

**T.C.
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**KOVARYANS MATRİSİNİN ÖZDEĞERLERİNİN ARALIK
TAHMİNLERİNİN SİMÜLASYONLA BELİRLENMESİ**

MELTEM ÜLKER

**YÜKSEK LİSANS TEZİ
İSTATİSTİK ANA BİLİM DALI**

Konya, 2009

ÖZET

Yüksek Lisans Tezi

KOVARYANS MATRİSİNİN ÖZDEĞERLERİNİN ARALIK TAHMİNLERİNİN SİMÜLASYONLA BELİRLENMESİ

Meltem ÜLKER

**Selçuk Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Ana Bilim Dalı**

Danışman: Yrd.Doç. Dr. Mustafa SEMİZ

2009, 92 sayfa

Bu tezde örneklem kovaryans matrisinin tüm karakteristik kökleri için güven aralıkları araştırıldı, bu kökler için güven aralıkları elde edildi. Monte Carlo simülasyonlarının sonuçlarına dayanan başka bir güven aralığı öngörüldü. Bu güven aralıkları simülasyona dayandığı için yığın kovaryans matrisinin en büyük ve en küçük karakteristik kökleri arasındaki fark büyük olduğu zaman bu yeni güven aralıklarının diğerlerine nazaran daha dar ve daha kesin bir doğrulukta olduğu gösterildi.

Anahtar Kelimeler: Karakteristik Kökler, Kovaryans Matrisi, Güven Aralığı, Wishart Dağılımı, Jackknife Yöntemi, Bootstrap Yöntemi

ABSTRACT

Ms Thesis

THE DETERMINATION OF INTERVALS ESTIMATIONS OF THE EIGENVALUES OF COVARIANCE MATRICES

Meltem ÜLKER

Selcuk University

Graduate School of Natural and Applied Sciences

Department of Statistics

Supervisor: Ass. Prof. Dr. Mustafa SEMİZ

2009, 92 page

Confidence intervals for all of the characteristic roots of a sample covariance matrix are derived. Using a perturbation expansion, we obtain a new confidence interval for these roots. Then, we propose another confidence interval based on the results of Monte Carlo simulations. Since it is based on simulations, this new confidence interval is both narrower and more accurate than others when the difference between the largest and smallest characteristic roots of the population covariance matrix is large.

Key Words: Characteristic Root, Confidence Interval, Wishart Distribution, Covariance Matrix, Jackknife Method, Bootstrap Method

ÖNSÖZ

Kovaryans matrisinin özdeğerlerinin aralık tahminlerinin simülasyonla belirlenmesi adlı tezimin seçimi ve gerçekleşmesinde yardımını esirgemeyen, çalışmalarında yol gösteren ve bilgileriyle katkıda bulunan değerli hocam Yrd.Doç.Dr. Mustafa SEMİZ'e, destek olan diğer tüm bölüm hocalarıma, tez çalışmam boyunca desteğini esirgemeyen aileme teşekkür ederim.

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT.....	ii
ÖNSÖZ.....	iii
ŞEKİLLER DİZİNİ.....	vi
TABLolar DİZİNİ.....	viii
GİRİŞ.....	1
KAYNAK ARAŞTIRMASI.....	2
1. TEMEL KAVRAMLAR.....	5
1.1. Tek Değişkenli Dağılımlar.....	5
1.1.1. Normal dağılım.....	5
1.1.2. Gama dağılımı.....	6
1.1.3. Üstel dağılım.....	6
1.1.4. Ki-kare dağılımı.....	7
1.1.5. Merkezsel olmayan ki-kare dağılımı.....	7
1.2. Çok Değişkenli Dağılımlar.....	8
1.2.1. İki değişkenli normal dağılım.....	8
1.2.2. p değişkenli normal dağılım.....	8
1.3. Wishart Dağılımı.....	9
1.3.1. Wishart olasılık yoğunluk fonksiyonu.....	10
1.3.2. Merkezsel olmayan wishart dağılımı.....	13
1.3.3. Wishart matrislerinin toplamının dağılımı.....	14
1.4. Hipergeometrik Fonksiyon.....	15
1.5. Jackknife Yöntemi.....	16
1.6. Bootstrap Yöntemi.....	18
2. ÖZDEĞERLERİN KULLANILDIĞI TEMEL İSTATİSTİK YÖNTEMLER.....	21
2.1. Temel Bileşenler Analizi.....	21
2.1.1. Temel bileşenlerin elde edilmesi.....	22
2.1.2. Temel bileşenlerin özellikleri ve sağladığı yararlar.....	27
2.2. Faktör Analizi.....	28
3. ÇOK DEĞİŞKENLİ DAĞILIMLARIN GEOMETRİSİ.....	31
3.1. Verilerin Düzenlenmesi.....	31
3.1.1. Tanımlayıcı istatistikler.....	32
3.1.2. Pozitif tanımlı matrisler.....	34
3.2. Çok Değişkenli Dağılımın Geometrik Yorumu.....	37
3.2.1. Uzaklıklar.....	38
3.2.2. Bir noktanın orijine olan uzaklığı.....	43
3.2.3. Gözlemlerin dağılımının geometrisi.....	44
3.2.4. Ortalamaların dağılımının geometrisi.....	48
3.2.5. Kovaryans matrisinin özdeğerleri ile elemanları arasındaki fonksiyonel ilişki.....	50

4. KOVARYANS MATRİSİNİN KARAKTERİSTİK KÖKLERİ İÇİN GÜVEN ARALIĞI	53
4.1. En Küçük ve En Büyük Karakteristik Kökler için Güven Aralığı	54
5. UYGULAMALAR	65
5.1. Uygulama 1	65
5.2. Uygulama 2	68
5.3. Uygulama 3	69
5.4. Uygulama 4	74
5.5. Uygulama 5	79
5.6. Uygulama 6	81
5.7. Uygulama 7	83
5.8. Uygulama 8	86
SONUÇ	89
KAYNAKLAR	90

ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 2.1. Temel Bileşenler Dönüşümünün İki Boyutlu Uzayda Gösterimi	22
Şekil 3.1. Bir noktanın orijine olan en kısa uzaklığı	38
Şekil 3.2. X ve Y noktalarının birbirine olan istatistiksel uzaklığı	41
Şekil 3.3. Merkezi orijinde olan ve yarım kısa eksen ve yarım uzun eksen uzunlukları $c\sqrt{\lambda_i}$ olan elips	42
Şekil 3.4. Merkezi orijinde olan ve yarım kısa eksen ve yarım uzun eksen uzunlukları $c\sqrt{s_{ii}}$ olan elips	42
Şekil 3.5. Merkezi orijinde, e_i yönündeki yarı uzunluk $\frac{c}{\sqrt{\lambda_i}}$ olan elips	44
Şekil 3.6. Gözlemlerin dağılımı için %90 ve %95 düzeyinde elips konturlar	48
Şekil 3.7. n gözlem ortalamasının dağılımı ($n=42$)	50
Şekil 3.8. Merkezi orijinde olmayan elips	52
Şekil 5.1. $n=50$ ve 500 deneme için S_1 matrisinin özdeğerlerinin dağılım grafikleri	69
Şekil 5.2. $n=100$ ve 500 deneme için S_1 matrisinin özdeğerlerinin dağılım Grafikleri	70
Şekil 5.3. $n=300$ ve 500 deneme için S_1 matrisinin özdeğerlerinin dağılım grafikleri	71
Şekil 5.4. $n=500$ ve 500 deneme için S_1 matrisinin özdeğerlerinin dağılım Grafikleri	72
Şekil 5.5. $n=1000$ ve 1000 deneme için S_1 matrisinin özdeğerlerinin dağılım grafikleri	73
Şekil 5.6. $n=50$ ve 500 deneme için S_2 matrisinin özdeğerlerinin dağılım grafikleri	74

	<u>Sayfa</u>
Şekil 5.7. $n=100$ ve 500 deneme için S_2 matrisinin özdeğerlerinin dağılım grafikleri	75
Şekil 5.8. $n=300$ ve 500 deneme için S_2 matrisinin özdeğerlerinin dağılım grafikleri	76
Şekil 5.9. $n=500$ ve 500 deneme için S_2 matrisinin özdeğerlerinin dağılım grafikleri	77
Şekil 5.10. $n=1000$ ve 1000 deneme için S_2 matrisinin özdeğerlerinin dağılım grafikleri	78
Şekil 5.11. Farklı örnek hacimleri için özdeğerlerin güven aralıkları	80
Şekil 5.12. Farklı örnek hacimleri için özdeğerlerin güven aralıklar	82
Şekil 5.13. Farklı örnek hacimleri için en küçük ve en büyük karakteristik kök arasındaki farkın büyük olduğu matrisin özdeğerlerinin güven aralıkları	85
Şekil 5.14. Farklı örnek hacimleri için en küçük ve en büyük karakteristik kök arasındaki farkın küçük olduğu matrisin özdeğerlerinin güven aralıkları	88

TABLolar DİZİNİ

	<u>Sayfa</u>
Tablo 3.1. Mikrodalga fırınlardan, kapı kapalı ve açıkken alınan radyasyon verileri	45
Tablo 3.2. Noktaların geometrik uzaklıkları	47
Tablo 4.1. Durum 1 için güven seviyeleri	62
Tablo 4.2. Durum 2 için güven seviyeleri	62
Tablo 4.3. Durum 1 için gerçek güven seviyeleri	62
Tablo 4.4. Durum 2 için gerçek güven seviyeleri	63
Tablo 4.5. Yüksek boyutlu durumlar için gerçek güven seviyeleri	63
Tablo 4.6. Tahminin gerçek güven seviyeleri	63
Tablo 5.1. 10 tekrar ve $n=30$ için en küçük ve en büyük özdeğerler	66
Tablo 5.2. 20 tekrar için en küçük ve en büyük özdeğerler	67
Tablo 5.3. $n=50$ ve 500 deneme için S_1 matrisinin özdeğerlerinin istatistikleri	69
Tablo 5.4. $n=100$ ve 500 deneme için S_1 matrisinin özdeğerlerinin istatistikleri	70
Tablo 5.5. $n=300$ ve 500 deneme için S_1 matrisinin özdeğerlerinin istatistikleri	71
Tablo 5.6. $n=500$ ve 500 deneme için S_1 matrisinin özdeğerlerinin istatistikleri	72
Tablo 5.7. $n=1000$ ve 1000 deneme için S_1 matrisinin özdeğerlerinin istatistikleri	73
Tablo 5.8. $n=50, 500$ deneme sonuçlarının istatistikleri (diagonal olmayan matris)	74
Tablo 5.9. $n=100, 500$ deneme sonuçlarının istatistikleri (diagonal olmayan matris)	75
Tablo 5.10. $n=300, 500$ deneme sonuçlarının istatistikleri (diagonal olmayan matris)	76
Tablo 5.11. $n=500, 500$ deneme sonuçlarının istatistikleri (diagonal olmayan matris)	77

	<u>Sayfa</u>
Tablo 5.12. $n=1000$, 1000 deneme sonuçlarının istatistikleri (diagonal olmayan matris)	78
Tablo 5.13. Farklı örnek hacimleri (n) ve 500 deneme için diagonal olmayan matrisin özdeğerlerin güven aralıkları tahmini	79
Tablo 5.14. Farklı örnek hacimleri (n) ve 500 deneme için diyagonal olmayan matrisin özdeğerlerin güven aralıkları tahmini	81
Tablo 5.15. Farklı örnek hacimleri (n) ve 500 deneme için en küçük ve en büyük karakteristik kök arasındaki farkın büyük olduğu matrisin özdeğerlerin güven aralıkları tahmini	83
Tablo 5.16. Farklı örnek hacimleri (n) ve 500 deneme için en küçük ve en büyük karakteristik kök arasındaki farkın küçük olduğu matrisin özdeğerlerin güven aralıkları tahmini	86

GİRİŞ

Çok deęişkenli daęılımın yayılımını ifade eden kovaryans matrisinin özdeęerleri ve özvektörlerinin kullanılmasıyla elipsoitler belirlenebilir. Bu elipsoitler aynı zamanda kovaryans matrisinin elemanlarına ya da başka bir ifadeyle deęişkenlerin varyanslarına ve deęişkenler arasındaki kovaryanslara baęlıdır. Deęişkenlerin varyansının daęılımının ki-kare daęılımına sahip olduęu gibi örneklem kovaryans matrisinin de bir daęılımı vardır ve bu daęılıma Wishart daęılımı denir.

Bu çalışmada Wishart daęılımı yerine elipsoid konturların belirlenmesinde özdeęerlerin daęılımlarının kullanılabilceęini ve bu sebeple özdeęerlerin daęılımlarının belirlenmesi amaçlanmıştır.

Örneklem kovaryans matrisinin karakteristik kökleri , çok deęişkenli analiz, temel bileşenler analizi gibi bir çok analizde önemli rol oynar. Bir çok araştırmacı karakteristik köklerin asimptotik daęılımlarını normallik varsayımı altında düşünürler. Bunun yanında karakteristik kökler için güven aralıkları bulunmuş ve ilgili hipotez testleri birçok durumda kullanılmıştır. (Anderson (2003)).

Bu çalışmada, örneklem kovaryans matrisin tüm karakteristik kökleri için güven aralıkları incelendi. Monte Carlo simülasyonları ile, en küçük ve en büyük karakteristik kökler arasındaki farkın çok büyük olduęu ve bu kökler arasındaki farkın küçük olduęu durumlardaki güven aralıkları incelendi.

Bu çalışmada Wishart daęılımı olarak bilinen matris daęılımı incelenmiştir. Daęılım genellikle ki-kare daęılımının çok deęişkenli uzantısı olarak kabul edilir.

KAYNAK ARAŞTIRMASI

- Anderson, T.W.(2003) “*An Introduction to Multivariate Statistical Analysis*” adlı kitabın 7. bölümünde Wishart dağılımı konusu ele alınmıştır. Dağılımı ikiden fazla boyuta genelleştiren ve yoğunluk fonksiyonunu bulan John Wishart (1928)’dir. Orijinal türetme şeklini bulan Wishart’ın yaklaşımı geometriktir ve merkezci bir durumu ele almaktadır. Bugün artık merkezci olmayan dağılım türleri de saptanmıştır. Örneğin, James (1954, 1955 ve 1964), Constantine (1963) ve Anderson’un (1945 ve 1946) da bu alanındaki çalışmalarından söz edilmesi gerekir. Wishart dağılımı üzerine modern tartışmalarda Rao (1973) tarafından geliştirilen ve yoğunluk fonksiyonu tanımına dayanmayan bir yaklaşım uygulanmaktadır. Kovaryans matrislerinin dağılımını ifade eden Wishart dağılımı 1. bölümde ayrıntılı olarak bu kaynak kullanılarak verilmiştir.
- Anderson, G.A., (1965), “*Asymptotic Expansion for The Distribution of The Latent Roots of The Estimated Covariance Matrix*” ortogonal matrislerin karakteristik köklerinin dağılımları incelenmiştir.
- Erfon, B., Gong, G., (1983). “*Aleisure Look at The Bootstrap, The Jackknife and Cross-Validation*” adlı makalesinde Bootstrap ve Jackknife metotları incelenmiştir.
- Gradshteyn, I. S., Ryzhik, I.M., (1965), “*Table of Integrals, Series and Products*” adlı kitapta hipergeometrik fonksiyonlar konusu anlatılmıştır. Bu çalışmada 1. bölümde hipergeometrik fonksiyonlardan söz edilmiştir.
- Graybill, F.A., (1998), “*Theory and Application of The Linear Model*” adlı kitabın 4. bölümünde dağılımlar konusuna ilişkin bilgilere yer verilmiştir. Bu çalışmada 1. bölümde dağılımlar konusuna ilişkin bilgilere yer verilmiştir.

- James, A.T., (1959), “*The Distribution of The Covariance Matrix*”, Yale University” makalesinde kovaryans matrisinin karakteristik köklerinin dağılımına ilişkin bilgilere yer verilmiştir.
- Johnson, R.A., Wichern, D.W., (1999), “*Applied Multivariate Statistical Analysis*” adlı kitap çalışmasında çok değişkenli dağılımlar üzerinde durulmuştur. Bu çalışmada 3. bölümde çok değişkenli dağılımın yayılımını ifade eden kovaryans matrisinin özdeğerleri ve özvektörlerinin kullanılmasıyla elipsoitlerin belirlenebileceği, bu elipsoitler aynı zamanda kovaryans matrisinin elemanlarına bağlı olduğuna ilişkin bilgilere yer vermiştir.
- Kollo, T., Rosen, D., (2005), “*Advanced Multivariate Statistics with Matrics*” adlı kitapta Wishart dağılımı ilgili bilgilere yer verilmiştir, bu çalışmada 1. bölümde Wishart dağılımı ve ilgili teoremlere ilişkin bilgilere yer verilmiştir.
- Miller, I., Miller, M., (2001), “*John E. Freund’dan Matematiksel İstatistik*” adlı kitabında Dağılımlar konusu incelenmiştir.
- Morrison, D.F., (1990), “*Multivariate Statistical Methods, United States of America*” adlı kitabın 3. bölümünde çok değişkenli normal dağılıma ilişkin bilgilere yer verilmiştir.
- Muirhead, R. J., Chikuse, Y., (1975), “*Asymptotic Expansion for The Joint and Marginal Distribution of The Latent Roots of The Covariance Matrix*” adlı çalışmada Wishart dağılımına sahip kovaryans matrislerin karakteristik köklerin dağılımları ele alınmıştır.
- Sakaori, F., Yamada, T., Kawamura, A., Sugiyama, T., (2007), “*A New Confidence Interval for All Characteristic Roots of A Covariance Matrix*” adlı makalesinde örnek kovaryans matrisinin tüm karakteristik kökleri için güven aralıkları araştırılmıştır. Örneğin Sugiyama (1970), Siotani (1976), Konishi (1977), Krishnaiah (1978), Konishi ve Sugiyama (1981) çalışmalarına bakılabilir.

Bunun yanında bu karakteristik kökler için güven aralıkları bulunmuş ve ilgili hipotez testleri birçok durumda kullanılmıştır (Anderson(2003)).

- Sazak, H.S., Okutan, D., (2009), “*Bootstrap Yönteminin Regresyon Analizinde Kullanımı ve Diğer Yöntemlerle Karşılaştırılması*” adlı çalışmasında Bootstrap methodu incelenmiştir.
- Semiz, M., (2007), “*İstatistik laboratuvarı ve SPSS-MP*” adlı kitap çalışmasında simülasyon konularına yer verilmiştir. Bu çalışmada Monte Carlo simülasyonları bu kaynakta anlatılan SPSS-MP programı kullanılarak hazırlanmıştır.
- Sugiyama, T., (1970), “*Joint Distribution of The Extreme Roots of A Covariance Matrix*” adlı çalışmada yığın kovaryans matrisinin en büyük ve en küçük karakteristik kökleri için güven aralıkları incelenmiştir. Bu çalışmanın 4. bölümünde güven aralıkları ile ilgili çalışılmıştır.
- Topuz, D., (2002), “*Regresyonda Yeniden Örnekleme Yöntemlerinin Karşılaştırmalı Olarak İncelenmesi*” adlı yüksek lisans tez çalışmasında Bootstrap ve Jaccknife methotları incelenmiştir.
- Yakupoğlu, Ç., Atıl, H., (2003), “*Bootstrap Metodu Ve Uygulanışı Üzerine Bir Çalışma*” adlı çalışmasında Bootstrap methodu incelenmiştir.

1. TEMEL KAVRAMLAR

1.1. Tek Değişkenli Dağılımlar

1.1.1. Normal dağılım

Normal dağılım pratikte çok sık karşılaşılan ve kullanılan sürekli dağılımlardan biridir. Normal dağılımın önemi, normal dağılımdan türetilen χ^2 , t, F gibi dağılımlarda da görülecektir. Özellikle merkezi limit teoreminden de bilindiği üzere, ele alınan yığında ilgilenilen X tesadüfî değişkeninin dağılım şekli ne olursa olsun, bu yığından ele alınacak belli bir büyüklüğün üstündeki örnekler için örnek ortalaması normal dağılmış bir tesadüfî değişken olarak ele alınmaktadır. Ayrıca normal dağılım örnek çapı büyüdüğü zaman binomial dağılımın bir yaklaşımı gibi görünür. Bu sonuç, çoğu kez pratikte, özellikle de örnekleme metotlarının uygulanmasında kullanılmaktadır. Zira bu kullanım, hesaplamaları büyük ölçüde azaltmaktadır (Er, F., Sönmez, H., 2001).

X rassal değişkeninin olasılık yoğunluğu, ancak ve ancak aşağıdaki gibiyse bu değişken normal dağılıma uyar ve rassal normal değişken adını alır.

$$-\infty < x < \infty \text{ için } n(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$$

burada $\sigma > 0$ 'dır. $\mu = 0$ ve $\sigma = 1$ olan normal dağılıma standart normal dağılım denir (Miller, I.M., 2001).

1.1.2. Gama dağılımı

Bir X rassal değişkeni, ancak ve ancak olasılık yoğunluğu aşağıdaki gibi ise gama dağılımına uyar ve gama rassal değişkeni adını alır.

$$g(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x > 0 \text{ ise} \\ 0 & \text{diğer durumlarda} \end{cases}$$

burada $\alpha > 0, \beta > 0$ 'dır. Gama dağılımının ortalaması $\mu = \alpha\beta$, varyansı $\sigma^2 = \alpha\beta^2$ dir (Miller, I.M., 2001).

1.1.3. Üstel dağılım

Bir X rassal değişkeni, ancak ve ancak olasılık yoğunluğu aşağıdaki gibi ise üstel dağılıma uyar ve üstel rassal değişkeni adını alır. Gama dağılımında $\alpha = 1$ ve $\beta = \theta$ yazılırsa;

$$g(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & x > 0 \text{ ise} \\ 0 & \text{diğer durumlarda} \end{cases}$$

burada $\theta > 0$ 'dır. Üstel dağılımının ortalaması $\mu = \theta$, varyansı $\sigma^2 = \theta^2$ dir (Miller, I.M., 2001).

1.1.4. Ki-kare dağılımı

Rassal bir X değişkeni, ancak ve ancak olasılık yoğunluğu aşağıdaki gibiyse bir ki-kare dağılımına uyar ve ki-kare rassal değişkeni adını alır. μ katsayısına serbestlik derecesi denir.

Gama dağılımında $\alpha = \frac{\mu}{2}$ ve $\beta = 2$ yazılırsa;

$$f(x) = \begin{cases} \frac{1}{2^{\mu/2} \Gamma(\mu/2)} x^{\frac{\mu-2}{2}} e^{-\frac{x}{2}} & x > 0 \text{ ise} \\ 0 & \text{diğer durumlarda} \end{cases}$$

Ki-kare dağılımının ortalaması $\mu = \mu$, varyansı $\sigma^2 = 2\mu$ dir (Miller, I.M., 2001).

