	5.1914062500000003
9.0	23.7929687500000004
11.5	12.291015625000000

Ek 11. Vücut yağ yüzdeleri veri kümesi

X_1	X_2	X_3	X_4
12.8	15.3	13.8	12.3
22.0	24.0	20.5	18.1
17.8	19.0	16.5	15.6
15.4	16.6	15.8	10.9
15.7	19.9	16.2	14.5
15.5	17.1	17.1	14.4
16.7	19.4	16.8	12.7
18.9	19.9	18.7	16.9
15.1	17.5	16.0	13.8
16.7	17.1	15.9	15.2
18.0	20.1	19.1	16.3
17.4	18.4	18.4	15.9
16.8	18.7	16.8	14.9
16.9	18.8	17.0	14.0
14.7	17.3	14.1	12.4
11.9	13.5	12.0	9.8
13.0	14.4	14.2	12.8
17.6	18.3	16.3	16.0
16.5	17.5	15.9	13.2
14.8	18.0	16.1	14.2
15.2	16.3	15.2	12.9
20.4	20.6	18.6	16.5
15.3	18.6	14.9	13.5
22.3	23.3	21.0	17.1
15.2	17.0	17.0	13.8
21.8	22.8	20.8	17.6

Ek 12. Vücut yağ yüzdeleri için oluşturulmuş olan Y matrisi (p = 4)

Y_1	Y_2	Y_3	Y_4
12.8	23.69570622257234	21.950042941353402	22.784731412173326
22.0	38.430120070046215	34.09568407729911	34.66556897911309
17.8	30.67527896576466	27.42509656299472	28.95234999973918
15.4	26.70108404903235	25.28108641519546	23.111006131475392
15.7	30.197858413623884	26.399974267379157	27.085754701895205
15.5	27.266675503896195	26.713038382514934	27.216478062992902
16.7	30.353772962262347	27.33245187044112	25.844451852718226
18.9	32.296784969266966	30.251042108488797	31.574567349699556
15.1	27.404309684440808	25.515939780204068	26.030837848769842
16.7	28.05377296226235	26.019688822931048	27.88338767527111
18.0	31.906461875492354	30.306946717090632	30.86656401912731
17.4	29.81291314630927	29.048535766114256	29.920176826403882
16.8	29.719364417126197	27.24904920427708	28.053342179416955
16.9	29.884955871990044	27.509216293768716	27.274138330568466
14.7	26.941943864985422	23.411163861151444	23.682598401916447
11.9	21.305383128797722	19.447023160816038	19.168134644631785
13.0	22.926889132300033	22.272968706310113	23.49035199459216
17.6	29.844096056036967	27.015031286726654	29.18015148929855
16.5	28.322590052534657	26.007031960689535	25.83290545834258
14.8	27.70753531984927	25.579008267384843	26.3947166538379
15.2	26.269901139304654	24.542806016755225	24.780399099191342
20.4	33.980656792224664	30.909978695207663	31.631307474914532
15.3	28.635492594168504	24.697789934299976	25.358777256041407
22.3	37.92689443463775	34.59672315120442	33.95238358222259
15.2	26.969901139304653	26.468429552953946	26.474991356314398
21.8	37.09893716031852	34.09588770374624	34.191381120976736