1.1.5. Merkezsel olmayan ki-kare dağılımı

n boyutlu rasgele Y vektörü $N(y : \mu, I)$ dağılımına sahip ve $U = Y'Y$ şeklinde tanımlansın.

$$f(x; \alpha, \beta) = \begin{cases} \sum_{j=0}^{\infty} \left(\frac{e^{-\beta} \beta^j}{j!} \right) \left(\frac{x^{(\alpha+2j-2)/2} e^{-x/2}}{\Gamma\left(\frac{\alpha+2j}{2}\right) 2^{j+(\alpha/2)}} \right) & x > 0 \text{ için} \\ 0 & x \geq 0 \text{ için} \end{cases}$$

$\beta = \frac{1}{2} \mu$ ($\beta \geq 0$), ve $\beta = 0$, $j = 0$ olduğunda $\beta^j = 1$ olarak tanımlansın. Bu

dağılıma merkezsel olmayan ki-kare dağılımı denir. α serbestlik derecesi, β merkezsel olmama parametresidir. $\beta = 0$ ve $\mu = 0$ ise dağılım merkezsel ki-kare dağılımına dönüşür (Graybill, 1998).

1.2. Çok Değişkenli Dağılımlar

1.2.1. İki değişkenli normal dağılım

Çok değişkenli yoğunluklar arasında, tek değişkenli normal dağılımın bir genellemesi olan çok değişkenli normal dağılımın özel bir önemi vardır.

X ve Y rassal değişken çifti, ancak ve ancak ortak olasılık yoğunlukları aşağıdaki gibi ise iki değişkenli normal dağılıma uyar ve ortak normal dağılmış rassal değişkenler olarak adlandırılırlar. Burada $\mu_1, \mu_2, \sigma_1, \sigma_2$ katsayıları X ve Y rassal değişkenlerinin ortalamalarıyla standart sapmalarıdır. $-\infty < x < \infty, -\infty < y < \infty$ için,

$$f(x, y) = \frac{e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

burada $\sigma_1 > 0, \sigma_2 > 0, -1 < \rho < 1$ 'dir (Miller, I.M., 2001).

1.2.2. p değişkenli normal dağılım

Çok değişkenli istatistiksel analiz tekniklerinin büyük çoğunluğunda, örneklemelerin çok değişkenli normal dağılımlı kitlelerden geldiği kabul edilmektedir. X, p elemanlı rastlantı değişken vektörü; μ ortalama vektörü, Σ kovaryans matrisi ile çok değişkenli normal dağılıma sahipse $X \square N_p(\mu; \Sigma)$ biçiminde gösterilir.

$$f(x) = ke^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

biçiminde yazılan olasılık yoğunluk fonksiyonu çok değişkenli olması halinde;

$$f(x) = k \exp \left[-\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right]; -\infty < \mu_j < \infty; -\infty < x_j < \infty; \Sigma > 0; j = 1, 2, \dots, p \text{ için}$$

olarak gösterilecektir. Burada $p \times p$ boyutlu, pozitif tanımlı Σ matrisinin elemanları ile $p \times 1$ boyutlu μ ortalama vektörünün elemanları sonludur ve dağılımın parametreleridir. $k = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}}$ için p değişkenli normal dağılımın olasılık

yoğunluk fonksiyonu yazılırsa;

$$f(x) = \begin{cases} \exp \left[-\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right] / (2\pi)^{p/2} |\Sigma|^{1/2} & ; -\infty < \mu_j \in \mu < \infty \\ & ; -\infty < x_j \in x < \infty \\ & ; \Sigma > 0; j = 1, 2, \dots, p \\ 0 & ; \text{öteki durumlarda} \end{cases}$$

Bulunan bu fonksiyon $p=1$ durumu için yazılacak olursa, tek değişkenli normal dağılımın olasılık yoğunluk fonksiyonu elde edilir (Orhunbilge, N., 2000).

1.3. Wishart Dağılımı

Wishart dağılımını ilk kez Fisher (1915) bulmuştur. Bu bulgu iki boyutlu vektörlere ilişkindir. Dağılımı ikiden fazla boyuta genelleştiren ve yoğunluk fonksiyonunu bulan John Wishart (1928)'dir. Bu nedenle dağılıma 'Wishart' adı verilmiştir. Orijinal türetme şeklini bulan Wishart'ın yaklaşımı geometriktir ve merkezsel bir durumu ele almaktadır. Bu yüzden ancak tarih kaygıları nedeniyle önem kazanmaktadır. Bugün artık merkezsel olmayan dağılım türleri de saptanmıştır. Örneğin, James (1954, 1955 ve 1964), Constantine (1963) ve Anderson (1945 ve 1946)'un da bu alandaki çalışmalarından söz edilmesi gerekir. Wishart dağılımı üzerine modern tartışmalarda Rao (1973) tarafından geliştirilen ve yoğunluk fonksiyonu tanımına dayanmayan bir yaklaşım uygulanmaktadır. Bu çalışmada da bu yöntem belirlenmiştir.

Günümüzde Wishart dağılımı olarak bilinen matris dağılımı ilk kez Wishart (1928) tarafından bulunmuştur ve bu dağılım genellikle, ki-kare dağılımının çok değişkenli uzantısı olarak kabul edilir. Wishart dağılımı ile ilgili ilginç yaklaşımlar Hassairi, Lajmi (2001) tarafından ele alınmıştır. Hassairi, Lajmi ve Zine (2004) çalışmalarında Wishart dağılımını anlatmak için bazı olasılıklardan söz edilmiştir (Kollo ve Rosen 2005).

1.3.1. Wishart olasılık yoğunluk fonksiyonu

X_1, \dots, X_N ($N > p$) birbirinden bağımsız ve her biri $N(\mu, \Sigma)$ dağılımına sahip rasgele değişkenler olsun. A'nın dağılımı, $A = \sum_{\alpha=1}^N (X_\alpha - \bar{X})(X_\alpha - \bar{X})'$ şeklindedir.

Burada A matrisi, $\sum_{\alpha=1}^n Z_\alpha Z_\alpha'$ şeklinde yazılabilir, $n = N - 1$ ve Z_1, \dots, Z_n birbirinden bağımsız ve her biri $N(0, \Sigma)$ dağılımına sahip olsun. A pozitif tanımlı matris için A'nın dağılımı (1.1) ile gösterilir (Anderson, 2003).

$$\frac{|A|^{\frac{1}{2}(n-p-1)} \exp\left(-\frac{1}{2}tr\Sigma^{-1}A\right)}{2^{\frac{1}{2}np} \pi^{p(p-1)/4} |\Sigma|^{\frac{1}{2}n} \prod_{i=1}^p \Gamma\left[\frac{1}{2}(n+1-i)\right]} \quad (1.1)$$

$\Sigma = I$ için

$$(Z_1, \dots, Z_n) = \begin{pmatrix} v_1' \\ \vdots \\ v_p' \end{pmatrix}$$

$A = (a_{ij})$ matrisinin elemanları, n elemanlı vektörlerin iç çarpımlarıdır. $a_{ij} = v_i' v_j'$ şeklindedir. v_1, \dots, v_p vektörleri $N(0, I_n)$ şeklinde birbirinden bağımsız olarak dağılır. Yeni koordinatların dönüşümü için $w_1 = v_1$ yazarsak,

$$w_i = v_i - \sum_{j=1}^{i-1} w_j \frac{w_j' v_i}{v_j' w_j}, \quad i = 2, \dots, p \quad (1.2)$$

burada $k < i$ için w_k, w_i 'ye diktir. $w_k' w_h = 0, k \neq h, k, h = 1, \dots, i-1$; w_k 'nin iç çarpımını alırsak (1.2)'den $w_k' w_i = 0, k = 1, \dots, i-1$ elde edilir ($\Pr \{\|w_i\| = 0\} = 0$).

$$t_{ii} = \|w_{ii}\| = \sqrt{w_i' w_i}, \quad i = 1, \dots, p \quad \text{ve} \quad t_{ij} = v_i' w_j / \|w_j\|, \quad j = 1, \dots, i-1, \quad j = 2, \dots, p$$

$$v_i = \sum_{j=1}^i (t_{ij} / \|w_j\|) w_j,$$

$$a_{hi} = v_h' v_i = \sum_{j=1}^{\min(h,i)} t_{hj} t_{ij} \quad T = (t_{ij})$$

matrisi alt üçgen matris olsun $t_{ij} > 0, i = 1, \dots, p$ ve $t_{ij} = 0, i < j$

$$A = TT'$$

Sonuç 1. Z_1, \dots, Z_n ($n \geq p$) her biri ayrı ayrı $N(0,1)$ dağılımına sahip rasgele değişkenler, $A = \sum_{\alpha=1}^n Z_{\alpha} Z_{\alpha}' = TT', t_{ij} = 0, i < j$ ve $t_{ii} > 0, i = 1, \dots, p$ olsun, $t_{11}, t_{21}, \dots, t_{pp}$ ayrı ayrı dağılıma sahip olsun öyle ki, $t_{ij}, i > j$ için $N(0,1)$ dağılımına sahip, t_{ii}^2 $n - i + 1$ serbestlik dereceli χ^2 dağılımına sahip olsun.

$$t_{ii} \sim 2^{-\frac{1}{2}(n-i+1)} t^{n-i} e^{-\frac{1}{2}t^2} / \Gamma\left[\frac{1}{2}(n+1-i)\right], \quad t_{ij}, \quad j = 1, \dots, i \quad i = 1, \dots, p$$

$$\prod_{i=1}^p \frac{t_{ii}^{n-i} \exp\left(-\frac{1}{2} \sum_{j=1}^i t_{ij}^2\right)}{\pi^{\frac{1}{2}(i-1)} 2^{\frac{1}{2}n-1} \Gamma\left[\frac{1}{2}(n+1-i)\right]} = \frac{\prod_{i=1}^p t_{ii}^{n-i} \exp\left(-\frac{1}{2} \sum_{i=1}^p \sum_{j=1}^i t_{ij}^2\right)}{2^{\frac{1}{2}p(n-2)} \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma\left[\frac{1}{2}(n+1-i)\right]} \quad (1.3)$$

C matrisi alt üçgen matris olsun, ($c_{ij} = 0, i < j$) $\Sigma = CC'$ ve $c_{ii} > 0$. Lineer dönüşüm

$T^* = CT$ şeklinde olsun.

$$t_{ij} = \sum_{k=j}^i c_{ik} t_{kj}, \quad i \geq j \text{ için}$$

$$= 0, \quad i < j \text{ için}$$

Matris şeklinde yazılırsa;

$$\begin{bmatrix} t_{11}^* \\ t_{21}^* \\ t_{31}^* \\ \vdots \\ t_{p1}^* \\ \vdots \\ t_{pp}^* \end{bmatrix} = \begin{bmatrix} c_{11} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ x & c_{22} & 0 & \cdots & 0 & \cdots & 0 \\ x & x & c_{22} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ x & x & x & \cdots & c_{pp} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ x & x & x & \cdots & x & \cdots & c_{pp} \end{bmatrix} \begin{bmatrix} t_{11} \\ t_{21} \\ t_{22} \\ \vdots \\ t_{p1} \\ \vdots \\ t_{pp} \end{bmatrix} \quad (1.4)$$

burada x ile gösterilen elemanlar sıfırdan farklıdır ve determinanı köşegen

üzerindeki elemanlarının çarpımıdır; $\prod_{i=1}^p c_{ii}^i$

$$\begin{aligned} \sum_{i=1}^p \sum_{j=1}^i t_{ij}^2 &= tr TT' = tr C^{-1} T^* T^* (C^{-1})' = tr T^* T^{*'} (C^{-1})' = tr T^* T^{*'} C'^{-1} C^{-1} \\ &= tr T^* T^{*'} \Sigma^{-1} = tr T^{*'} \Sigma^{-1} T^* \end{aligned}$$

burada $\prod_{i=1}^p c_{ii}^2 = |C| |C'| = |\Sigma|$

Teorem 1. Z_1, \dots, Z_n her biri ayrı ayrı $N(0, \Sigma)$ dağılımına sahip rasgele değişkenler

olsun. $A = \sum_{\alpha=1}^n Z_{\alpha} Z_{\alpha}'$ nın dağılımı;

$$\frac{|A|^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}tr\Sigma^{-1}A}}{2^{\frac{1}{2}np} \pi^{p(p-1)/4} |\Sigma|^{\frac{1}{2}n} \prod_{i=1}^p \Gamma\left[\frac{1}{2}(n+1-i)\right]}, \text{ A pozitif tanımlı ise}$$

0, diğer durumlarda (1.5)

(Anderson, 2003).

X_1, \dots, X_N ($N > p$) her biri ayrı ayrı $N(\mu, \Sigma)$ dağılımına sahip rasgele değişkenler ve $n = N - 1$ için $A = \sum_{\alpha=1}^N (X_{\alpha} - \bar{X})(X_{\alpha} - \bar{X})'$ olsun. (1.5) eşitliği $W(\Sigma, n)$ ile gösterilir.

1.3.2. Merkezsel olmayan wishart dağılımı

W matrisi $p \times p$ boyutlu Wishart dağılımına sahip, $W = XX'$ olsun öyle ki; X matrisi; $X \sim N_{p,n}(\mu, \Sigma, I), \Sigma \geq 0$ olsun. Eğer $\mu = 0$ ise $W \sim W_p(\Sigma, n)$ merkezi Wishart dağılımını, $\mu \neq 0$ ise $W \sim W_p(\Sigma, n, \Delta)$ ve $\Delta = \mu\mu'$ olan, merkezi olmayan Wishart dağılımını gösterir (Kollo ve Rosen 2005).

Σ parametresi genellikle bilinmeyen, ikinci parametre olan n ise serbestlik derecesini gösteren genellikle bilinen olarak kabul edilir. Üçüncü parametre olan Δ ise , merkezi olmayan Wishart dağılımında kullanılan, merkezi olmayan parametre olarak adlandırılır. Genel olarak $\Delta \neq 0$ olduğunda durumun daha karmaşık olmaya başladığını söyleyebiliriz. Wishart dağılımı, matris dağılımları sınıfına ait bir dağılımdır. Eğer $p = 1$, $\mu = 0$ ve $\Sigma = 1$ ise Wishart matrisi ile n serbestlik dereceli χ^2 değişkeni aynıdır ve formül (1.6) yazılır.

$$f_{\chi^2}(x) = \left(2^{n/2} \Gamma(n/2)\right)^{-1} x^{n/2-1} e^{-x/2}, \quad x > 0, \quad (1.6)$$

burada $\Gamma(\cdot)$, gama fonksiyonudur. Eğer $p=1$, $\Sigma=1$ ama $\mu \neq 0$ ise merkezsiz olmayan n serbestlik dereceli χ^2 dağılımı elde edilmiş olur ve merkezsiz olmayan parametre $\delta = \mu^2$ dir (Kollo ve Rosen, 2005).

1.3.3. Wishart matrislerinin toplamının dağılımı

A_i $i=1,2$ için ayrı ayrı $W(\Sigma, n_i)$ dağılımına sahip olsun. $Z_1, \dots, Z_{n_1+n_2}$ her biri ayrı ayrı $N(0, \Sigma)$ dağılımına sahip olsun ve $A_1, \sum_{\alpha=1}^{n_1} Z_\alpha Z'_\alpha$ ve $A_2, \sum_{\alpha=n_1+1}^{n_1+n_2} Z_\alpha Z'_\alpha$ gibi dağılsın. $A = A_1 + A_2$, $n = n_1 + n_2$ için $\sum_{\alpha=1}^n Z_\alpha Z'_\alpha$ gibi dağılsın. Burada iki matrisin toplamı şeklinde yazılan A matrisi $W(\Sigma, n_i)$ dağılımına sahiptir. Aynı şekilde q adet matrisin toplamı şeklinde yazılan matris de Wishart dağılımına sahiptir ve serbestlik derecesi, toplanan her bir matrisin serbestlik dereceleri toplamına eşittir.

Teorem 2. A_1, \dots, A_q her biri ayrı ayrı A_i $W(\Sigma, n_i)$ dağılımına sahip olsun,

$A = \sum_{i=1}^q A_i$, $W\left(\Sigma, \sum_{i=1}^q n_i\right)$ dağılımına sahiptir.

$$A = \sum_{i=1}^q A_i \quad (1.7)$$

$W\left(\Sigma, \sum_{i=1}^q n_i\right)$ 'a göre dağılır (Anderson, 2003)

1.4. Hipergeometrik Fonksiyon

Hipergeometrik serinin genel formu aşağıdaki gibidir;

$$F(\alpha, \beta; \gamma; z) = 1 + \frac{\alpha \cdot \beta}{\gamma \cdot 1} z + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1) \cdot 1 \cdot 2} z^2 + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{\gamma(\gamma+1)(\gamma+2) \cdot 1 \cdot 2 \cdot 3} z^3 + \dots \quad (1.8)$$

Hipergeometrik seriler, α veya β negatif tamsayıya veya sıfıra eşit olduğunda sınırlı olur. $\gamma = -n$ ($n = 0, 1, 2, \dots$), için α veya $\beta = -m$ ($m < n$ ve m doğal sayı) 'e eşit olmadığı sürece hipergeometrik seri sınırsızdır.

$$\lim_{\gamma \rightarrow -n} \frac{F(\alpha, \beta; \gamma; z)}{\Gamma(\gamma)} = \frac{\alpha(\alpha+1)\dots(\alpha+n)\beta(\beta+1)\dots(\beta+n)}{(n+1)!} z^{n+1} F(\alpha+n+1, \beta+n+1; n+2; z)$$

genelleştirilmiş hipergeometrik fonksiyon; ${}_p F_q(a_1, \dots, a_p; b_1, \dots, b_q; x)$ şeklindedir.

Bu fonksiyonda $p = 2$ ve $q = 1$ için yazılırsa;

$$\begin{aligned} {}_2 F_1(a, b; c; z) &= 1 + \frac{ab}{1!c} z + \frac{a(a+1)b(b+1)}{2!c(c+1)} z^2 + \dots \\ &= \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!} \end{aligned} \quad (1.9)$$

elde edilir (Gradshteyn, Ryzhik, 1965).

Örnek 1.1.

$$\begin{aligned} {}_2 F_1(1, 1; 1; z) &= 1 + \frac{1 \cdot 1}{1!1} z + \frac{1 \cdot 2 \cdot 1 \cdot 2}{2!1 \cdot 2} z^2 + \dots \\ &= 1 + z + z^2 + z^3 + \dots \\ &= \frac{1}{1-z} \end{aligned}$$

Örnek 1.2.

$$\begin{aligned} {}_2F_1(1,2;1, z) &= 1 + \frac{1.2}{1!.1}z + \frac{1.2.2.3}{2!.1.2}z^2 + \frac{1.2.3.2.3.4}{3!.1.2.3}z^3 + \dots \\ &= 1 + 2z + 3z^2 + 4z^3 + \dots \\ &= \frac{1}{(1-z)^2} \end{aligned}$$

Örnek 1.3.

$$\begin{aligned} {}_2F_1(1,2;2, z) &= 1 + \frac{1.2}{1!.2}z + \frac{1.2.2.3}{2!.2.3}z^2 + \frac{1.2.3.2.3.4}{3!.2.3.4}z^3 + \dots \\ &= 1 + z + z^2 + z^3 \\ &= \frac{1}{1-z} \end{aligned}$$

1.5. Jackknife Yöntemi

Jackknife tekniği ilk kez 1949 yılında Quenouilli tarafından ortaya atılmıştır. Yöntem, populasyon parametrelerinin tahmin edilmesinde örnek hatasının en aza indirilmesinde , tahmin edicinin sapmasının hesaplanmasında kullanılmasına ek olarak güçlü güven aralıkları oluşturulması amacıyla yönelik geliştirilmiştir.

Bootstrap gibi jackknife da, parametrik varsayımlar kullanmak yerine, örnek değişkenliğinin açıklanması yoluyla elde edilen tahmin edicilerin güvenilirliğini arttırmaya yöneliktir. Jackknife , örneğin değişkenliğini farklı bir yolla açıklamaya çalışan bir yöntemdir (Topuz, 2002).

Bilinmeyen bir $F(x)$ olasılık dağılımından bağımsız ve genişliği n olan $X = (x_1, x_2, \dots, x_n)$ örnek veri setinden, her defasında örnekteki verilerden biri dışarıda bırakılarak geriye kalan $n-1$ gözlemi olan n tane jackknife örneği oluşturulur.

$\hat{\theta}_{(-j)}$; j.gözlem çıkarıldıktan sonra geriye kalan n-1 gözlemden hesaplanan j.

Jackknife tahmin değeri, $\bar{\hat{\theta}}_{(-j)}$; n tane jackknife tahmin değerlerinin ortalaması olmak

üzere, $\hat{\theta}$ tahmin edicisinin, oluşan jackknife örneklerinde aldığı değerler,

$$\hat{\theta}_{(-j)} = \hat{\theta}(X_{(j)}) = \hat{\theta}(X_1, X_2, \dots, X_{j-1}, X_{j+1}, \dots, X_n) \quad j = 1, 2, \dots, n$$

elde edilir. Bu değerler kullanılarak,

$$\bar{\hat{\theta}}_{(-j)} = \frac{\sum_{j=1}^n \hat{\theta}_{(-j)}}{n}$$

değeri hesaplanır.

Jackknife sapma tahmini, $b(\hat{\theta}_{(-j)})_{jack}$; j. Jackknife tahmin edicisinin sapma değeri

$$\begin{aligned} b(\hat{\theta}_{(-j)})_{jack} &= (n-1)(\bar{\hat{\theta}}_{(-j)} - \hat{\theta}) \\ &= (n-1) \left\{ \bar{\hat{\theta}}_{(-j)} - \hat{\theta}(x_1, x_2, \dots, x_n) \right\} \end{aligned}$$

dır. Jackknife standart hata tahmini,

$$s_{jackk} = \sqrt{\frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{-j} - \bar{\hat{\theta}}_{(-j)})^2}$$

dır (Efron 1982).

Jackknife yöntemi modeller ile ilgili varsayımlara daha az bağlı olan, teorik formüller gerektirmeyen bir yöntemdir. Ancak, bir tahmin edicinin n defa hesaplanmasını gerektirmektedir.

Jackknife güven aralıkları ilk kez 1958 yılında Tukey tarafından geliştirilmiştir. %100(1-2 α) güven düzeyindeki Jackknife güven aralıkları,

$$P(\hat{\theta}_{(-j)} - t_{n-1,\alpha} S_{jackk} < \theta < \hat{\theta}_{(-j)} + t_{n-1,\alpha} S_{jackk}) = 1 - 2\alpha$$

şeklindedir.

1.6. Bootstrap Yöntemi

Bootstrap yöntemi ilk kez Efron tarafından öne sürülmüştür. Bu yöntem, eldeki örnek verilerinin iadeli ve rastsal örnekleme yöntemiyle oluşturulmasına, bu şekilde oluşturulan her bir örneklem için ilgili istatistiklerin tahmin edilmesine ve bu işlemin defalarca tekrarına dayanan bir süreçtir.

Bootstrap yöntemi, matematik formülleri içermeyen, çok basit bir yöntemdir. Veri dağılımı ile ilgili ağır varsayımlar gerektirmediğinden bilinen istatistiksel yöntemlerin ve varsayımların yetersiz kaldığı durumlarda güvenilir sonuçlar vermektedir. Veri dağılımına dayanan varsayımların doğruluğundan kesin olarak emin olunamadığı için bu yaklaşım uygulamada çok yararlı olabilmektedir. Regresyon analizinde hata terimleri ve bağımsız değişkenlerle ilgili yapılan varsayımların gerçekleşmemesi durumunda bir düzeltme işlemi amacıyla da kullanılan bootstrap yöntemi, daha küçük tahmin hatalarının elde edilmesi, standart sapmaların küçülmesi ve buna bağlı olarak da daha güvenilir parametre tahminlerinin elde edilmesi ve güven aralıklarının oluşturulması amacıyla geliştirilmiştir (Sazak ve Okutan, 2009).

Günümüzde ise istatistiksel hesaplamalardaki modern gelişmelere paralel olarak, Bootstrap metodunda da ilerlemeler kazanılmış ve uygulamalı istatistik alanında kullanımı giderek artmıştır.

Bootstrap metodu uygulanma kolaylığı ve yararlılığının yanı sıra başka avantajlara da sahiptir. Klasik istatistikte, incelenen değişkenlerin normal dağılışı gösterdiği varsayımına dayalı olarak tahminlemeler yapılmaktadır. Bootstrap metodunda ise veri setinden şansa bağlı örnekler alınarak istatistiksel tahminlemeler yapılmaktadır. Bu metod ile, çok küçük veri setlerinde bile doğru yanıtlar alınabilmekte, büyük veri setlerinde ise klasik istatistik sonuçlarla paralellik sağlanmakta, hemen hemen tüm istatistikler analiz edilebilmektedir.

Bootstrap metodu, çok karmaşık matematik formüllerin çözülmesinde hesaplama yükünü azaltması sebebiyle de avantajlıdır. Ayrıca, verilerin dağılışı hakkında herhangi bir varsayım taşımamakta ve herhangi bir istatistiğin değişkenliği hakkında bilgi verebilmektedir. Bu nedenle, diğer metotların kullanımının uygun olmadığı ya da bilinen varsayımların geçersiz olduğu durumlarda Bootstrap metodu tercih edilebilmektedir. Bunun yanısıra, popülasyonlardan alınan örnek veri setlerine dayalı bir tahminin güvenilirliğini ortaya koymak amacıyla da bu yöntem önem taşımaktadır. Ayrıca, mevcut veri setinden şansa bağlı olarak tekrar tekrar örneklemeler yapılarak parametrelerin yaklaşık güven aralıklarının tahminlemesi bakımından da bu metod kullanışlıdır.

Bootstrap metodu standart sapma, güven aralığı gibi istatistiklerde ve parametrik olmayan tahminleme problemlerinde kullanılan basit ve güvenilir bir metoddur. Bu metod aşağıdaki gibi açıklanabilmektedir.

Herhangi bir $S(x)$ istatistiğini hesaplamak üzere, N adet gözlemden oluşan veri setini $x = (x_1, x_2, x_3, x_4, \dots, x_N)$ ele alalım. Bu orjinal veri setinde gözlemlerin yer değiştirilerek, $1/N$ kadar olasılıkla şansa bağlı seçimiyle bootstrap örnek veri seti $x_i^* = (x_1, x_2, x_3, \dots, x_N)$ elde edilmektedir. Bu işlem istenildiği kadar yinelenerek birbirinden farklı B adet Bootstrap gözlemler seti oluşturulmaktadır. İlgili istatistik bu yeni veri setleri kullanılarak hesaplanmaktadır.

Bootstrap metoduna göre $S(x)$ istatistiğinin standard sapmasını hesaplamak için x veri setinden yer değiştirmeyeyle N bireylik B adet Bootstrap örnekleri $(x_1^*, x_2^*, x_3^*, \dots, x_B^*)$ oluşturulur. Her bir Bootstrap örneğinde söz konusu istatistik hesaplanır.

$$\hat{\theta}_i^* = S(x_i^*) \quad i=1,2,\dots,B$$

(1.10) ile standard sapma hesaplanır. (Yakupoğlu ve Atıl 2003).

$$\sigma^* = \left[\frac{1}{B-1} \sum_{i=1}^B (\theta_i^* - \langle \theta^* \rangle)^2 \right]^{1/2}, \quad \langle \theta^* \rangle = \sum_{i=1}^B \theta_i^* / B \quad (1.10)$$

2. ÖZDEĞERLERİN KULLANILDIĞI TEMEL İSTATİSTİK YÖNTEMLER

2.1. Temel Bileşenler Analizi

Çok değişkenli istatistiksel analizde n tane nesneye ilişkin p tane değişken incelenmektedir. Bu özelliklerden birçoğunun birbiriyle ilişkili ve p değişken sayısının çok büyük olması analizde sorun yaratmaktadır. Örneğin insanın anatomik özellikleri değişkenleri ifade ediyor olsun. Bu durumda karın çevresi, ağırlık, boy uzunluğu, kol uzunluğu, omuz genişliği, bacak uzunluğu gibi çok sayıda değişken bulunmaktadır. Bu değişkenlerin bazıları birbirleri ile ilişkilidir. Oysa ki bu durum değişkenlerin bağımsızlığı kuralını zedeler. Ayrıca çok sayıda değişkenle çalışmak, işlem yükünü artıracaktır ve elde edilecek sonuçların yorumunda bazı güçlükler neden olacağı için arzulanan bir durum değildir. Bilgisayar olanaklarının çok geliştiği günümüzde, işlem yükü bir sorun olarak görülmesi de, çok sayıda değişkene ilişkin analiz sonuçlarının yorumlanması gerçekten zor olabilmektedir. Böyle durumlarda başvurulan tekniklerden en önemlisi Temel Bileşenler Analizi olarak adlandırılır. Genel olarak değişkenler arasındaki bağımlılık yapısının yok edilmesi veya boyut indirgeme amacıyla kullanılan Temel Bileşenler Analizi başlı başına bir analiz olduğu gibi, başka analizler için veri hazırlama tekniği olarak da kullanılmaktadır (Tatlidil, 2002).

n gözlem ve p değişkenden oluşan veri matrisi X 'in p boyutlu uzaydaki durumu düşünülecek olursa, veri matrisi çok sayıda noktadan oluşan bir topluluk olarak ifade edilebilir. Değişkenler arasında tam bağımsızlık söz konusu olamayacağı için bulut biçiminde ifade edilen geometrik şeklin eksenleri birbirine dik olmayacak ve tanımı da yapılamayacaktır. Oysa ki, bu noktaları eksenleri birbirine dik bir elipsoid içerisine almak daha ayrıntılı ve açıklayıcı bilgi verecektir. Bu amaçla uygulanan dönüştürmede, noktaların ilk eksenler boyunca sahip oldukları toplam varyans değişmediği gibi yeni eksenler birbirine dik olmaktadır. Ham veri matrisinin kullanılması durumunda temel bileşenlerin bulunmasında varyans-kovaryans

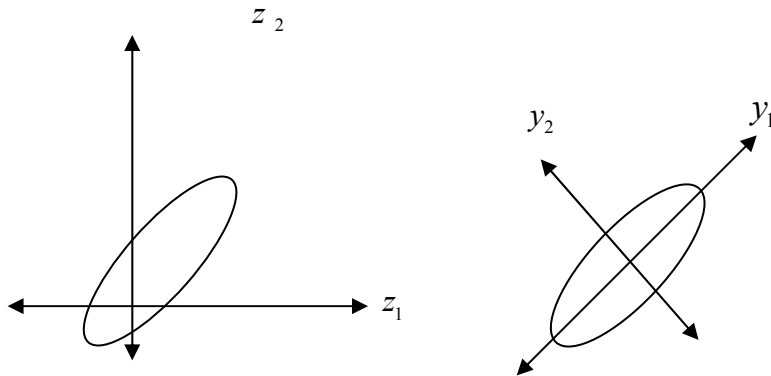
matrisinden, standartlaştırılmış veri matrisinin kullanılması durumunda ise korelasyon matrisinden yararlanılmaktadır. Oldukça farklı sonuçlar verebilen bu iki yoldan hangisinin seçileceği konusunda en önemli belirleyici özellik verilerin ölçü birimleridir. Eğer verilerin (değişkenlerin) ölçü birimleri ve varyansları birbirine yakın ise kovaryans matrisinden, değilse korelasyon matrisinden yararlanılması önerilir (Tatlídil, 2002).

2.1.1. Temel bileşenlerin elde edilmesi

Değişkenlerin ölçü birimlerinin birbirine yakın olması pratikte pek olağan olmayacağına göre, veri matrisi olarak X_{pxn} boyutlu ham veri matrisi değil, bunların standartlaştırılmış değerlerinden oluşan Z_{pxn} standart matrisi kullanılacaktır. Bu durumda yukarıda sözü edilen dönüştürme, T_{pxn} bir dönüşüm matrisi olmak üzere,

$$Y_{pxn} = T'_{pxn} Z_{pxn}$$

biçiminde olmaktadır. Yani birbiri ile ilişkili z_{ij} değerlerinden dönüştürme sonunda, birbiri ile ilişkisiz y_{ij} değerlerine ulaşılmaktadır. Bu dönüşüm aşağıdaki şekilde gösterilmiştir.



Şekil 2.1. Temel Bileşenler Dönüşümünün İki Boyutlu Uzayda Gösterimi

Y tesadüfi vektörünün beklenen değer ve kovaryans matrisi

$$E(Y) = E(T'Z) = T'E(Z) = 0 \text{ ve } Var(Y) = T'E(ZZ')T = T'RT$$

olur. Burada R : $p \times p$ boyutlu deęişkenler arası korelasyon matrisidir. Dönüştürülmüş Y matrisinin tesadüfi vektörlerinin (deęişken) birbirine dik olabilmeleri için $Var(Y)$ matrisinin köşegen olması gerekir. Bu matrisinin köşegenleştirilmesinde çok sayıda T dönüşüm matrisinin kullanılması söz konusudur. Birbirinden farklı bu dönüşüm matrislerinden amaca en uygun olanının seçilebilmesi için y vektörleri üzerine bazı kısıtlayıcıların konması gerekir. Bu kısıtlayıcılar şöyledir:

y vektörlerinin ilki olan y_1 öyle seçilmelidir ki varyansı en büyük olsun. Yani;

$$Var(y_1) = \max \frac{1}{n-1} \sum_{i=1}^n (y_{1i})^2$$

olmalıdır.

y_1 vektörünün bulunmasında kullanılan t_1 vektörünün elemanlarının kareleri toplamı 1 olmalıdır.

$$t_1' t_1 = 1$$

Bu kısıtlayıcılar yardımıyla z_i vektöründen, dönüşüm sonucu elde edilen y_1 vektörünün i 'nci elemanı,

$$y_{1i} = t_1' z_i$$

biçiminde bulunur. İlk kısıtlayıcı nedeniyle y_1 vektörünün varyansı,

$$\frac{1}{n-1} \sum_{i=1}^n (y_{1i})^2 = \frac{1}{n-1} \sum_{i=1}^n (t_1' z_i)^2 = \frac{1}{n-1} \sum_{i=1}^n t_1' z_i z_i' t_1$$

$$Var(y_1) = t_1' \frac{1}{n-1} ZZ' t_1 = t_1' R t_1$$

olarak bulunur. y_1 vektörünün varyans deęeri olan $t_1' R t_1$ 'in ikinci kısıtlayıcıdan da yararlanılarak en büyük yapılması söz konusudur. Bu amaçla,

$$\varphi_1 = t_1' R t_1 - \lambda_1 (t_1' t_1 - 1)$$

fonksiyonu verilen kısıt altında çözülür. Fonksiyonun t_1 'e göre türevi alınıp sıfıra eşitlenecek olursa;

$$\frac{\partial \varphi_1}{\partial t_1} = 2R t_1 - 2\lambda_1 t_1 = 0$$

ise

$$(R - \lambda_1 I) t_1 = 0$$

elde edilir. Burada λ_1 değeri, R matrisinin özdeğeri, t_1 vektörü de R matrisinin özvektörü olarak adlandırılır. Özdeğerleri elde etmek için;

$$|R - \lambda I| = 0 \tag{2.1}$$

açılımından elde edilen p 'inci dereceden polinom denklem çözülerek p tane λ değeri bulunur. R matrisi pozitif tanımlı ve simetrik olduğu için elde edilecek değerlerin tümü gerçek değerler olacaktır. (2.1)'den elde edilen p tane özdeğer kullanılarak her birine karşılık gelen p tane özvektör elde edilir.

Bağıntının kullanımı ile elde edilen özdeğerlerden birini λ_1 olarak ve ilgili özvektörü de t_1 olarak göstermiş olalım. Bu bağıntı soldan t_1' ile çarpılacak olursa;

$$t_1' R t_1 - \lambda_1 t_1' t_1 = 0$$

elde edilir. Buradan $t_1' t_1 = 1$ olması nedeniyle (kısıtlayıcı), $t_1' R t_1 = \lambda_1$ olacaktır. Sonuç olarak;

$$Var(y_1) = Var(\sqrt{\lambda_1 t_1}) = E(\sqrt{\lambda_1 t_1})(\sqrt{\lambda_1 t_1})' = \lambda_1 t_1 t_1' = \lambda_1 \quad (2.2)$$

bulunur (Tatlıldil, 2002). Yani, y_1 değişkeninin varyansı λ_1 'dir. Temel bileşenler analizinde y_1 'in varyansının en büyük olması istendiğinden, λ_1 değeri λ_j değerleri arasında en büyük değerli olarak seçilir. Seçilen λ_1 değerinin kullanımı ile elde edilen t_1 vektörüne birinci özvektör adı verilecektir. Birinci özdeğer λ_1 ve birinci özvektör t_1 olmak üzere; t_1 ile orijinal veri matrisi Z 'nin çarpımından elde edilen $y_1 = t_1'Z$ dönüştürülmüş vektöre de birinci temel bileşen adı verilir. Ayrıca bazı kaynaklarda t_1 vektörü yerine $\sqrt{\lambda_1} t_1$ vektörü kullanılmakta varyans ağırlıklarının katsayılarına etki etmesi sağlanarak birinci asıl temel bileşen adı verilmektedir.

İkinci temel bileşen y_2 bulunurken, y_1 vektörünün bulunmasında kullanılan iki kısıtlayıcı yanında üçüncü bir kısıtlayıcı da göz önüne alınır. Bu kısıtlayıcılar:

$$\begin{aligned} y_2 \text{ vektörünün varyansı, } y_1 \text{ 'den sonra en büyük olsun,} \\ t_2 \text{ vektörü birim normal bir vektör olsun. } (t_2' t_2 = 1), \\ y_1 \text{ ve } y_2 \text{ vektörleri birbirine dik olsun. } (t_2' t_1 = 0), \end{aligned}$$

biçimindedir. Bu üç kısıtlayıcı yardımıyla aşağıdaki bağıntıda verilen ,

$$y_2 = t_2' Z$$

vektörü bulunacaktır. İlk olarak üçüncü kısıtlayıcıdan,

$$\frac{1}{n-1} \sum_{i=1}^n y_{1i} y_{2i} = \frac{1}{n-1} \sum_{i=1}^n (t_1' z_i)(t_2' z_i)' = t_1' \frac{1}{n-1} \sum_{i=1}^n z_i z_i' t_2 = t_1' R t_2$$

ve $R t_i = t_i \lambda_i$ 'den

$$t_1' R t_2 = \lambda_1 t_1' t_2 = 0 \quad (2.3)$$

sonucuna ulaşılır. y_2 vektörünün varyansı,

$$Var(y_2) = E(t_2'Z)^2 = \frac{1}{n-1} \sum_{i=1}^n t_2' z_i z_i' t_2 = t_2' R t_2$$

dir. Varyans değeri $t_2' t_2 = 1$ ve $t_1' t_2 = t_2' t_1 = 0$ koşulları altında en büyük yapılmak istenirse yine lagranj çarpanlarından yararlanır.

$$\varphi_2 = t_2' R t_2 - \lambda_2 (t_2' t_2 - 1) - \gamma_1 (t_2' t_1 - 0) = 0$$

burada λ_2 ve γ_1 lagranj çarpanlarıdır. φ_2 'nin t_2 'ye göre türevi alınıp sıfıra eşitlendiğinde,

$$\frac{\partial \varphi_2}{\partial t_2} = 2R t_2 - 2\lambda_2 t_2 - \gamma_1 t_1 = 0$$

bulunur. Bu eşitlik, (2.1) bağıntısından da yararlanılarak aşağıdaki biçimde gösterilir.

$$2R t_2 - 2\lambda_2 t_2 - \gamma_1 t_1 = 2(R - \lambda_2 I) t_2 - \gamma_1 t_1 = 0$$

bu eşitlikte $\gamma_1 t_1 = c = 0$ olmalı ki,

$$(R - \lambda_2 I) t_1 = 0$$

eşitliği yazılabilir. t_1 vektörü sıfır olamayacağından, bu koşulun sağlanabilmesi için $\gamma_1 = 0$ olmalıdır. Bu durumda ikinci özdeğer yine,

$$|R - \lambda_2 I| = 0;$$

bağıntısından elde edilir. Bu sonuca göre λ_2 , R matrisinin ikinci özdeğeri ve t_2 , R matrisinin ikinci özvektörüdür. Elde edilen y_2 vektörüne ikinci temel bileşen, $\sqrt{\lambda_2} t_2$ değerine de ikinci asıl temel bileşen katsayıları adı verilir. İkinci temel bileşen, varyansı y_1 in varyansından sonra en büyük ve y_1 vektörüne diktir. Bu biçimde devam edilecek olursa, $j=1,2,\dots,p$ için tüm λ_j , t_j ve y_j değerleri elde edilir. Bu

durumda, λ_p en küçük değere sahip özdeğer ve y_p ise en küçük varyanslı temel bileşendir.

$RTT' = T\Lambda T' \Rightarrow R = T\Lambda T'$ olduğu bilinmektedir. Burada $\Lambda = p \times p$ boyutlu, köşegen elemanları λ_j 'ler, köşegen dışı elemanları sıfır olan bir matristir. Bu durumda;

$$|R| = |T\Lambda T'| = |T||\Lambda T'| = |T||\Lambda||T'| = |\Lambda||T||T'| = |\Lambda| = \prod_{i=1}^p \lambda_j \quad (2.4)$$

ve

$$\dot{I}z(R) = \dot{I}z(T\Lambda T') = \dot{I}z(T'T\Lambda) = \dot{I}z(\Lambda) = \sum_{j=1}^p \lambda_j = p \quad (2.5)$$

olduğu gösterilmiş olur (Tatlıldil, 2002).

2.1.2. Temel bileşenlerin özellikleri ve sağladığı yararlar

İstatistiksel analizlerde, y_j vektörlerinin birçok özelliğinden yararlanılmaktadır. Bu özelliklerden bazıları şöyle sıralanabilir:

Gerek ham veri matrisi X ve gerekse standartlaştırılmış biçimi olan Z matrisinde değişkenler arasında bağımlılık söz konusu iken, y_j vektörleri birbirinden bağımsızdır. Geometrik olarak y_j değerleri dik eksenlere göre elde edilmiştir. Oysa ki z_j değerleri eğik eksenler üzerinde bulunmakta idi.

Noktaların z_j eksenlerine göre varyansı değişiktir ve eksenler arası kovaryans terimi de bulunmaktadır. Oysa ki y_j eksenlerinin varyansları büyükten küçüğe doğru sıralanır. Ayrıca eksenler birbirine dik olduğundan, kovaryans terimi yoktur ve noktaların dağılımı yalnız varyansla açıklanmaktadır.

Bu özelliklere ek olarak, eğer ilk m tane temel bileşen toplam varyansın büyük kısmını açıklıyorsa, geriye kalan $p-m$ tane temel bileşen ihmal edilebilir. Bu durumda az bir varyans kaybıyla üzerinde çalışılan uzayın boyutu p 'den m 'ye ($p < m$) indirgenmiş olur.

z_j değişkenlerinin varyansının tümü y_j değişkenleri tarafından açıklanmaktadır. Bu nedenle, p tane y_j temel bileşenin kullanılması durumunda boyut indirgeme kazancı sağlamasa bile, hiçbir varyans kaybı olmaksızın p tane bağımsız yeni değişken elde edilmiş olur.

2.2. Faktör Analizi

Faktör analizi, başlıca amacı aralarında ilişki bulunduğu düşünülen çok sayıdaki değişken arasındaki ilişkilerin anlaşılmasını ve yorumlanmasını kolaylaştırmak için daha az sayıdaki temel boyuta indirmek veya özetlemek olan bir grup çok değişkenli analiz tekniğine verilen genel bir isimdir. Diğer bir ifade ile faktör analizi, aralarında ilişki bulunan çok sayıda değişkenden oluşan bir veri setine ait temel faktörlerin (ilişkinin yapısının) ortaya çıkarılarak araştırmacı tarafından veri setinde yer alan kavramlar arasındaki ilişkilerin daha kolay anlaşılmasına yardımcı olmaktır.

Faktör analizinde araştırmacı öncelikle araştırma bağlamında kullanılan değişkenler setinin temelini oluşturan ana (temel) faktörlerin neler olduğu ve bu faktörlerden her birinin değişkenlerden her birini açıklama derecesini de görme imkanına sahip olmaktadır. Bu yolla, araştırmacı elinde bulunan çok sayıda değişkenden oluşan değişkenler setini daha az sayıda yeniden oluşturulmuş değişkenler (faktörler) cinsinden ifade etme ve anlama imkânına sahip olacaktır.

Etkin bir faktör analizinin toplam veri setini en iyi temsil edebilen ancak mümkün olduğunca az sayıda faktörden oluşan bir çözüm olması arzu edilen

özelliğidir. İyi bir faktör analizi sonucunun hem basit hem de yorumlama becerisinin iyi olması arzu edilir.

Faktör analiziyle ilgili teknikleri Açıklayıcı Faktör Analizi ve Onaylayıcı Faktör Analizi olmak üzere iki başlık altında incelemek mümkündür.

Açıklayıcı faktör analizinde araştırmacı araştırma yaptığı konuyla ilgili olarak değişkenler arasındaki ilişkiye yönelik olarak herhangi bir fikrinin veya öngörüsünün olmaması sebebiyle değişkenler arasındaki muhtemel ilişkiyi ortaya çıkarmaya çalışır.

Onaylayıcı faktör analizinde ise araştırmacı tarafından daha önceden belirlenen bir ilişkinin doğruluğunu test etmek amaçlanmaktadır.

Faktör analizi çeşitli aşamalardan oluşan bir analiz tekniğidir. Tipik bir faktör analizinde yer alan aşamalar aşağıdaki şekilde özetlenebilir.

Problem tanımı ve veri toplama aşaması faktör analizi için gerekli olan hazırlık çalışmalarını kapsayan ilk aşamadır. Bu aşamada faktör analizinin amacı ve faktör analizinde kullanılacak olan değişkenlerin teori, mevcut araştırmalar ve araştırmacının bilgi ve tecrübeleri veya yaptığı ön çalışmalar (kalitatif veya kantitatif türdeki çalışmalarla) ışığında geliştirilmesi ve uygun ölçüm araçları ile ölçülmesi ve makul yöntemlerle verilerin toplanması işlemleri yapılmaktadır.

Korelasyon matrisinin oluşturulması aşaması analiz sürecinin başladığı aşama olup, bu aşamada korelasyon matrisi oluşturulur. Korelasyon matrisi faktör analizinde yer alan değişkenler arasındaki ilişkiyi gösteren bir matristir.

Faktör sayısına karar verme aşaması ise, söz konusu veri seti için faktör analizinin uygun olduğuna karar verdikten sonra, oluşturulan korelasyon matrisini baz alarak, faktör çözümünü ortaya koymak amacıyla uygun bir faktör çıkarma

(oluřturma) ynteminin seilmesi ve bařlangı zmnn oluřturulmasını kapsamaktadır.

Faktr eksen dndrme, bařlangı faktr analizi zmne ulařıldıktan sonra (bařlangı faktr matrisi) ortaya ıkan faktrlerin yorumlanması ve isimlendirilmesini kolaylařtırmak iin faktrleri temsil eden eksenlerde eřitli maniplasyonlar veya eksen kaydırmaları yapma yoluna gidilir.

Faktr matrisi faktrler cinsinden standardize edilmiř deęiřkenleri ifade eden katsayıları ierir. Bu katsayılar faktr ykleri olarak adlandırılıp, deęiřkenlerle faktrler arasındaki korelasyonları temsil eder. Mutlak deęer olarak faktr yknn byklę arttıka deęiřken ve faktrn birbiriyle yakından iliřkili olduęunu gstermektedir (Altunıřık, Bayraktaroęlu, 2005).

Faktr analizi pek ok alanda sıkka kullanılan ok deęiřkenli analiz tekniklerinden biridir. Faktr analizi p deęiřkenli bir olayda (p boyutlu uzay) birbiri ile iliřkili deęiřkenleri bir araya getirerek, az sayıda yeni iliřkisiz deęiřkenler bulmayı amalar. Yani, temel bileřenler analizi gibi bir boyut indirgeme ve baęımlılık yapısını yok etme yntemidir.

Faktr analizinde de yine kovaryans matrisi ya da korelasyon matrisi ile iře bařlanır. Bu matrislerden hangisinin kullanılacaęına yine temel bileřenler analizi konusunda verilen uyarılar ıřıęında karar verilir. Ancak genelde korelasyon matrisi kullanılmaktadır. Korelasyon matrisinin faktrleřtirilmesi esasına dayalı faktr analizinde faktrleřtirmede kullanılan pek ok yntem bulunmaktadır.

3. ÇOK DEĞİŞKENLİ DAĞILIMLARIN GEOMETRİSİ

3.1. Verilerin Düzenlenmesi

Bu bölümde çeşitli değişken veya karakterler üzerinde yapılacak ölçümlerin analizi üzerinde durulacaktır. Bu ölçümler (genellikle veri olarak isimlendirilir) bir düzene sokulmalı ve farklı şekillerde ortaya konulmalıdır. Örneğin grafiksel ve tablosal düzenlemeler veri analizinde önemli yardımcılardır. Şimdi veri düzenlemesinin ilk basamağında yer alan öncül kavramları tanımlayacağız.

Çok değişkenli veriler, bir araştırmacı sosyal veya fiziksel bir olayı anlamaya çalıştığında ortaya çıkar, belli bir sayıda $p \geq 1$ değişken veya karakteri seçerek kaydeder. Bu değişkenlerin değerleri her bir birey, veri veya deney için kaydedilir. Çok değişkenli istatistiksel yöntemlerin çoğunda değişkenlerin matrisler şeklinde düzenlenmiş özet bilgileri kullanılarak çözüme ulaşılmaktadır.

x_{ij} ile j . verideki i . değişken ölçümü gösterilsin.

p değişkeni üzerindeki n ölçümü aşağıdaki gibi gösterilebilir.

	ölçüm1	ölçüm2	...	ölçüm j	...	ölçüm n
1.değişken:	x_{11}	x_{12}	...	x_{1j}	...	x_{1n}
2.değişken:	x_{21}	x_{22}	...	x_{2j}	...	x_{2n}
⋮	⋮	⋮		⋮		⋮
i.değişken:	x_{i1}	x_{i2}	...	x_{ij}	...	x_{in}
⋮	⋮	⋮		⋮		⋮
p.değişken:	x_{p1}	x_{p2}	...	x_{pj}	...	x_{pn}

p satır ve n sütundan oluşan matris şeklinde de gösterilir. X matrisi:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pj} & \dots & x_{pn} \end{bmatrix} \quad (3.1)$$

X serisi bütün deęişkenler üzerindeki bütün gözlemlerden oluşan verileri içerir.

3.1.1. Tanımlayıcı istatistikler

Geniş bir veri düzeninde gözlemsel bilgilerin çıkarılması oldukça zordur. Veriler içerisindeki bilgilerin birçoğunu deęerlendirmek için belli özellięi gösteren sayılar üzerinde hesaplar yapılabilir. Bunlara tanımlayıcı istatistikler denir. Örneęin aritmetik ortalama, örnek ortalaması bir tanımlayıcı istatistiktir ve bize konumsal bir ölçüm verir.

Birinci deęişkendeki ölçümlerimiz $x_{11}, x_{12}, \dots, x_{1n}$ olsun. Her p deęişkeni için n ölçümün örnek ortalaması aşıęıdaki gibi hesaplanabilir:

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij} \quad i = 1, 2, \dots, p$$

Burada \bar{x}_i , i 'inci deęişkenin örnek ortalaması olmak üzere, p deęişken için deęişkenlerin örnek varyanslarını aşıęıdaki eşitlikle hesaplarız.

$$s_i^2 = \frac{1}{n} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \quad i = 1, 2, \dots, p$$

s_i^2 yerine matris formatı kullanıldığında ana köşegen üzerindeki elemanları ifade etmesi açısından s_{ii} gösteriminin kullanımı da tercih edilir.

$$s_i^2 = s_{ii} = \frac{1}{n} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \quad i = 1, 2, \dots, p$$

Örnek varyansının karekökü $\sqrt{s_{ii}}$, standart sapma olarak bilinir. Değişkenler arasındaki örnek kovaryansı aşağıdaki eşitlikle hesaplanır.

$$s_{ik} = \frac{1}{n} \sum_{j=1}^n (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k) \quad i = 1, 2, \dots, p \quad k = 1, 2, \dots, p \quad (3.2)$$

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \quad (3.3)$$

Son olarak değişkenler arasındaki doğrusal ilişkinin ölçüsü olarak değişkenlerin varyansları ve kovaryanslarının bir fonksiyonu olan Pearson korelasyon katsayısı aşağıdaki eşitlikle hesaplanır.

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}} \sqrt{s_{kk}}} = \frac{\sum_{j=1}^n (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k)}{\sqrt{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2} \sqrt{\sum_{j=1}^n (x_{kj} - \bar{x}_k)^2}} \quad (3.4)$$

$i=1, 2, \dots, p$ ve $k=1, 2, \dots, p$ her i ve k için $r_{ik} = r_{ki}$

$$r = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (3.5)$$

3.1.2. Pozitif tanımlı matrisler

Simetrik, $k \times k$ boyutlu A matrisi her $x' = [x_1, x_2, \dots, x_k]$ için $0 \leq x'Ax$ oluyorsa A matrisi negatif tanımlı olmayan matristir, A matrisi her $x' \neq 0$ için $0 < x'Ax$ oluyorsa A matrisi pozitif tanımlı matrisidir.

Çok değişkenli verilerin arasındaki ilişki ve değişim konusu, uzaklık ve verilerin çok değişkenli ve normal olarak dağılımı varsayımına dayanır. Karesel uzaklık ve çok değişkenli normal dağılım matris şeklinde gösterilebilir buna da karesel form denir. Sonuç olarak karesel formların çok değişkenli analizde önemli rol oynaması şaşırtıcı değildir. Bu bölümde karesel formları negatif olmayan ve pozitif tanımlı matrisler için kullanılacaktır.

Karesel formları ve simetrik matrisleri içeren sonuçlar birçok durumda spektral ayrışım olarak bilinen simetrik matrislerin genişlemesinin direk sonucudur. $k \times k$ boyutlu simetrik S matrisinin spektral ayrışımı;

$$S = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \dots + \lambda_k e_k e_k' \quad (3.6)$$

Burada $\lambda_1, \lambda_2, \dots, \lambda_k$ S matrisinin özdeğerleri, e_1, e_2, \dots, e_k bu özdeğerlere karşılık gelen normalize edilmiş özvektörlerdir.

Örnek 3.1. S simetrik matrisi aşağıdaki gibi olsun;

$$S = \begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix}$$

S matrisinin özdeğerleri $|S - \lambda I| = 0$ karakteristik denklemin çözümünden elde edilir. Buradan $\lambda_1 = 9$ $\lambda_2 = 9$ $\lambda_3 = 18$ bulunur. Matrisin normalize edilmiş e_1, e_2, e_3 özvektörleri ise $i=1,2,3$ için $S e_i = \lambda_i e_i$ eşitliğinden bulunur.

$$S = \begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{21} \\ e_{31} \end{bmatrix} = 9 \begin{bmatrix} e_{11} \\ e_{21} \\ e_{31} \end{bmatrix}$$

veya;

$$13e_{11} - 4e_{21} + 2e_{31} = 9e_{11}$$

$$-4e_{11} + 13e_{21} - 2e_{31} = 9e_{21}$$

$$2e_{11} - 2e_{21} + 10e_{31} = 9e_{31}$$

Üç bilinmeyenli üç denklemden oluşan denklem sistemi çözüldüğünde $e_{11} = 1$, $e_{21} = 1$, $e_{31} = 0$ bulunur. Böylece normalize edilmiş özvektör bulunur;

$$\begin{aligned} e'_1 &= \left[\frac{1}{\sqrt{1^2+1^2+0^2}} \quad \frac{1}{\sqrt{1^2+1^2+0^2}} \quad \frac{0}{\sqrt{1^2+1^2+0^2}} \right] \\ &= \left[\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}} \quad 0 \right] \end{aligned}$$

$$\begin{aligned} e'_2 &= \left[\frac{1}{\sqrt{1^2+(-1)^2+4^2}} \quad \frac{-1}{\sqrt{1^2+(-1)^2+4^2}} \quad \frac{4}{\sqrt{1^2+(-1)^2+4^2}} \right] \\ &= \left[\frac{1}{\sqrt{18}} \quad \frac{-1}{\sqrt{18}} \quad \frac{-4}{\sqrt{18}} \right] \end{aligned}$$

$$\begin{aligned} e'_3 &= \left[\frac{2}{\sqrt{2^2+(-2)^2+1^2}} \quad \frac{-2}{\sqrt{2^2+(-2)^2+1^2}} \quad \frac{1}{\sqrt{2^2+(-2)^2+1^2}} \right] \\ &= \left[\frac{2}{3} \quad \frac{-2}{3} \quad \frac{1}{3} \right] \end{aligned}$$

S matrisi $S = \lambda_1 e_1 e'_1 + \lambda_2 e_2 e'_2 + \lambda_3 e_3 e'_3$ şeklinde yazılabilir.

$$\begin{aligned} S &= \begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix} = 9 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \left[\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}} \quad 0 \right] + 9 \begin{bmatrix} \frac{1}{\sqrt{18}} \\ \frac{-1}{\sqrt{18}} \\ \frac{-4}{\sqrt{18}} \end{bmatrix} \\ &\quad \left[\frac{1}{\sqrt{18}} \quad \frac{-1}{\sqrt{18}} \quad \frac{-4}{\sqrt{18}} \right] + 18 \begin{bmatrix} \frac{2}{3} \\ \frac{-2}{3} \\ \frac{1}{3} \end{bmatrix} \left[\frac{2}{3} \quad \frac{-2}{3} \quad \frac{1}{3} \right] \end{aligned}$$

Örnek 3.2. Aşağıdaki karesel formun pozitif tanımlı olduğunu gösterelim.

$$3x_1^2 + 2x_2^2 - 2\sqrt{2}x_1x_2$$

Öncelikle karesel formu, matris gösterimi ile yazalım:

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x'Ax$$

A matrisinin özdeğerleri $|A-\lambda I|=0$ veya $(3-\lambda)(2-\lambda)-2=0$ eşitliğinden bulunur. Buradan $\lambda_1=4$ $\lambda_2=1$ bulunur. Spektral ayrışım kullanarak aşağıdaki gibi yazılabilir.

$$A = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' = 4e_1 e_1' + e_2 e_2'$$

e_1 ve e_2 normalize edilmiş $\lambda_1 = 4$ $\lambda_2 = 1$ özdeğerlerine karşılık gelen ortogonal özvektörlerdir. $x' = [x_1 \ x_2]$ vektörü sıfırdan farklı olsun eşitliği sağdan ve soldan çarparsak:

$$x'Ax = 4x'e_1 e_1'x + x'e_2 e_2'x$$

$$y_1 = x'e_1 = e_1'x \text{ ve } y_2 = x'e_2 = e_2'x \text{ olduğundan}$$

$$x'Ax = 4y_1^2 + y_2^2 \geq 0$$

y_1 ve y_2 her ikisi de sıfırdan farklı olduğundan $x'Ax = 4y_1^2 + y_2^2 > 0$ yani A pozitif tanımlıdır. y_1 ve y_2 nin tanımından:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} e_1' \\ e_2' \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\underset{(2 \times 1)}{y} = \underset{(2 \times 2)}{E} \underset{(2 \times 1)}{x} \tag{3.7}$$

Şimdi E' tersine sahip E ortogonal matrisini ele alalım. $x = E'y$ (x sıfırdan farklı vektör ve $0 \neq x = E'y$, $y \neq 0$ iken)

Spektral ayrışmaları kullanmak, $k \times k$ boyutlu A simetrik matrisinin pozitif tanımlı olduğunu göstermekten daha kolaydır, tabii eğer bu A matrisinin bütün özdeğerleri pozitifse, negatif tanımlı olmayan matrislerde tüm özdeğerler sıfırdan büyük veya eşittir.

3.2. Çok Değişkenli Dağılımın Geometrik Yorumu

Çok değişkenli dağılımların geometrik yorumunu incelerken, iki değişkenli dağılımların incelenmesi tercih edilmiştir. Daha fazla değişkenli dağılımların incelenmesi grafiklerinin çiziminin zorluğu ve yorum güclüğü sebebiyle tercih edilmemiştir.

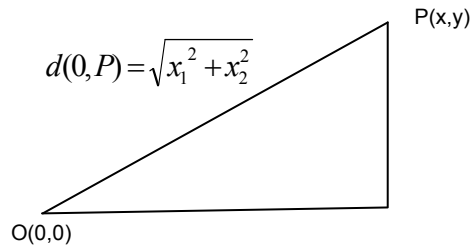
Tek değişkenli dağılımda gözlemlerin bir eksen üzerinde (doğru) belirli bir noktaya (konum ya da ortalama) uzaklığı önem taşır. Noktaların ortalamadan uzaklığının ölçüsü olarak da yaygınlık (genellikle varyans) değerleri kullanılır.

İki değişkenli dağılımda bir birimden alınan iki ölçüm değeri düzlemde bir nokta ifade eder. Bu noktaların düzlem üzerindeki ortalama koordinatına uzaklığına bağlı olarak yayılımın ölçüsü kovaryans matrisiyle ifade edilir.

İki değişkenli dağılımın geometrik yorumunda, düzlem üzerindeki noktaların konumu, merkeze uzaklığı ve yayılımı ile kovaryans matrisi arasındaki ilişkiyi inceleyeceğiz.

3.2.1. Uzaklıklar

İlk bakışta zorlu görünmesine rağmen, birçok çok değişkenli basit bir uzaklık kavramı üzerine kurulmuştur. Düz çizgi veya öklidyen uzaklıklar birbirine benzer. Bu noktada phythagorem teoremine göre düşünersek $P=(x_1,x_2)$ noktasının orijine olan en kısa mesafesi aşağıdaki gibidir:



Şekil 3.1. Bir noktanın orijine olan en kısa uzaklığı

Genel olarak eğer P noktası $P=(x_1, x_2, \dots, x_p)$, p koordinata sahipse P 'den orijine en kısa mesafe aşağıdaki gibi bulunur.

$$d(0, P) = \sqrt{x_1^2 + x_2^2 + \dots + x_p^2}$$

Orijinden başlayıp bütün (x_1, x_2, \dots, x_p) 'ye kadar olan noktaların kareler toplamı aşağıdaki eşitliğe uyarlar.

$$d^2(0, P) = x_1^2 + x_2^2 + \dots + x_p^2 = c^2 \tag{3.8}$$

$P=(x_1, x_2, \dots, x_p)$, $Q=(y_1, y_2, \dots, y_p)$ noktaları arasındaki öklidyen uzaklık aşağıdaki eşitliğe uyarlar.

$$d(P, Q) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \tag{3.9}$$

Seçimler örnek varyansı ve kovaryansına bağlı olduğu için istatistiksel uzaklığın kullanılması tercih edilir.

Burada standart sapma kullanılarak standartlaştırılmış koordinatlar elde edilebilir.

$$x_1^* = \frac{x_1}{\sqrt{s_{11}}} \text{ ve } x_2^* = \frac{x_2}{\sqrt{s_{22}}}$$

$P(x_1, x_2)$ noktasının orijin $O(0,0)$ noktasına olan uzaklığı standartlaştırılmış koordinatlarla hesaplanırsa:

$$d(P,Q) = \sqrt{(x_1^*)^2 + (x_2^*)^2} = \sqrt{\left(\frac{x_1}{\sqrt{s_{11}}}\right)^2 + \left(\frac{x_2}{\sqrt{s_{22}}}\right)^2} = \sqrt{\frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}}} \quad (3.10)$$

(3.10) eşitliği kullanılarak (x_1, x_2) koordinatlarına sahip tüm noktaların orijine olan uzaklığın karesinin c^2 sayısına eşit olduğu gösterilir.

$$\frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}} = c^2 \quad (3.11)$$

(3.11) eşitliği merkezi orijinde olan uzun ve kısa eksenleri koordinat eksenleri olan elips denklemdir.

Sabit Q ve herhangi bir P noktası arasındaki istatistiksel uzaklık hesabı için, (3.11) deki ifade genelleştirilebilir. P ile Q noktaları arasındaki uzaklık:

$$Id(P,Q) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}}} \quad (3.12)$$

İkiden daha fazla boyutlar için bu istatistiksel uzaklığın ifadesi daha kolaydır. P koordinatlı nokta; $P=(x_1, x_2, \dots, x_p)$ ve $Q=(y_1, y_2, \dots, y_p)$ olsun. Q 'nun sabit bir nokta (bu nokta orijin de olabilir) ve koordinat değişkenlerinin birbiri arasında bağımsız

olarak deđiřtiđini varsayalım. $s_{11}, s_{22}, \dots, s_{pp}$, x_1, x_2, \dots, x_p ölçümlerinden oluşturulmuş örnek varyanslar olsun. P ve Q arasındaki istatistiksel uzaklık (3.12) gibi genişletilebilir.

$$Id(P, Q) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}} + \dots + \frac{(x_p - y_p)^2}{s_{pp}}} \quad (3.13)$$

Orijin ile P noktaları arasındaki uzaklık $y_1 = y_2 = \dots = y_p = 0$ verilerek elde edilir. Eđer $s_{11} = \dots = s_{pp}$ ise öklidyen uzaklık formülü geçerlidir (Johnson, 1999).

Örnek 3.3. $X(x_1, x_2) = (2, 4)$ ve $Y(y_1, y_2) = (3, 1)$ noktaları için deđiřik uzaklıkları hesaplayalım.

Öklid uzaklıkları:

Noktaların $O(0, 0)$ noktasına olan öklid uzaklıđı:

X noktasının orijine olan uzaklıđı; $d(0, X) = \sqrt{x_1^2 + x_2^2} = \sqrt{2^2 + 4^2} = \sqrt{20} = 4.47$

Y noktasının orijine olan uzaklıđı; $d(0, Y) = \sqrt{y_1^2 + y_2^2} = \sqrt{3^2 + 1^2} = \sqrt{10} = 3.16$

X ve Y Noktaların birbirine olan öklid uzaklıđı:

X noktasının Y noktasına olan uzaklıđı;

$$d(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} = \sqrt{(2 - 3)^2 + (4 - 1)^2} = \sqrt{10} = 3.16$$

İstatistiksel uzaklıklar

$s_{11} = 4, s_{22} = 1$ olmak üzere

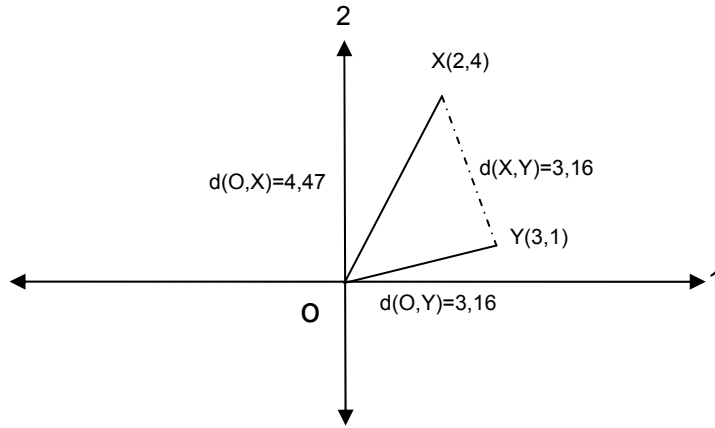
Noktaların $O(0, 0)$ noktasına olan istatistiksel uzaklıđı:

$$X \text{ noktasının orijine olan uzaklığı; } \sqrt{\frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}}} = \sqrt{\frac{2^2}{4} + \frac{4^2}{1}} = \sqrt{17} = 4.12$$

$$Y \text{ noktasının orijine olan uzaklığı; } \sqrt{\frac{y_1^2}{s_{11}} + \frac{y_2^2}{s_{22}}} = \sqrt{\frac{3^2}{4} + \frac{1^2}{1}} = \sqrt{3.25} = 1.80$$

X noktasının Y noktasına olan istatistiksel uzaklığı:

$$Id(X, Y) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}}} = \sqrt{\frac{(2-3)^2}{4} + \frac{(4-1)^2}{1}} = 3.041$$



Şekil 3.2. X ve Y noktalarının birbirine olan istatistiksel uzaklığı

Örnek 3.4. $S = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ olmak üzere $s_{11} = 4$, $s_{22} = 1$ olan (x_1, x_2) ölçümleri için

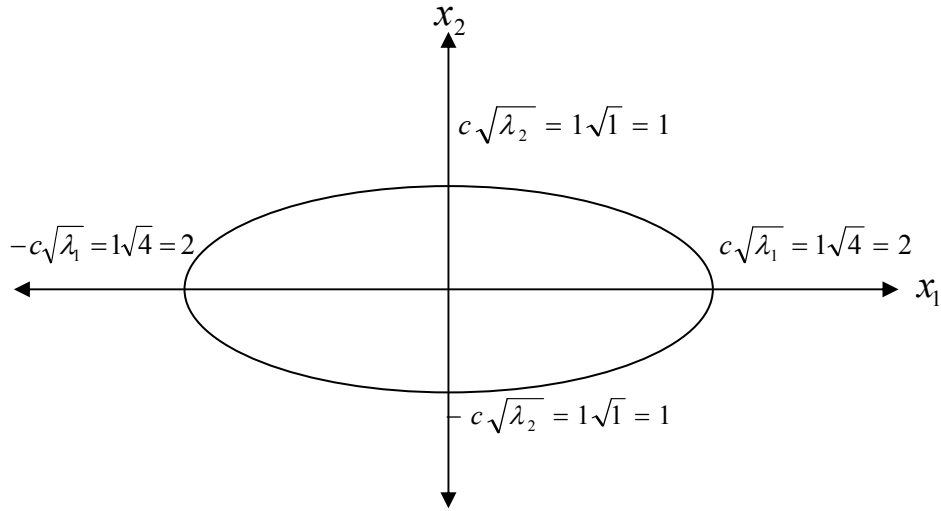
$\bar{x}_1 = \bar{x}_2 = 0$ veriliyor. $P(x_1, x_2)$ noktasının orijin $O(0,0)$ noktasına olan uzaklığı aşağıdaki gibidir.

$$d^2(O, P) = \frac{x_1^2}{4} + \frac{x_2^2}{1} = 1$$

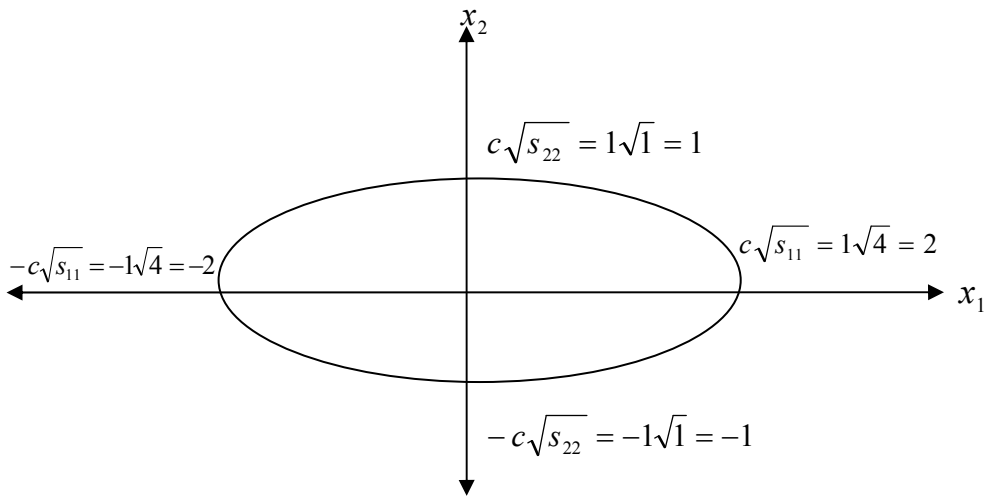
Bu eşitlik merkezi orijinde ve kısa eksenini x_2 , uzun eksenini x_1 eksenini boyunca uzanan elips denklemdir.

$$\frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}} = c^2 \text{ olduğundan } \sqrt{4} = 2 \text{ ve } \sqrt{1} = 1 \text{ değerleri elipsin yarım kısa eksen}$$

uzunluğu ve yarım uzun eksen uzunluklarıdır.



Şekil 3.3. Merkezi orijinde olan ve yarım kısa eksen ve yarım uzun eksen uzunlukları $c\sqrt{\lambda_i}$ olan elips



Şekil 3.4. Merkezi orijinde olan ve yarım kısa eksen ve yarım uzun eksen uzunlukları $c\sqrt{s_{ii}}$ olan elips

3.2.2. Bir noktanın orijine olan uzaklığı

$0 < d^2 = x'Ax$ ($x \neq 0$) ifadeden de görüldüğü gibi $p \times p$ boyutlu A matrisi pozitif tanımlıdır. Yani uzaklık pozitif tanımlı karesel form olan $x'Ax$ den bulunur. Tersisi de doğrudur yani karesel form matris şeklinde de yazılabilir. Herhangi bir $\mu' = [\mu_1, \dots, \mu_p]$ için genel ifade $(x - \mu)'A(x - \mu)$ yazılabilir. A matrisinin özdeğer ve özvektörlerine bağlı olarak pozitif tanımlı kuadratik formun mesafesi bize geometrik olarak da yorum getirir. Örneğin $p = 2$ ve $x' = [x_1, x_2]$ noktasının orijine olan uzaklığı,

$$x'Ax = a_{11}x_1^2 + a_{12}x_2^2 + 2a_{12}x_1x_2 = c^2 \quad (3.14)$$

şeklindedir.

Spektral açılımlar;

$$A = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2'$$

$$x'Ax = \lambda_1 (x_1' e_1)^2 + \lambda_2 (x_2' e_2)^2$$

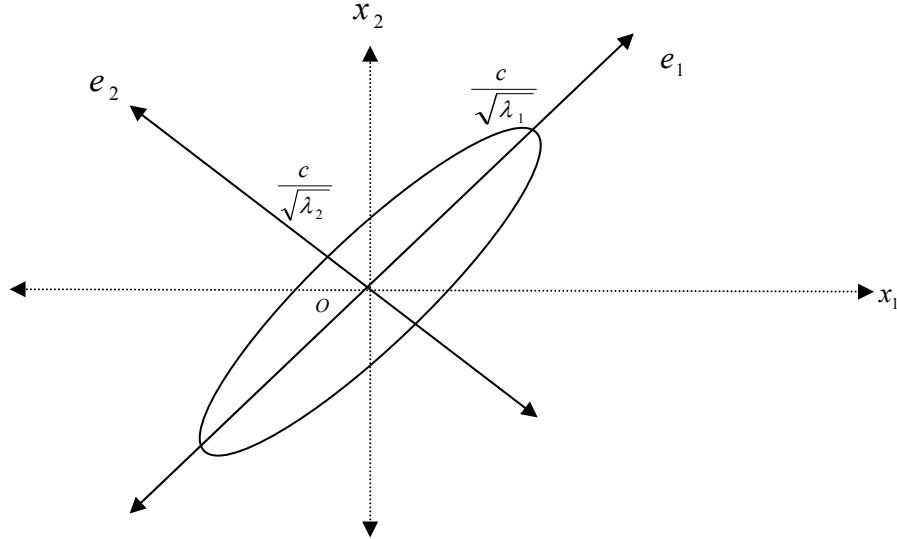
dır.

$c^2 = \lambda_1 y_1^2 + \lambda_2 y_2^2$ elips denklemdir. ($y_1 = x'e_1 = e_1'x$ ve $y_2 = x'e_2 = e_2'x$ ve $\lambda_1, \lambda_2 > 0$) burada $x = c\lambda_1^{-1/2}e_1$ $x'Ax = \lambda_1(c\lambda_1^{-1/2}e_1'e_1)^2 = c^2$ denklemini sağlar. Benzer olarak $x = c\lambda_2^{-1/2}e_2$ de denklemi sağlar ve e_2 doğrultusundaki uzaklığı verir. c uzaklığındaki noktalar bir elips üzerinde dağılır. Bu elipsin eksenleri de A matrisinin özvektörleridir. Bu özvektörler de özdeğerlerin karekökü ile orantılıdır.

Eğer $p > 2$ ise $x' = [x_1, x_2, \dots, x_p]$ noktaları için $x'Ax = \lambda_1 (x_1' e_1)^2 + \lambda_2 (x_2' e_2)^2 + \dots + \lambda_p (x_p' e_p)^2 = c^2$ eşitliği yazılabilir. Bu eşitlik; merkezi orijinde, eksenleri A matrisinin özvektörleri olan ve $\lambda_1, \lambda_2, \dots, \lambda_p$, A

matrisinin özdeğerleri olmak üzere, elipsin merkezinden e_i yönündeki yarı uzunluk

$\frac{c}{\sqrt{\lambda_i}}$, $i=1,2,3,\dots,p$ olan elips denklemdir.



Şekil 3.5. Merkezi orijinde, e_i yönündeki yarı uzunluk $\frac{c}{\sqrt{\lambda_i}}$ olan elips

3.2.3. Gözlemlerin dağılımının geometrisi

P boyutlu normal dağılıma sahip bir x gözlem noktasının, ortalama μ 'ya olan uzaklığı için $(x - \mu)' \Sigma^{-1} (x - \mu) = c^2$ yazılabilir. Burada elipsin merkezi μ ve eksenleri $\pm c \sqrt{\lambda_i} e_i$ dir. $i = 1, 2, \dots, p$ için $\sum e_i = \lambda_i e_i$, $(x - \mu)' \Sigma^{-1} (x - \mu) = c^2$ eşitliğinde $c^2 = \chi_p^2(\alpha)$ yazılabilir. $(x - \mu)' \Sigma^{-1} (x - \mu) \leq \chi_p^2(\alpha)$

Örnek 3.5. Mikrodalga fırınlardan, kapı kapalı ve açıkken alınan radyasyon verileri Tablo 3.1.'dedir (Johnson, 1999).

Tablo 3.1. mikrodalga fırınlardan, kapı kapalı ve açıkken alınan radyasyon verileri

Fırın numarası	Kapı kapalıyken ölçümü	Kapı açıkken ölçümü	x_1	x_2
1	0.15	0.3	0.622332977	0.740082804
2	0.16	0.4	0.622332978	0.740082805
3	0.17	0.5	0.622332979	0.740082806
4	0.18	0.6	0.622332980	0.740082807
5	0.19	0.7	0.622332981	0.740082808
6	0.20	0.8	0.622332982	0.740082809
7	0.21	0.9	0.622332983	0.740082810
8	0.22	0.10	0.622332984	0.740082811
9	0.23	0.11	0.622332985	0.740082812
10	0.24	0.12	0.622332986	0.740082813
11	0.25	0.13	0.622332987	0.740082814
12	0.26	0.14	0.622332988	0.740082815
13	0.27	0.15	0.622332989	0.740082816
14	0.28	0.16	0.622332990	0.740082817
15	0.29	0.17	0.622332991	0.740082818
16	0.30	0.18	0.622332992	0.740082819
17	0.31	0.19	0.622332993	0.740082820
18	0.32	0.20	0.622332994	0.740082821
19	0.33	0.21	0.622332995	0.740082822
20	0.34	0.22	0.622332996	0.740082823
21	0.35	0.23	0.622332997	0.740082824
22	0.36	0.24	0.622332998	0.740082825
23	0.37	0.25	0.622332999	0.740082826
24	0.38	0.26	0.622333000	0.740082827
25	0.39	0.27	0.622333001	0.740082828
26	0.40	0.28	0.622333002	0.740082829
27	0.41	0.29	0.622333003	0.740082830
28	0.42	0.30	0.622333004	0.740082831
29	0.43	0.31	0.622333005	0.740082832
30	0.44	0.32	0.622333006	0.740082833
31	0.45	0.33	0.622333007	0.740082834
32	0.46	0.34	0.622333008	0.740082835
33	0.47	0.35	0.622333009	0.740082836
34	0.48	0.36	0.622333010	0.740082837
35	0.49	0.37	0.622333011	0.740082838
36	0.50	0.38	0.622333012	0.740082839
37	0.51	0.39	0.622333013	0.740082840
38	0.52	0.40	0.622333014	0.740082841
39	0.53	0.41	0.622333015	0.740082842
40	0.54	0.42	0.622333016	0.740082843
41	0.55	0.43	0.622333017	0.740082844
42	0.56	0.44	0.622333018	0.740082845

$$x_1 = (\text{ kapı kapalı iken radyasyon ölçümleri})^{1/4}$$

$$x_2 = (\text{ kapı açıkken radyasyon ölçümleri})^{1/4}$$

Mikrodalga fırınların radyasyon değerleri Tablo 3.1.'de verilmiştir. $n=42$ gözlem için aşağıdaki değerler bulunmuştur.

$$\bar{x} = \begin{bmatrix} 0.564 \\ 0.603 \end{bmatrix}, \quad S = \begin{bmatrix} 0.0144 & 0.0117 \\ 0.0117 & 0.0146 \end{bmatrix}, \quad S^{-1} = \begin{bmatrix} 203.018 & -163.391 \\ -163.391 & 200.228 \end{bmatrix}$$

S matrisi için özdeğerler ve bu özdeğerlere karşılık gelen özvektörler:

$$\lambda_1 = 0.026, \quad e'_1 = [0.704 \quad 0.710]$$

$$\lambda_2 = 0.002, \quad e'_2 = [-0.710 \quad 0.704]$$

$(x - \mu)' \Sigma^{-1} (x - \mu) = c^2$ eşitliğinde $c^2 = \chi_p^2(\alpha)$ belirlenmiş uzaklığın karesi olmak üzere bu mesafeden daha kısa mesafedeki uzaklıkların karesi şu eşitliğin sağlanması ile belirlenir.

$$(x - \mu)' \Sigma^{-1} (x - \mu) \leq \chi_p^2(1 - \alpha)$$

%95 elips kontur için eksen uzunlukları

$$c^2 = \chi_2^2(0.95) = 5.991 \text{ ise } c = \sqrt{5.991} = 2.4476$$

$$c\sqrt{\lambda_1} = 2.4476\sqrt{0.026} = (2.4476)(0.161) = 0.394$$

$$c\sqrt{\lambda_2} = 2.4476\sqrt{0.002} = (2.4476)(0.044) = 0.107$$

%90 elips kontur için eksen uzunlukları

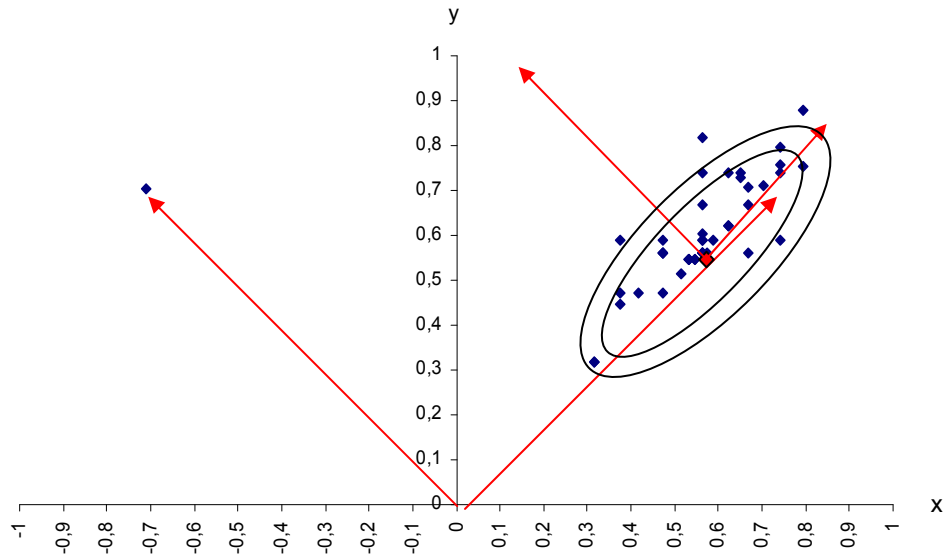
$$c^2 = \chi_2^2(0.90) = 4.61 \text{ ise } c = \sqrt{4.61} = 2.147$$

$$c\sqrt{\lambda_1} = 2.147\sqrt{0.026} = (2.147)(0.161) = 0.345$$

$$c\sqrt{\lambda_2} = 2.147\sqrt{0.002} = (2.147)(0.044) = 0.094$$

Tablo 3.2.Noktaların geometrik uzaklıkları

Ölçüm no	x_1	x_2	Ölçüm noktasının ortalamaya uzaklığının karesi	$\chi^2(0.90)$	$\chi^2(0.95)$
1	0.622332977	0.740082804	1.846	4.61	5.991
2	0.547722558	0.547722558	0.368	4.61	5.991
3	0.651355562	0.740082804	0.402	4.61	5.991
4	0.562341325	0.562341325	0.306	4.61	5.991
5	0.472870805	0.562341325	0.813	4.61	5.991
6	0.588566191	0.588566191	0.276	4.61	5.991
7	0.531829590	0.547722558	0.239	4.61	5.991
8	0.472870805	0.562341325	0.813	4.61	5.991
9	0.531829590	0.547722558	0.239	4.61	5.991
10	0.562341325	0.562341325	0.306	4.61	5.991
11	0.514368672	0.514368672	0.633	4.61	5.991
12	0.376060309	0.472870805	2.578	4.61	5.991
13	0.316227766	0.316227766	5.712	4.61	5.991
14	0.562341325	0.819036259	9.483	4.61	5.991
15	0.562341325	0.588566191	0.033	4.61	5.991
16	0.562341325	0.668740305	0.908	4.61	5.991
17	0.376060309	0.447213595	2.469	4.61	5.991
18	0.562341325	0.562341325	0.306	4.61	5.991
19	0.316227766	0.316227766	5.712	4.61	5.991
20	0.795270729	0.880111737	5.291	4.61	5.991
21	0.562341325	0.588566191	0.033	4.61	5.991
22	0.472870805	0.562341325	0.813	4.61	5.991
23	0.416179145	0.472870805	1.545	4.61	5.991
24	0.472870805	0.472870805	1.200	4.61	5.991
25	0.622332977	0.622332977	0.392	4.61	5.991
26	0.562341325	0.740082804	3.850	4.61	5.991
27	0.622332977	0.622332977	0.392	4.61	5.991
28	0.547722558	0.547722558	0.368	4.61	5.991
29	0.531829590	0.547722558	0.239	4.61	5.991
30	0.651355562	0.727427153	1.099	4.61	5.991
31	0.562341325	0.562341325	0.306	4.61	5.991
32	0.668740305	0.562341325	3.935	4.61	5.991
33	0.575901449	0.562341325	0.513	4.61	5.991
34	0.740082804	0.740082804	2.162	4.61	5.991
35	0.376060309	0.588566191	6.346	4.61	5.991
36	0.668740305	0.707106781	0.832	4.61	5.991
37	0.668740305	0.668740305	0.837	4.61	5.991
38	0.740082804	0.795270729	2.631	4.61	5.991
39	0.740082804	0.757928931	2.181	4.61	5.991
40	0.795270729	0.752120619	4.029	4.61	5.991
41	0.740082804	0.588566191	7.146	4.61	5.991
42	0.472870805	0.588566191	1.307	4.61	5.991



Şekil 3.6. Gözlemlerin dağılımı için %90 ve %95 düzeyinde elips konturlar.

3.2.4. Ortalamaların dağılımının geometrisi

P boyutlu normal dağılıma sahip bir \bar{x} gözlem noktasının, ortalama μ 'ya olan uzaklığı, p değişken ve n gözlem için $n(\bar{x} - \mu)'S^{-1}(\bar{x} - \mu) \leq \frac{p(n-1)}{(n-p)} F_{p,n-p}(1-\alpha)$ yazılabilir.

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j, \quad S = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})', \text{ ve } x_1, x_2, \dots, x_n$$

S matrisinin λ_i özdeğerleri ve e_i özvektörleri için elipsin eksenleri aşağıdaki gibi bulunur.

$$n(\bar{x} - \mu)'S^{-1}(\bar{x} - \mu) \leq c^2 = \frac{p(n-1)}{(n-p)} F_{p,n-p}(1-\alpha)$$

$$\sqrt{\lambda_i} c / \sqrt{n} = \sqrt{\lambda_i} \sqrt{p(n-p) F_{p,n-p}(1-\alpha) / n(n-p)} \quad (3.15)$$

Merkezi \bar{x} olan elipsin eksenlerinin uzunlukları

$$\begin{aligned} & -\sqrt{\lambda_i} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(1-\alpha)} e_i \quad i = 1, 2, \dots, p \text{ için, } Se_i = \lambda_i e_i \end{aligned} \quad (3.16)$$

%95 güven aralığında (μ_1, μ_2) 'nin bütün değerleri elipsin içinde kalır.

$$42 \begin{bmatrix} 0.564 - \mu_1 & 0.603 - \mu_2 \\ -163.391 & 200.228 \end{bmatrix} \begin{bmatrix} 0.564 - \mu_1 \\ 0.603 - \mu_2 \end{bmatrix}$$

$$\leq \frac{2 \cdot (41)}{40} F_{2,40}(0.05)$$

$$F_{2,40}(0.05) = 3.23$$

$$42 \cdot (203.018)(0.564 - \mu_1)^2 + 42(200.228)(0.63 - \mu_2)^2 - 84(163.391)(0.564 - \mu_1)(0.603 - \mu_2) \leq 6.62$$

$\mu' = [0.562 \quad 0.589]$ noktasının, güven aralığında olup olmadığını hesaplayalım.

$$42 \cdot (203.018)(0.564 - 0.562)^2 + 42(200.228)(0.63 - 0.589)^2 - 84(163.391)(0.564 - 0.562)(0.603 - 0.589) \leq 6.62$$

$\mu' = [0.562 \quad 0.589]$ noktasının elipsin içinde olduğu sonucuna varılır.

$$H_0 : \mu = \begin{bmatrix} 0.562 \\ 0.589 \end{bmatrix}, \quad H_1 : \mu \neq \begin{bmatrix} 0.562 \\ 0.589 \end{bmatrix} \quad \alpha = 0.05 \text{ anlam düzeyinde}$$

Elipsin merkezi $\bar{x} = \begin{bmatrix} 0.564 \\ 0.603 \end{bmatrix}$ noktasıdır. Elipsin merkezinden elipsin en uzak

noktasına olan uzaklığı ve en kısa noktasına olan uzaklığı aşağıdaki gibi bulunur.

$$\sqrt{\lambda_1} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(1-\alpha)} = \sqrt{0.026} \sqrt{\frac{2(41)}{42(40)}} (3.23) = 0.064$$

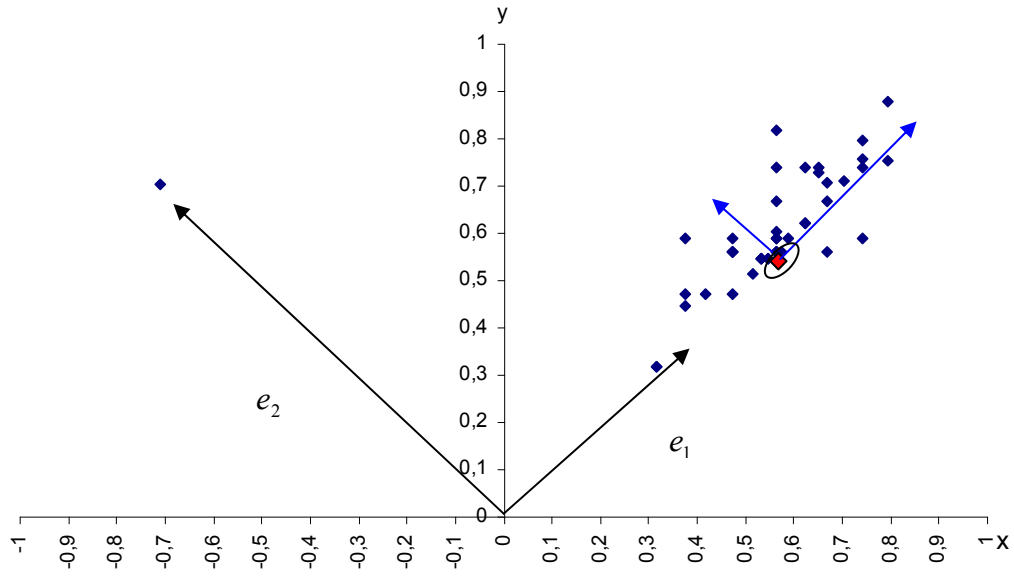
$$\sqrt{\lambda_2} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(1-\alpha)} = \sqrt{0.002} \sqrt{\frac{2(41)}{42(40)}} (3.23) = 0.018$$

Elipsin merkezini belirledikten sonra elipsin merkezi orijin gibi düşünülerek elipsin eksenleri $e'_1 = [0.704 \quad 0.710]$ $e'_2 = [-0.710 \quad 0.704]$ vektörlerinin yönleri

doğrultusunda çizilir. Ayrıca elipsin eksenlerinin uzunlukları arasında bir oran vardır. Bu oran aşağıdaki gibi bulunur.

$$\frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} = \frac{0.161}{0.045} = 3.6$$

Elipsin uzun ekseninin uzunluğu kısa ekseninin uzunluğundan 3.6 katıdır.



Şekil 3.7. n gözlem ortalamasının dağılımı ($n=42$)

3.2.5. Kovaryans matrisinin özdeğerleri ile elemanları arasındaki fonksiyonel ilişki

Değişken sayısı $p=2$ için,

$$S = \begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix}$$

dir.

S matrisinin özdeğerleri $|S-\lambda I|=0$ karakteristik denkleminin çözümünden elde edilir.

$$\begin{vmatrix} s_{11} - \lambda & s_{12} \\ s_{12} & s_{22} - \lambda \end{vmatrix} = 0$$

$$(s_{11} - \lambda)(s_{22} - \lambda) - s_{12}^2 = 0$$

$$\lambda^2 - (s_{11} + s_{22})\lambda + s_{11}s_{22} - s_{12}^2 = 0$$

denkleminin kökleri

$$\lambda_1 = \frac{(s_{11} + s_{22}) - \sqrt{s_{11}^2 + s_{22}^2 - 2s_{11}s_{22} + 4s_{12}^2}}{2}$$

$$\lambda_2 = \frac{(s_{11} + s_{22}) + \sqrt{s_{11}^2 + s_{22}^2 - 2s_{11}s_{22} + 4s_{12}^2}}{2}$$

olarak bulunur.

Örneğin; $\mu = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $S = \begin{bmatrix} 7 & -2 \\ -2 & 4 \end{bmatrix}$ olmak üzere kovaryans matrisinin

özdeğerlerini bulalım.

$$\begin{aligned} \lambda_1 &= \frac{(s_{11} + s_{22}) + \sqrt{s_{11}^2 + s_{22}^2 - 2s_{11}s_{22} + 4s_{12}^2}}{2} \\ &= \frac{(7 + 4) + \sqrt{7^2 + (-2)^2 - 2 \cdot 7 \cdot 4 + 4(-2)^2}}{2} = \frac{11 + \sqrt{25}}{2} = \frac{11 + 5}{2} = 8 \end{aligned}$$

$$\begin{aligned} \lambda_2 &= \frac{(s_{11} + s_{22}) - \sqrt{s_{11}^2 + s_{22}^2 - 2s_{11}s_{22} + 4s_{12}^2}}{2} \\ &= \frac{(7 + 4) - \sqrt{7^2 + (-2)^2 - 2 \cdot 7 \cdot 4 + 4(-2)^2}}{2} = \frac{11 - \sqrt{25}}{2} = \frac{11 - 5}{2} = 3 \end{aligned}$$

Kovaryans matrisinin özdeğerlerine bağlı olarak yaygınlığının belirlenmesi için elipsin uzun ve kısa eksen uzunlukları,

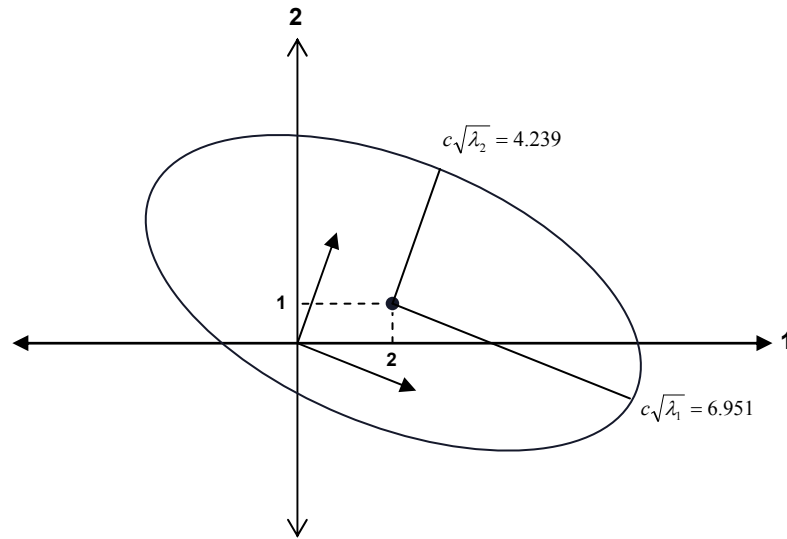
$$c\sqrt{\lambda_1} = 2.4476\sqrt{8} = (2.4476)(2.828) = 6.921$$

$$c\sqrt{\lambda_2} = 2.4476\sqrt{3} = (2.4476)(4.236) = 4.236$$

olarak hesaplanır.

$$e_1 = \begin{bmatrix} 0.894 \\ -0.447 \end{bmatrix} \text{ özvektörüne paralel olarak } 6.921, \quad e_2 = \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix} \text{ özvektörüne}$$

paralel olarak 4.236 yarı eksen uzunlukları ile elipsi Şekil 3.8.'deki gibi çizilebilir.



Şekil 3.8. Merkezi orijinde olmayan elips

4. KOVARYANS MATRİSİNİN KARAKTERİSTİK KÖKLERİ İÇİN GÜVEN ARALIĞI

Bu bölümde örneklem kovaryans matrisinin karakteristik kökleri için güven aralıkları araştırıldı. Monte Carlo simülasyonlarının sonuçlarına dayanan başka bir güven aralığı öngörüldü. Bu güven aralıkları simülasyona dayandığı için yığın kovaryans matrisinin en büyük ve en küçük karakteristik kökleri arasındaki fark büyük olduğu zaman bu yeni güven aralıklarının diğerlerine nazaran daha dar olduğu gösterildi.

S örnek kovaryans matrisi, p boyutlu normal dağılıma sahip, ortalaması μ ve kovaryans matrisi Σ olan N büyüklüğünde bir örneklem olsun. nS Wishart dağılımına sahip olsun $W_p(n, \Sigma)$. Burada $n = N - 1$ dir. Σ 'nin karakteristik kökleri $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$ ve S nin karakteristik kökleri $l_1 > l_2 > \dots > l_p > 0$ olsun. Anderson (2003), Σ 'nin karakteristik kökleri için güven aralığını (4.1) ile formüle etmiştir.

$$\frac{l_p}{u} \leq \lambda_p \leq \lambda_1 \leq \frac{l_1}{l} \quad (4.1)$$

$P_r \{nl \leq \chi_n^2\} P_r \{\chi_{n-p+1}^2 \leq nu\} = 1 - \varepsilon$ burada χ_f^2 le f serbestlik dereceli χ^2 dağılımına sahip değişken gösterilmektedir. Anderson birçok güven aralıkları içinde en dar olanının, (4.1) ile gösterilen güven aralığı olduğunu göstermiştir.

Fakat birçok durumda (4.1) den daha dar güven aralığının bulunabileceği gösterilmiştir. Bu yeni güven aralıklarının daha dar olduğu simülasyonlarla gösterilmiş bu güven aralıkları hakkındaki temel sonuçlar verilmiştir. Simülasyonlar için “SPSS Programı” kullanılmıştır.

4.1. En Küçük ve En Büyük Karakteristik Kökler için Güven Aralığı

Yığın kovaryans matrisinin karakteristik kökü için güven aralıkları normallik varsayımı altında araştırılmaktadır. Fakat, tüm karakteristik kökler için güven aralığı elde etmek gerçekten zordur.

Örneklem kovaryans matrisinin tüm karakteristik köklerinin yoğunluk fonksiyonu (4.2) de verilmiştir (James, 1960).

$$\left(\frac{n}{2}\right)^{np/2} \frac{\pi^{p^2/2} (\det \Sigma)^{-n/2}}{\Gamma_p\left(\frac{1}{2}n\right)\Gamma_p\left(\frac{1}{2}p\right)} \prod_{i=1}^p l_i^{(n-p-1)} \prod_{i<j}^p (l_i - l_j) {}_0F_0^{(p)}\left(-\frac{1}{2}nL, \Sigma^{-1}\right) \quad (4.2)$$

Burada $L = \text{diag}(l_1, l_2, \dots, l_p)$ ve ${}_0F_0^{(p)}\left(-\frac{1}{2}nL, \Sigma^{-1}\right)$ aşağıdaki gibi hipergeometrik fonksiyondur:

$${}_0F_0^{(p)}\left(-\frac{1}{2}nL, \Sigma^{-1}\right) = \sum_{k=0}^{\infty} \sum_K \frac{C_K\left(-\frac{1}{2}nL\right) C_K(\Sigma^{-1})}{k! C_K(I_p)} \quad (4.3)$$

Bu şekilde l_1 ve l_p 'nin marjinal dağılımları elde edilebilir. Fakat hipergeometrik fonksiyonların açılımlarından dolayı, tüm karakteristik kökler için güven aralığı elde etmek mümkün değildir.

Yığın kovaryans matrisinin diagonal ve tüm karakteristik köklerin farklı olduğu bilindiğinde, geniş hacimli örneklem için hipergeometrik fonksiyonun tahmini aşağıdaki gibidir.(Anderson 1965).

$${}_0F_0^{(p)}\left(-\frac{1}{2}nL, \Sigma^{-1}\right) \approx \frac{\Gamma_p\left(\frac{1}{2}p\right)}{\pi^{p^2/2}} \exp\left(-\frac{1}{2}n \sum_{i=1}^p \frac{l_i}{\lambda_i}\right) \prod_{i<j} \left(\frac{2\pi}{nc_{ij}}\right)^{1/2} \quad (4.4)$$

Burada $c_{ij} = (l_i - l_j)(\lambda_i - \lambda_j)/(\lambda_i \lambda_j)$ 'dir. Şu durumda hala tüm karakteristik kökler için güven aralığı elde etmek güçtür.

Karakteristik köklerin fonksiyonunun dağılımı için asimptotik açılımlar kullanılabilir ancak çok karmaşık olduğu için güven aralığı elde edemeyiz. Limitsel dağılım kullanılarak güven aralığı elde edilebilir, ancak bu dağılım, limitsel dağılıma yavaş bir şekilde yakınsadığı için, bu güven aralığı sadece büyük örneklem hacmi için kullanılabilir.

l_1 ve l_p 'nin yakınsak dağılımları normallik varsayımı altında elde edilmiştir. Bu dağılımın tahmini küçük hacimli örneklerde dahi doğru yanıt verir ve bu dağılım kısa bir formdadır. Yakınsak dağılımlar kullanılarak, tüm karakteristik kökler için yeni güven aralığı türetilir (Sakaori, Yamada, 2007).

$$\sqrt{n} (l_1^{1/3} - \lambda_1^{1/3}) / \left(\frac{\sqrt{2}}{3} \lambda_1^{1/3} \right) \quad (4.5)$$

(4.5) ile gösterilen fonksiyonun limit dağılımı normal dağılıma sahiptir. Bu yüzden bu dağılımı (4.6) gibi yazabiliriz.

$$\Pr \left(\lambda_1 \leq l_1 / \left(1 - \frac{1}{3} \sqrt{\frac{2}{n}} z_\alpha \right)^3 \right) = 1 - \alpha \quad (4.6)$$

burada, standart normal dağılımın üstten $100\alpha\%$ nıncı değere karşılık gelen z_α değerini gösterir. Ayrıca limit dağılımı, sınırlı normal dağılımdır.

$$\sqrt{n} \left(l_p^{(p+2)/4} - \lambda_p^{(p+2)/4} \right) / \left(\frac{\sqrt{2}(p+2)}{4} \lambda_p^{(p+2)/4} \right) \quad (4.7)$$

bu yüzden tahmini (4.8) gibi yazabiliriz.

$$\Pr\left(l_p / \left(1 + \frac{p+2}{4} \sqrt{\frac{2}{n} z_\alpha}\right)^{4/(p+2)} \leq \lambda_p\right) = 1 - \alpha \quad (4.8)$$

Ayrıca (4.5)'in hassaslığının, $p=2$ olduğunda diğer tahminlerden daha iyi olduğu sonucuna varılmıştır.

Bu bölümde en küçük ve en büyük karakteristik köklerin korelasyon katsayısını bulmak için, l_1 ve l_p için korelasyon katsayısının karmaşık açılımları yapıldı.

Burada $\Sigma = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ ve $S = \Lambda + \frac{1}{\sqrt{n}}V$ olsun. Karakteristik köklerin reel değerli fonksiyonu için karmaşık ve zorlu bir açılım kullanımı Konishi (1977) tarafından yapıldı.

$$\begin{aligned} & (l_1 - E(l_1))(l_p - E(l_p)) \\ &= (\lambda_1 - E(l_1))(\lambda_p - E(l_p)) + \frac{1}{n^2} F_1 + \frac{1}{n} F_2 + \frac{1}{n^{\frac{3}{2}}} F_3 + \frac{1}{n^2} F_4 + \dots, \end{aligned} \quad (4.9)$$

Burada;

$$\begin{aligned} F_1 &= v_{11}(\lambda_p - E(l_p)) + v_{pp}(\lambda_1 - E(l_1)) \\ F_2 &= \sum_{\alpha \neq 1} \lambda_{1\alpha} v_{1\alpha}^2 (\lambda_p - E(l_p)) + \sum_{\alpha \neq p} \lambda_{p\alpha} v_{p\alpha}^2 (\lambda_1 - E(l_1)) + v_{11} v_{pp} \\ F_3 &= \sum_{\alpha \neq 1} \sum_{\beta \neq 1} \lambda_{1\alpha} \lambda_{1\beta} v_{1\alpha} v_{1\beta} v_{\alpha\beta} (\lambda_p - E(l_p)) + \sum_{\alpha \neq p} \sum_{\beta \neq p} \lambda_{p\alpha} \lambda_{p\beta} v_{p\alpha} v_{p\beta} v_{\alpha\beta} (\lambda_1 - E(l_1)) \\ & - \sum_{\alpha \neq 1} \lambda_{1\alpha}^2 v_{11} v_{1\alpha}^2 (\lambda_p - E(l_p)) - \sum_{\alpha \neq p} \lambda_{p\alpha}^2 v_{pp} v_{p\alpha}^2 (\lambda_1 - E(l_1)) + v_{11} \sum_{\alpha \neq 1} \lambda_{1\alpha} v_{1\alpha}^2 + v_{pp} \sum_{\alpha \neq p} \lambda_{p\alpha} v_{p\alpha}^2 \end{aligned}$$

$$\begin{aligned}
 F_4 = & \sum_{\alpha \neq 1} \sum_{\beta \neq 1} \sum_{\gamma \neq 1} \lambda_{1\alpha} \lambda_{1\beta} \lambda_{1\gamma} \nu_{1\alpha} \nu_{\alpha\beta} \nu_{\beta\gamma} \nu_{\gamma 1} (\lambda_p - E(l_p)) \\
 & + \sum_{\alpha \neq p} \sum_{\beta \neq p} \sum_{\gamma \neq p} \lambda_{p\alpha} \lambda_{p\beta} \lambda_{p\gamma} \nu_{p\alpha} \nu_{\alpha\beta} \nu_{\beta\gamma} \nu_{\gamma p} (\lambda_1 - E(l_1)) \\
 & - 2 \sum_{\alpha \neq 1} \sum_{\beta \neq 1} \lambda_{1\alpha} \lambda_{1\beta}^2 \nu_{11} \nu_{1\alpha} \nu_{\alpha\beta} \nu_{\beta 1} (\lambda_p - E(l_p)) \\
 & - 2 \sum_{\alpha \neq p} \sum_{\beta \neq p} \lambda_{p\alpha} \lambda_{p\beta}^2 \nu_{pp} \nu_{p\alpha} \nu_{\alpha\beta} \nu_{\beta p} (\lambda_1 - E(l_1)) \\
 & + \sum_{\alpha \neq 1} \lambda_{1\alpha}^3 \nu_{1\alpha}^2 \nu_{11}^2 (\lambda_p - E(l_p)) + \sum_{\alpha \neq p} \lambda_{p\alpha}^3 \nu_{p\alpha}^2 \nu_{pp}^2 (\lambda_1 - E(l_1)) \\
 & + \nu_{11} \nu_{pp} \left(\sum_{\alpha \neq 1} \lambda_{1\alpha}^2 \nu_{1\alpha}^2 + \sum_{\alpha \neq p} \lambda_{p\alpha}^2 \nu_{p\alpha}^2 \right) + \left(\sum_{\alpha \neq 1} \lambda_{1\alpha} \nu_{1\alpha}^2 \right)^2 \\
 & + \nu_{pp} \sum_{\alpha \neq 1} \sum_{\beta \neq 1} \lambda_{1\alpha} \lambda_{1\beta} \nu_{1\alpha} \nu_{\alpha\beta} \nu_{\beta 1} + \nu_{11} \sum_{\alpha \neq p} \sum_{\beta \neq p} \lambda_{p\alpha} \lambda_{p\beta} \nu_{p\alpha} \nu_{\alpha\beta} \nu_{\beta p}
 \end{aligned}$$

dır. Burada $V = (\nu_{ij})$ ve $\lambda_{ij} = (\lambda_i - \lambda_j)^{-1}$ dir. Ayrıca $\alpha = 1$ ve p için $(l_p - \lambda_p)$ nin açılımını ve $\nu_{\alpha p}$ nin açılımını aşağıdaki gibi elde edebiliriz.

$$(l_p - \lambda_p) = \frac{1}{n^2} \nu_{\alpha\alpha} + \frac{1}{n} \sum_{\beta \neq \alpha} \lambda_{\alpha\beta} \nu_{\alpha\beta}^2 + \frac{1}{n^2} \left(\sum_{\beta \neq \alpha} \sum_{\gamma \neq \alpha} \lambda_{\alpha\beta} \lambda_{\alpha\gamma} \nu_{\alpha\beta} \nu_{\beta\gamma} \nu_{\gamma\alpha} - \sum_{\beta \neq \alpha} \lambda_{\alpha\beta}^2 \nu_{\alpha\alpha} \nu_{\alpha\beta}^2 \right) + \dots, \quad (4.10)$$

$$E(\nu_{ab}) = 0,$$

$$E(\nu_{ab} \nu_{cd}) = \lambda_a \lambda_b (\delta_{ac} \delta_{bd} + \delta_{ad} \delta_{bc}) (\equiv \alpha(ab, cd)),$$

$$\begin{aligned}
 E(\nu_{ab} \nu_{cd} \nu_{ef}) = & n^{-1/2} \{ (\delta_{ac} \delta_{be} \delta_{df} \lambda_a \lambda_b \lambda_d + \delta_{ac} \delta_{bf} \delta_{ce} \lambda_a \lambda_b \lambda_c) \\
 & + (\delta_{ac} \delta_{bf} \delta_{df} \lambda_a \lambda_b \lambda_d + \delta_{ad} \delta_{be} \delta_{cf} \lambda_a \lambda_b \lambda_c) \\
 & + (\delta_{ad} \delta_{bf} \delta_{ce} \lambda_a \lambda_b \lambda_c + \delta_{ae} \delta_{bc} \delta_{df} \lambda_a \lambda_b \lambda_d) \\
 & + (\delta_{ae} \delta_{bd} \delta_{cf} \lambda_a \lambda_b \lambda_c + \delta_{af} \delta_{bc} \delta_{de} \lambda_a \lambda_b \lambda_d) \} + O(n^{-3/2}),
 \end{aligned}$$

$$E(\nu_{ab} \nu_{cd} \nu_{ef} \nu_{gh}) = \alpha(ab, cd) \alpha(ef, gh) + \alpha(ab, ef) \alpha(cd, gh) + \alpha(ab, gh) \alpha(cd, ef) + O(n^{-1}),$$

Burada $\delta_{\alpha\beta}$ Kroniker deltayı göstermektedir (Siotani, 1985). $E(l_\alpha) - \lambda_\alpha$ 'nin asimptotik açılımı, (4.10)'dan faydalanarak (4.11)'deki gibi elde edilebilir.

$$E(l_p) - \lambda_\alpha = \frac{1}{n} \lambda_\alpha \sum_{\beta \neq \alpha} \lambda_{\alpha\beta} \lambda_\beta + \frac{1}{n^2} \lambda_\alpha \left\{ \left(\sum_{\beta \neq \alpha} \lambda_{\alpha\beta} \lambda_\beta \right)^2 + \sum_{\beta \neq \alpha} (\lambda_{\alpha\beta} \lambda_\beta)^2 \right\} + \dots \quad (4.11)$$

(4.11)'in yerine (4.9) yazılarak

$$(l_1 - E(l_1))(l_p - E(l_p)) = \frac{1}{n^{\frac{3}{2}}} \tilde{F}_1 + \frac{1}{n^2} \tilde{F}_2 + \dots,$$

elde edilir. Burada,

$$\begin{aligned} \tilde{F}_1 &= -\nu_{11} \lambda_p \sum_{\alpha \neq p} \lambda_{p\alpha} \lambda_\alpha - \nu_{pp} \lambda_1 \sum_{\alpha \neq 1} \lambda_{1\alpha} \lambda_\alpha + \nu_{11} \sum_{\alpha \neq 1} \lambda_{1\alpha} \nu_{1\alpha}^2 + \nu_{pp} \sum_{\alpha \neq p} \lambda_{p\alpha} \nu_{p\alpha}^2, \\ \tilde{F}_2 &= \lambda_1 \lambda_p \left(\sum_{\alpha \neq 1} \lambda_{1\alpha} \lambda_\alpha \right) \left(\sum_{\alpha \neq p} \lambda_{p\alpha} \lambda_\alpha \right) - \lambda_p \left(\sum_{\alpha \neq 1} \lambda_{1\alpha} \nu_{1\alpha}^2 \right) \left(\sum_{\alpha \neq p} \lambda_{p\alpha} \lambda_\alpha \right) - \lambda_1 \left(\sum_{\alpha \neq p} \lambda_{p\alpha} \nu_{p\alpha}^2 \right) \left(\sum_{\alpha \neq 1} \lambda_{1\alpha} \lambda_\alpha \right) \\ &\quad + \nu_{11} \nu_{pp} \left(\sum_{\alpha \neq 1} \lambda_{1\alpha}^2 \nu_{1\alpha}^2 + \sum_{\alpha \neq p} \lambda_{p\alpha}^2 \nu_{p\alpha}^2 \right) + \left(\sum_{\alpha \neq 1} \lambda_{1\alpha} \nu_{1\alpha}^2 \right)^2 \\ &\quad + \nu_{pp} \sum_{\alpha \neq 1} \sum_{\beta \neq 1} \lambda_{1\alpha} \lambda_{1\beta} \nu_{1\alpha} \nu_{\alpha\beta} \nu_{\beta 1} + \nu_{11} \sum_{\alpha \neq p} \sum_{\beta \neq p} \lambda_{p\alpha} \lambda_{p\beta} \nu_{p\alpha} \nu_{\alpha\beta} \nu_{\beta p} \end{aligned}$$

dır. (4.10) kullanılarak $E(l_1 - E(l_1))(l_p - E(l_p))$ 'nin asimptotik açılımı (4.12)'de verilmiştir.

$$E(l_1 - E(l_1))(l_p - E(l_p)) = \frac{2\lambda_1 \lambda_p}{n} \frac{\lambda_1 \lambda_p}{n(\lambda_1 - \lambda_p)^2} + O\left(n^{-\frac{5}{2}}\right) \quad (4.12)$$

Üstelik l_α karakteristik kökün varyansı (4.13)'deki gibi belirtilebilir ve genişletilebilir.

$$Var(l_\alpha) = \frac{2}{n} \lambda_\alpha^2 + O\left(n^{-\frac{3}{2}}\right) \quad (4.13)$$

burada (4.12) ve (4.13) için ilgili korelasyon katsayısı (4.14)'deki gibidir.

$$Cor(l_1, l_p) = \frac{1}{n} \frac{\lambda_1 \lambda_p}{(\lambda_1 - \lambda_p)^2} + O\left(n^{-\frac{3}{2}}\right) \quad (4.14)$$

l_1 ve l_p için korelasyon katsayısı,

$$Cor(l_1, l_p) = \frac{1}{n} \frac{\lambda_1 \lambda_p}{(\lambda_1 - \lambda_p)^2} + O\left(n^{-\frac{3}{2}}\right) \quad (4.15)$$

şeklinde verilir. Bu korelasyon katsayısı örneklem hacmi büyük ve $\lambda_1 - \lambda_p$ çok büyük olduğunda, l_1 ve l_p 'nin korelasyonunun çok küçük olduğunu gösterir. (4.15) için (4.16)'daki ifade kullanılabilir.

$$\begin{aligned} & \Pr\left\{l_p / \left(1 + \frac{p+2}{4} \sqrt{\frac{2}{n} Z_\alpha}\right)^{4/(p+2)} \leq \lambda_p\right\} \Pr\left\{\lambda_1 \leq l_1 / \left(1 - \frac{1}{3} \sqrt{\frac{2}{n} Z_\alpha}\right)^3\right\} \\ & \approx \Pr\left\{l_p / \left(1 + \frac{p+2}{4} \sqrt{\frac{2}{n} Z_\alpha}\right)^{4/(p+2)} \leq \lambda_p < \lambda_1 < l_1 / \left(1 - \frac{1}{3} \sqrt{\frac{2}{n} Z_\alpha}\right)^3\right\} \\ & = 1 - \varepsilon \end{aligned} \quad (4.16)$$

Teorem 4.1. Σ 'nin tüm karakteristik kökler için güven aralığı tahmini, $1 - \varepsilon$ anlam düzeyinde (4.17)'deki gibidir (Sakaori, Yamada, 2007).

$$l_p / \left(1 + \frac{p+2}{4} \sqrt{\frac{2}{n} Z_\alpha}\right)^{4/(p+2)} \leq \lambda_p < \lambda_1 \leq l_1 / \left(1 - \frac{1}{3} \sqrt{\frac{2}{n} Z_\alpha}\right)^3 \quad (4.17)$$

Burada $\alpha = 1 - \sqrt{1 - \varepsilon}$ 'dir.

Anderson (2003) (4.18)'deki gibi başka bir güven aralığı formüle etmiştir .

$$\Pr\left(nl_p / \chi_{n-p+1}^2(\alpha) \leq \lambda_p < \lambda_1 \leq nl_1 / \chi_n^2(1-\alpha)\right) = 1 - \varepsilon \quad (4.18)$$

Burada $\chi_f^2(\alpha)$ ile ifade edilen, f serbestlik dereceli χ^2 dağılımının üstten $100\alpha\%$ 'ıncı değere karşılık gelen değerdir.

Monte Carlo simülasyonlarıyla hesaplanan değişkenler için güven aralıkları simülasyonlarla gösterilmiştir. Bu simülasyonların sonuçları, (4.17)'de elde edilen alt sınırın (4.18)'dekinden daha kötü olduğu görülmüştür. Burada (4.18)'in alt sınırı kullanılarak güven aralığı belirlenir.

$$\Pr\left\{nl_p/\chi_{n-p+1}^2(\alpha) \leq \lambda_p\right\} \Pr\left\{\lambda_1 \leq l_1 / \left(1 - \frac{1}{3}\sqrt{\frac{2}{n}Z_\alpha}\right)^3\right\}$$

$$\Pr\left\{nl_p/\chi_{n-p+1}^2(\alpha) \leq \lambda_p < \lambda_1 \leq l_1 / \left(1 - \frac{1}{3}\sqrt{\frac{2}{n}Z_\alpha}\right)^3\right\} = 1 - \varepsilon$$

Teorem 4.2. Σ 'nın karakteristik kökleri için $1 - \varepsilon$ anlam düzeyinde güven aralığı tahmini aşağıdaki gibidir (Sakaori, Yamada, 2007).

$$nl_p/\chi_{n-p+1}^2(\alpha) \leq \lambda_p < \lambda_1 \leq l_1 / \left(1 - \frac{1}{3}\sqrt{\frac{2}{n}Z_\alpha}\right)^3 \quad (4.19)$$

Burada $\alpha = 1 - \sqrt{1 - \varepsilon}$ 'dır ve $\chi_f^2(\alpha)$ ile ifade edilen, f serbestlik dereceli χ^2 dağılımının üstten $100\alpha\%$ 'ıncı değere karşılık gelen değerdir.

Monte Carlo simülasyonları ile (4.19)'un doğruluğu, (4.18) ve (4.17) ile kıyaslanarak görülür.

Yüksek boyutlu durumlarda, genellikle büyük karakteristik köklere sahip başlıca birkaç bileşenlerle ilgilenilir, geriye kalanlar anlamsız olduğundan göz önünde bulundurulmaz. Ayrıca l_k ve l_{k+1} arasındaki korelasyon katsayısı, l_1, l_k 'nın l_p, l_{k+1} 'in yerine kullanılarak elde edilen sonuçlarla aynı formdadır. Bu nedenle k .ıncı ve $(k+1)$.inci kökler yeteri derecede farklı ise, sıradaki varsayım geçerlidir.

Varsayım 4.1. Çok boyutlu durumlarda, eğer k .nıncı ve $(k+1)$.inci karakteristik kökler yeteri derecede farklı ise, l_1 yerine l_k , l_p yerine l_{k+1} yazılarak λ_p ve λ_k 'nin güven aralığı yaklaşık olarak (4.19)'da alınmıştır.

Monte Carlo simülasyonları ile varsayımları düzenlemek için bazı kanıtlar gösterilecektir.

Yukarıdaki güven aralıkları, n büyük olduğunda en küçük ve en büyük karakteristik köklerin bağımsız olduğu varsayımına dayanarak oluşturulmuştur. Bu nedenle, aralıkların esas güven aralıkları, n küçük veya orta hacimli olduğundaki durumdan farklı olabilir. Aşağıda simülasyonlar kullanılarak esas güven aralıkları bulunmuştur.

A_c ve N_c sırayla (4.1) ve (4.17)'nin güven aralıkları olsun. A_c ve N_c 'nin güven seviyeleri, 1.000.000 örneklem kullanılarak hesaplanmıştır. N yerine 50,100 ve 200, p yerine 5 ve ε yerine 0,05 yazıldı. Aşağıdaki durumları düşünülün.

$$\text{Durum1 } \Sigma = \text{diag}(500, 50, 10, 5, 1) \quad (4.20)$$

$$\text{Durum2 } \Sigma = \text{diag}(200, 20, 15, 10, 5) \quad (4.21)$$

Tablo 4.1. ve 4.2.'de A_c ve N_c için güven sınırları ve genişlikler verilmiştir. Bu aralıklar içinde ki l_1 ve l_p örneklem karakteristik kökleri 10.000 tekrarın medyanı kullanılarak elde edilmiştir. Bu denemeler Fortran programı kullanılarak ve hesaplamalar da Mathematica programı kullanılarak yapıldı. Bu hesaplamalar yapılırken l ve u değerleri sırayla $100(1-\sqrt{1-\varepsilon})$ ve $100\sqrt{1-\varepsilon}$ yüzde oranları şeklinde atandı. Bu kıyas yoluyla küçük örnek hacmi için N_c 'nin daha dar olduğu açıkça görülür.

Bu sonuç, N_c 'nin A_c aralık tahmininden daha doğru olduğunu gösterir. Buna rağmen N_c 'nin alt limiti A_c 'den daha küçüktür. Burada en dar güven aralığına

ulaşmak için N_c yerine A_c 'nin alt limiti kullanılarak yeni bir güven aralığı olan N_c' formüle edilebilir.

Tablo 4.3. ve 4.4.'de gerçek güven seviyeleri verilmiştir. Bu simülasyonlar Mathematica programı kullanılarak yapılmıştır. Bu simülasyonlar 1.000.000 tekrar üzerinden yapılmıştır.

Tablo 4.1. Durum 1 için güven seviyeleri

n	A_c			N_c		
	Üst sınır	Alt sınır	Fark	Üst sınır	Alt sınır	Fark
50	763.386	0.676	762.710	751.945	0.668	751.277
100	669.718	0.760	668.958	664.862	0.757	664.105
200	612.632	0.824	611.808	610.462	0.824	609.638

Tablo 4.2. Durum 2 için güven seviyeleri

n	A_c			N_c		
	Üst sınır	Alt sınır	Fark	Üst sınır	Alt sınır	Fark
50	306.082	3.261	302.820	301.494	3.223	298.272
100	268.192	3.738	264.454	266.247	3.726	262.521
200	245.202	4.094	241.107	244.334	4.093	240.240

Tablo 4.3. Durum 1 için gerçek güven seviyeleri

n	A_c	N_c	N_c'
50	0.9537	0.9530	0.9498
100	0.9526	0.9510	0.9497
200	0.9517	0.9500	0.9498

Tablo 4.4. Durum 2 için gerçek güven seviyeleri

n	A_c	N_c	N'_c
50	0.9657	0.9638	0.9619
100	0.9600	0.9583	0.9573
200	0.9567	0.9549	0.9547

Tablo 4.5. Yüksek boyutlu durumlar için gerçek güven seviyeleri

A_c	N_c	N'_c
0.9889	0.9866	0.9866

Tablo 4.6. Tahminin gerçek güven seviyeleri

n	$p=20$	$p=50$
50	0.9570	0.9475
100	0.9527	0.9451
200	0.9511	0.9463

Asıl güven seviyelerinin örnek hacmi genişledikçe $1 - \varepsilon$ 'yakınsadığı ve N'_c 'nin durumlar içerisinde üçü arasında en doğrusu olduğu açıktır. Ayrıca en büyük ve en küçük karakteristik kökler arasındaki farkın daha büyük olduğunda, (4.20)'deki yakınsama oranı (4.21)'den daha hızlıdır.

Tablo 4.5.'de, yüksek boyutlu durumda $\varepsilon = 0.05$, $p = 40$, $n = 50$ ve $\Sigma = \text{diag}(100, 49, 16, 9, 4, 4 \times 0.95, \dots, 4 \times 0.95^{35})$ olduğunda asıl güven aralıkları verilmiştir. Burada 100.000 tekrar kullanılmıştır. Bu nedenle bu güven aralıkları doğru olabilir ancak uygun olan aralıktan daha geniş de olmuş olabilir. Bu sonucun böyle olması normaldir çünkü, yaklaşık dağılımlar sadece n sonsuza giderken, p sınırlı olduğunda elde edilmiştir.

Tablo 4.6.'da, yüksek boyutlu durumda $p=20$ ve $p=50$ için $\Sigma = \text{diag}(100, 49, 16, 1, 1, \dots, 1)$ olduğunda asıl güven seviyeleri verilmiştir. Bu durumda λ_3 ve λ_4 birbirinden farklı görünmektedir. Varsayımlardan faydalanarak λ_1, λ_2 ve λ_3 'ün güven aralıklarını hesaplayabiliriz. Burada 100.000 tekrar kullanılmıştır. $n=50$ olduğunda bile asıl güven aralıkları $p=20$ ve $p=50$ için, $1 - \varepsilon$ değerine yakın olur. Bu nümerik sonuçlar bu varsayımı desteklemektedir

5. UYGULAMALAR

Bu bölümde önceden belirlenen kovaryans matrislerinin Jacknife ve Bootstrap yöntemi ile tahmin edilmesi ve güven aralıklarının belirlenmesi için Monte Carlo simülasyon çalışmaları yapılmıştır.

5.1. Uygulama 1

Kovaryans matrisinin özdeğerlerini ve özvektörlerini veren, spss programı yazıldı, 3 değişkenli normal dağılımdan 30 sayı üretildi;

$$n=30 \text{ ve } \mu = \begin{bmatrix} 10 \\ 20 \\ 50 \end{bmatrix} \text{ ve } S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix} \text{ için } S\text{'nin en küçük ve en büyük}$$

özdeğerlerini ve tüm özdeğerlerine karşılık gelen özvektörleri bulmak için spss programı kullanılmıştır.

matrix.

compute n=30.

compute ort={10;20;50}.

compute km={1,0,0;0,4,0;0,0,9}.

compute x11=uniform(n,1).

compute x12=uniform(n,1).

compute z1=(-2*ln(x11))&**0.5&*sin(6.28*x12)).

compute x21=uniform(n,1).

compute x22=uniform(n,1).

compute z2=(-2*ln(x21))&**0.5&*sin(6.28*x22)).

compute z={z1,z2}.

```
compute x31=uniform(n,1).
compute x32=uniform(n,1).
compute z3=((-2*ln(x31))&*&0.5)&*sin(6.28*x32)).
compute z={z1,z2,z3}.

/*Kovaryans matrisinin Cholesky yapısını kullanarak */
compute ckm=chol(km).
compute x=t((t(ckm)*t(z))+(kroneker(make(1,n,1),ort))).
save x /outfile=*.

compute oort=t(csum(x)/n).
print oort.
compute s=(t(x)*x/n-oort*t(oort))*n/(n-1). /* i. gruba ilişkin kovaryans matrisi */

print s.
call eigen(s,sov,sod).
compute lmax=mmax(sod).
compute lmin=mmin(sod).
compute sonuc={lmin,lmax}.
print sonuc.

print sod.
end matrix.
```

S kovaryans matrisinin en küçük özdeğeri l_1 ve en büyük özdeğeri l_3 10 tekrarla tablo(5.1) de verildiği gibi bulunmuştur.

Tablo 5.1. 10 tekrar ve $n=30$ için en küçük ve en büyük özdeğerler

l_1	l_3
0.847325152	7.307374835
0.97329263	10.04166251
0.652684834	9.368833714
0.833730904	8.026180585
0.93262954	12.96791435
0.890073428	5.660639553
1.40171629	11.26771939
0.68464786	11.12394837
0.87723787	14.48237536
0.61540782	11.93951288

l_1 özdeğerinin medyanı= 0.862281511

l_3 özdeğerinin medyanı= 10.58280544

20 tekrar için $l_1.l_3$ arasındaki korelasyon 10 tekrar için $l_1.l_3$ arasındaki korelasyondan daha yakındır; 20 tekrar için tablo(5.2)'de verildiği gibi bulunmuştur.

$$\begin{aligned} \text{Cor}(l_1, l_p) &= \frac{1}{n} \frac{\lambda_1 \cdot \lambda_p}{(\lambda_1 - \lambda_p)^2} + o(n^{-\frac{3}{2}}) \\ &= \frac{1}{29} \frac{1.9}{(1-9)^2} = 0.0048 \end{aligned}$$

Tablo 5.2. 20 tekrar için en küçük ve en büyük özdeğerler

l_1	l_3
0.847325152	7.307374835
0.97329263	10.04166251
0.652684834	9.368833714
0.833730904	8.026180585
0.93262954	12.96791435
0.890073428	5.660639553
1.40171629	11.26771939
0.68464786	11.12394837
0.87723787	14.48237536
0.61540782	11.93951288
0.668275831	0.668275831
1.132232355	1.132232355
0.68490577	0.68490577
0.849696487	0.849696487
0.773697382	0.773697382
1.009350749	1.009350749
0.8147744	0.8147744
0.676185665	0.676185665
0.624300485	0.624300485
0.784466208	0.784466208

l_1 özdeğerinin medyanı= 0.824252652

l_3 özdeğerinin medyanı= 10.10115591

5.2. Uygulama 2

Wishart Dağılımı ile ilgili program yazıldı. Wishart Dağılımına sahip A matrisinin dağılımı;

$$\frac{|A|^{\frac{1}{2}(n-p-1)} \exp\left(-\frac{1}{2} \text{tr} \Sigma^{-1} A\right)}{2^{\frac{1}{2}np} \pi^{p(p-1)/4} |\Sigma|^{\frac{1}{2}n} \prod_{i=1}^p \Gamma\left[\frac{1}{2}(n+1-i)\right]}$$

şeklindedir.

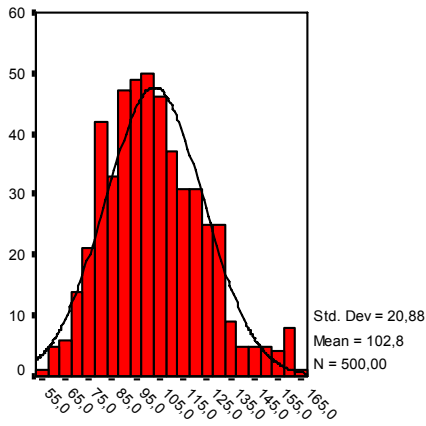
```
matrix.  
compute n=6.  
compute S={1,0,0;0,4,0;0,0,25}.  
compute W={1,0,0;0,4,0;0,0,24}.  
compute p=nrow(S).  
compute a=det(S)**(n/2).  
compute b=det(W)**(n-p-1)/2.  
compute c=exp(-trace(inv(S)*W)/2).  
print det(s).  
print det(w).  
print a.  
print b.  
print c.  
compute gama33=3.14**1.5*2*(1.5*0.5*3.14**0.5)*1.  
compute f=(1/(2**(p*n/2)*gama33*a))*b*c.  
print f.  
end matrix.
```

5.3. Uygulama 3

$$S_1 = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 40 & 0 \\ 0 & 0 & 9 \end{bmatrix} \quad \lambda_1=100 \quad \lambda_2=40 \quad \lambda_3=9$$

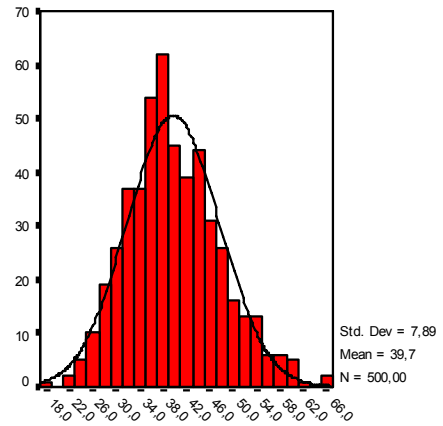
Tablo 5.3. n=50 ve 500 deneme için S_1 matrisinin özdeğerlerinin istatistikleri

n=50	λ_1	λ_2	λ_3
Ortalama	102.785	39.742	8.624
Medyan	101.030	38.914	8.498
Alt sınır	100.951	39.049	8.473
Üst sınır	104.620	40.435	8.776



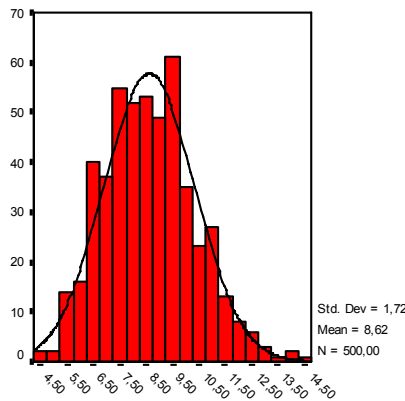
L1

a) n=50 için S_1 matrisinin λ_1 dağılım grafiği



L2

b) n=50 için S_1 matrisinin λ_2 dağılım grafiği



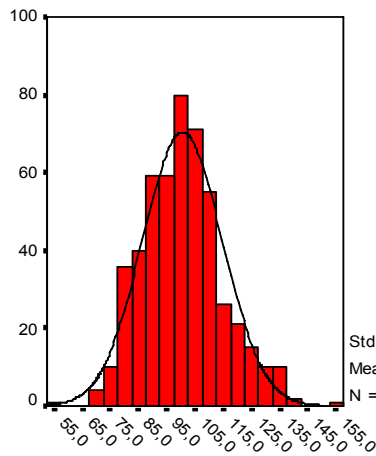
L3

c) n=50 için S_1 matrisinin λ_3 dağılım grafiği

Şekil 5.1. n=50 ve 500 deneme için S_1 matrisinin özdeğerlerinin dağılım grafikleri

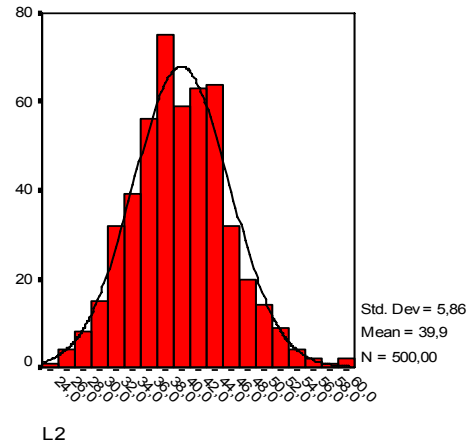
Tablo 5.4. n=100 ve 500 deneme için S_1 matrisinin özdeğerlerinin istatistikleri

n=100	λ_1	λ_2	λ_3
Ortalama	100.511	39.936	8.944
Medyan	100.456	39.602	8.841
Alt sınır	99.268	39.421	8.838
Üst sınır	101.754	40.451	9.050



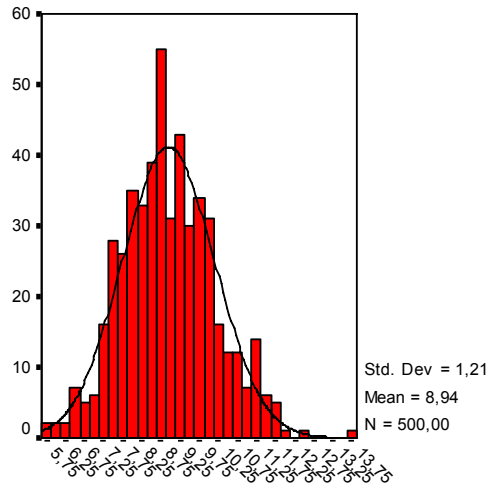
L1

a) n=100 için S_1 matrisinin λ_1 dağılım grafiği



L2

b) n=100 için S_1 matrisinin λ_2 dağılım grafiği



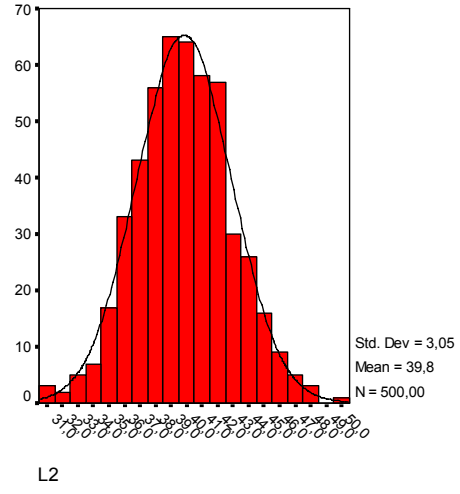
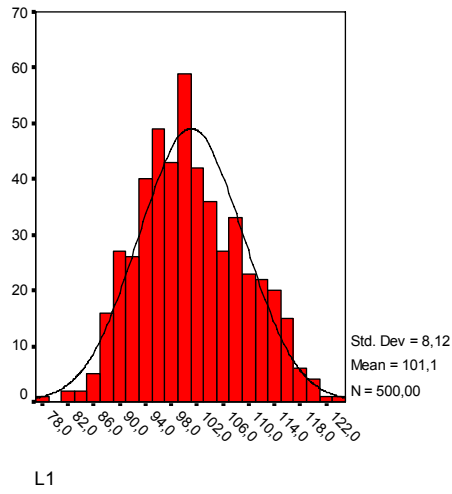
L3

c) n=100 için S_1 matrisinin λ_3 dağılım grafiği

Şekil 5.2. n=100 ve 500 deneme için S_1 matrisinin özdeğerlerinin dağılım grafikleri

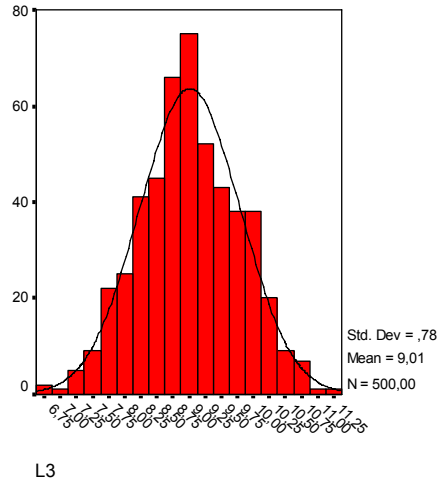
Tablo 5.5. n=300 ve 500 deneme için S_1 matrisinin özdeğerlerinin istatistikleri

n=300	λ_1	λ_2	λ_3
Ortalama	101.096	39.844	9.014
Medyan	100.382	39.772	9.009
Alt sınır	100.382	39.576	8.945
Üst sınır	101.809	40.112	9.082



a) n=300 için S_1 matrisinin λ_1 dağılım grafiği

b) n=300 için S_1 matrisinin λ_2 dağılım grafiği

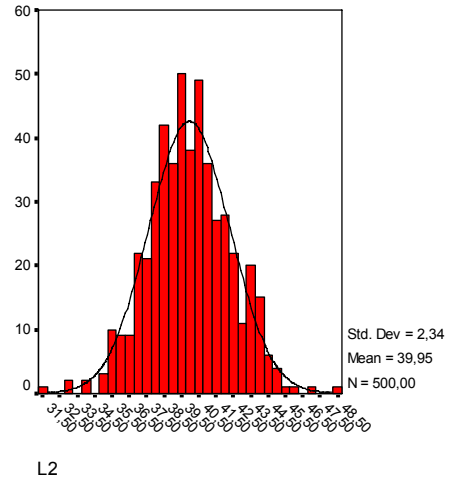
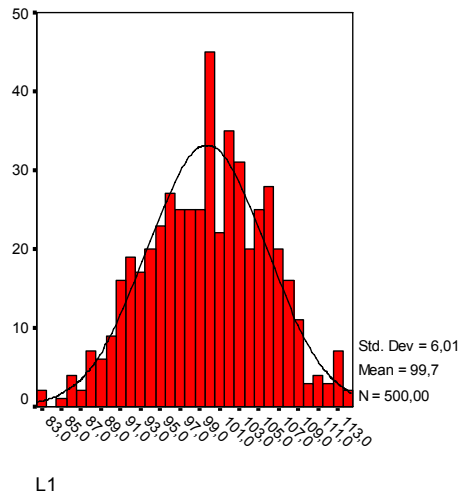


c) n=300 için S_1 matrisinin λ_3 dağılım grafiği

Şekil 5.3. n=300 ve 500 deneme için S_1 matrisinin özdeğerlerinin dağılım grafikleri

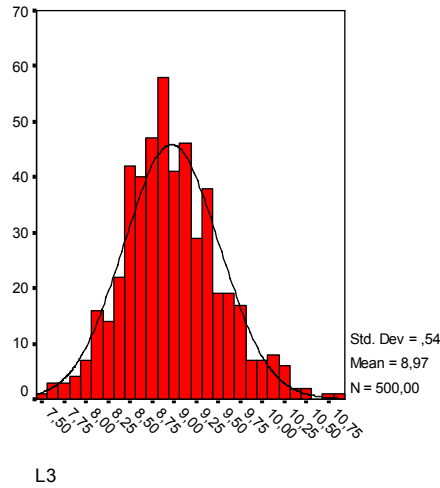
Tablo 5.6. n=500 ve 500 deneme için S_1 matrisinin özdeğerlerinin istatistikleri

n=500	λ_1	λ_2	λ_3
Ortalama	99.733	39.950	8.972
Medyan	99.965	39.885	8.914
Alt sınır	99.206	39.744	8.924
Üst sınır	100.261	40.155	9.019



a) n=500 için S_1 matrisinin λ_1 dağılım grafiği

b) n=500 için S_1 matrisinin λ_2 dağılım grafiği

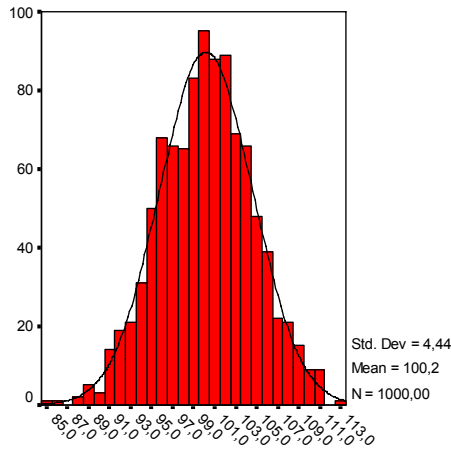


c) n=500 için S_1 matrisinin λ_3 dağılım grafiği

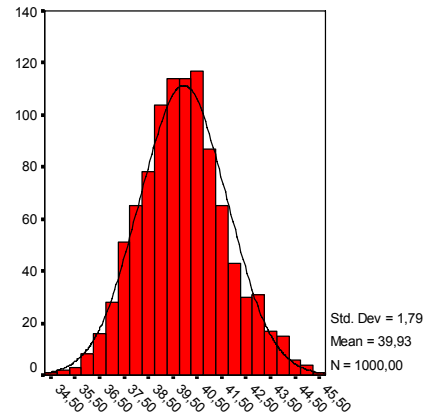
Şekil 5.4. n=500 ve 500 deneme için S_1 matrisinin özdeğerlerinin dağılım grafikleri

Tablo 5.7. n=1000 ve 1000 deneme için S_1 matrisinin özdeğerlerinin istatistikleri

n=1000	λ_1	λ_2	λ_3
Ortalama	100.184	39.933	8.984
Medyan	100.259	39.886	8.995
Alt sınır	99.909	39.822	8.960
Üst sınır	100.460	40.044	9.009



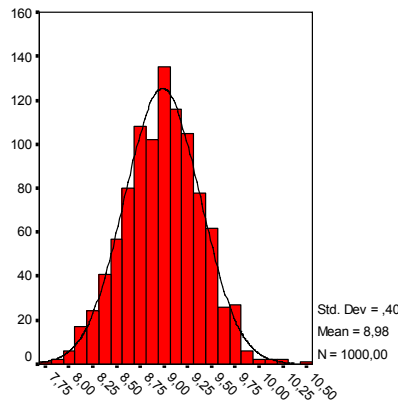
L1



L2

a) n=1000 için S_1 matrisinin λ_1 dağılım grafiği

b) n=1000 için S_1 matrisinin λ_2 dağılım grafiği



L3

c) n=1000 için S_1 matrisinin λ_3 dağılım grafiği

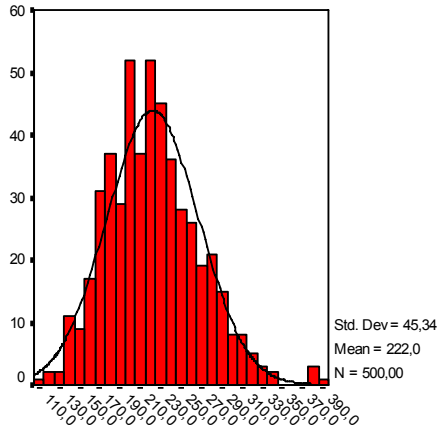
Şekil 5.5. n=1000 ve 1000 deneme için S_1 matrisinin özdeğerlerinin dağılım grafikleri

5.4. Uygulama 4

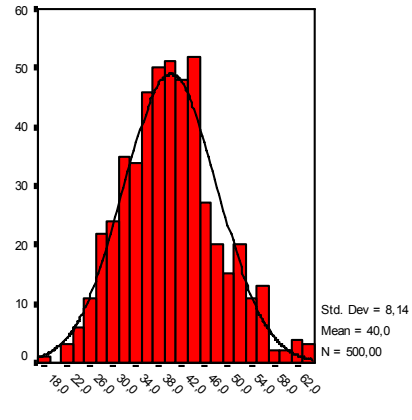
$$S_2 = \begin{bmatrix} 200 & 40 & 30 \\ 40 & 50 & 10 \\ 30 & 10 & 5 \end{bmatrix} \quad \lambda_1=214.74 \quad \lambda_2=40.14 \quad \lambda_3=0.116$$

Tablo 5.8. n=50, 500 deneme sonuçlarının istatistikleri (diagonal olmayan matris)

n=50	λ_1	λ_2	λ_3
Ortalama	221.984	40.033	0.113
Medyan	218.170	39.783	0.111
Alt sınır	218.000	39.318	0.111
Üst sınır	225.968	40.748	0.115



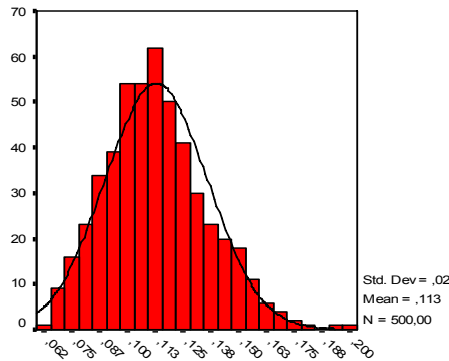
L1



L2

a) n=50 için S_2 matrisinin λ_1 dağılım grafiği

b) n=50 için S_2 matrisinin λ_2 dağılım grafiği



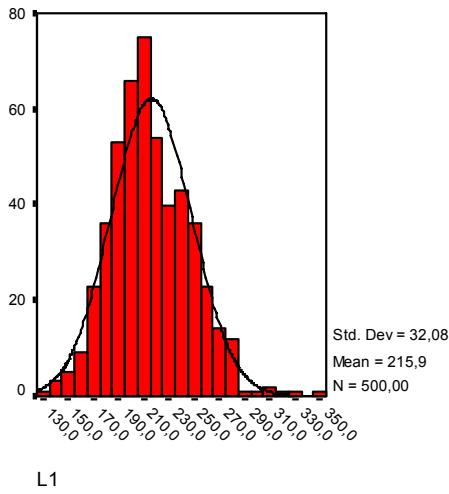
L3

c) n=50 için S_2 matrisinin λ_3 dağılım grafiği

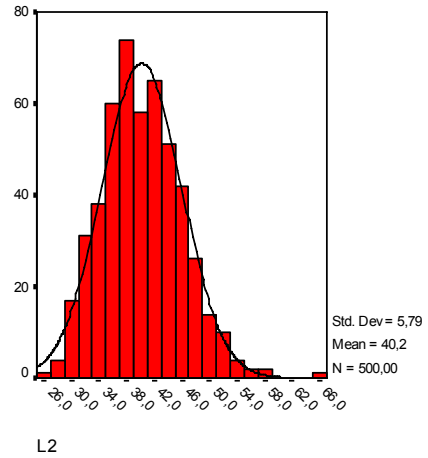
Şekil 5.6. n=50 ve 500 deneme için S_2 matrisinin özdeğerlerinin dağılım grafikleri

Tablo 5.9. n=100, 500 deneme sonuçlarının istatistikleri (diagonal olmayan matris)

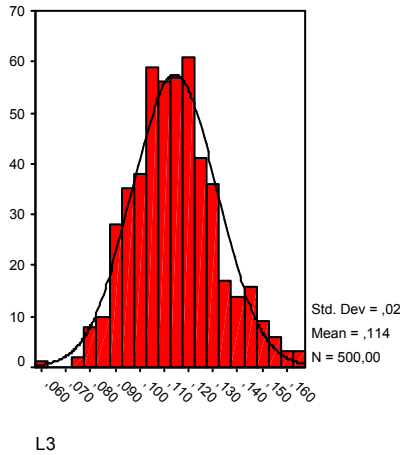
n=100	λ_1	λ_2	λ_3
Ortalama	215.902	40.178	0.114
Medyan	211.933	39.780	0.113
Alt sınır	213.083	39.669	0.112
Üst sınır	218.721	40.687	0.115



a) n=100 için S_2 matrisinin λ_1 dağılım grafiği



b) n=100 için S_2 matrisinin λ_2 dağılım grafiği

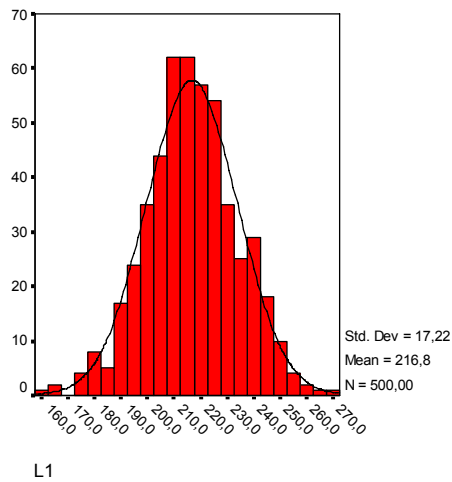


c) n=100 için S_2 matrisinin λ_3 dağılım grafiği

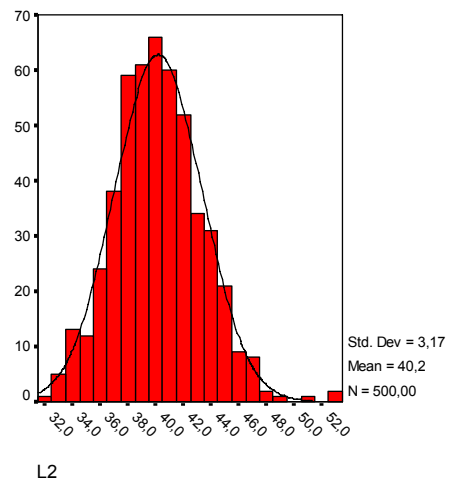
Şekil 5.7. n=100 ve 500 deneme için S_2 matrisinin özdeğerlerinin dağılım grafikleri

Tablo 5.10. n=300, 500 deneme sonuçlarının istatistikleri (diagonal olmayan matris)

n=100	λ_1	λ_2	λ_3
Ortalama	216.759	40.190	0.115
Medyan	216.147	39.952	0.114
Alt sınır	215.246	39.912	0.1144
Üst sınır	218.272	40.469	0.1161



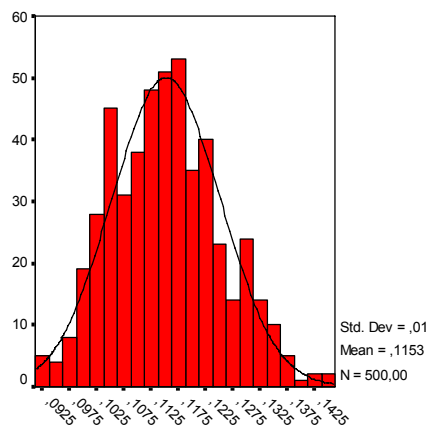
L1



L2

a) n=300 için S_2 matrisinin λ_1 dağılım grafiği

b) n=300 için S_2 matrisinin λ_2 dağılım grafiği



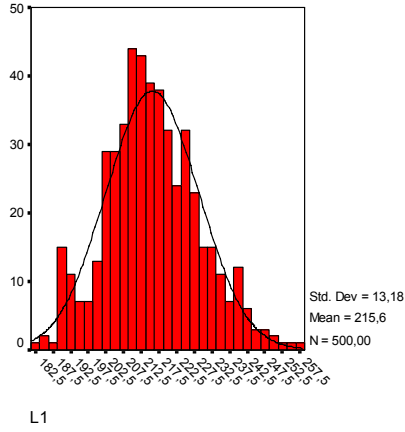
L3

c) n=300 için S_2 matrisinin λ_3 dağılım grafiği

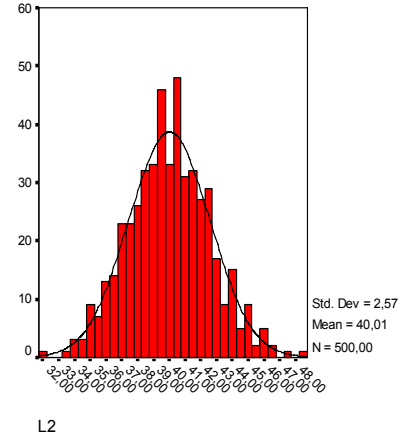
Şekil 5.8. n=300 ve 500 deneme için S_2 matrisinin özdeğerlerinin dağılım grafikleri

Tablo 5.11. n=500, 500 deneme sonuçlarının istatistikleri (diagonal olmayan matris)

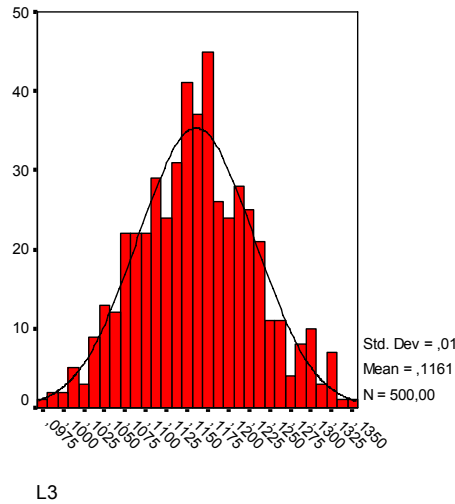
n=100	λ_1	λ_2	λ_3
Ortalama	215.632	40.006	0.116
Medyan	214.632	40.031	0.115
Alt sınır	214.474	39.780	0.115
Üst sınır	216.790	40.232	0.1163



a) n=500 için S_2 matrisinin λ_1 dağılım grafiği



b) n=500 için S_2 matrisinin λ_2 dağılım grafiği

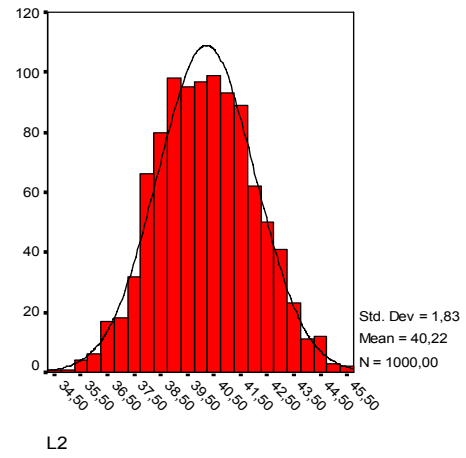
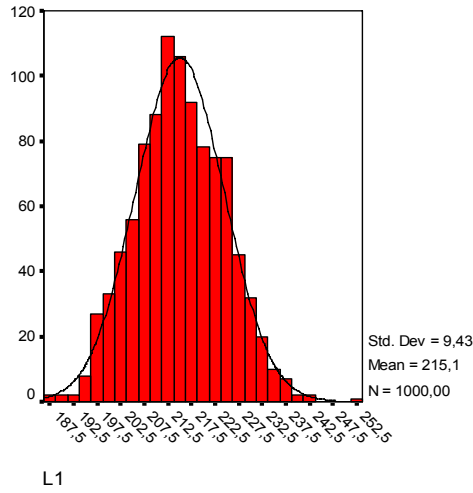


c) n=500 için S_2 matrisinin λ_3 dağılım grafiği

Şekil 5.9. n=500 ve 500 deneme için S_2 matrisinin özdeğerlerinin dağılım grafikleri

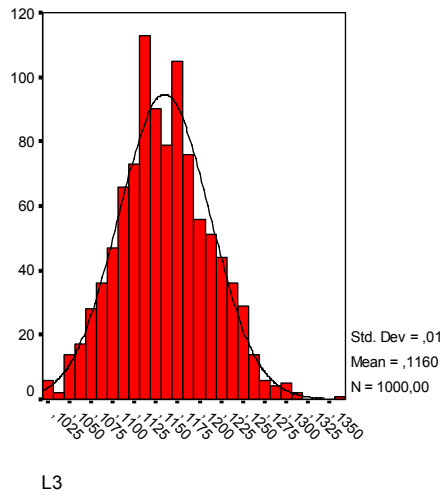
Tablo 5.12. n=1000, 1000 deneme sonuçlarının istatistikleri (diyagonal olmayan matris)

n=100	λ_1	λ_2	λ_3
Ortalama	215.071	40.215	0.116
Medyan	214.630	40.188	0.115
Alt sınır	214.485	40.102	0.115
Üst sınır	215.656	40.329	0.116



a) n=1000 için S_2 matrisinin λ_1 dağılım grafiği

b) n=1000 için S_2 matrisinin λ_2 dağılım grafiği



c) n=1000 için S_2 matrisinin λ_3 dağılım grafiği

Şekil 5.10. n=1000 ve 1000 deneme için S_2 matrisinin özdeğerlerinin dağılım grafikleri

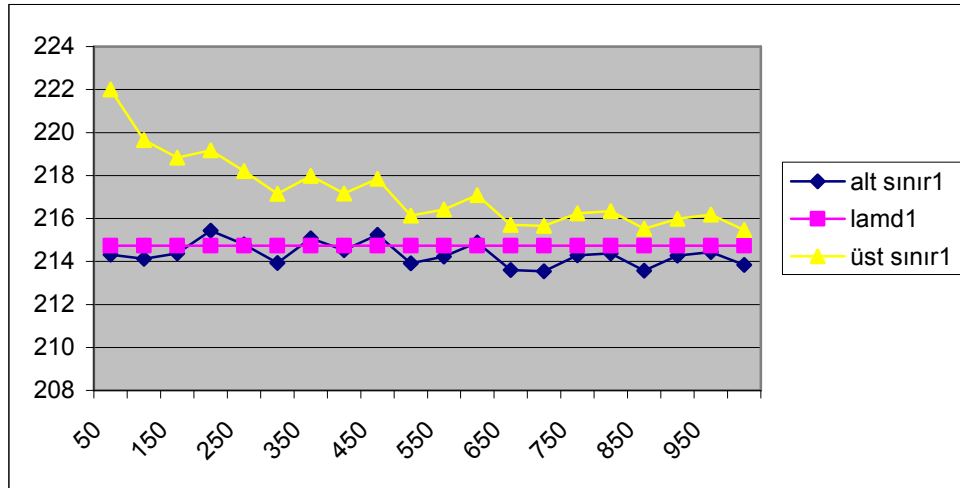
5.5. Uygulama 5

$$S_2 = \begin{bmatrix} 200 & 40 & 30 \\ 40 & 50 & 10 \\ 30 & 10 & 5 \end{bmatrix} \quad \lambda_1 = 214,74 \quad \lambda_2 = 40,14 \quad \lambda_3 = 0,116$$

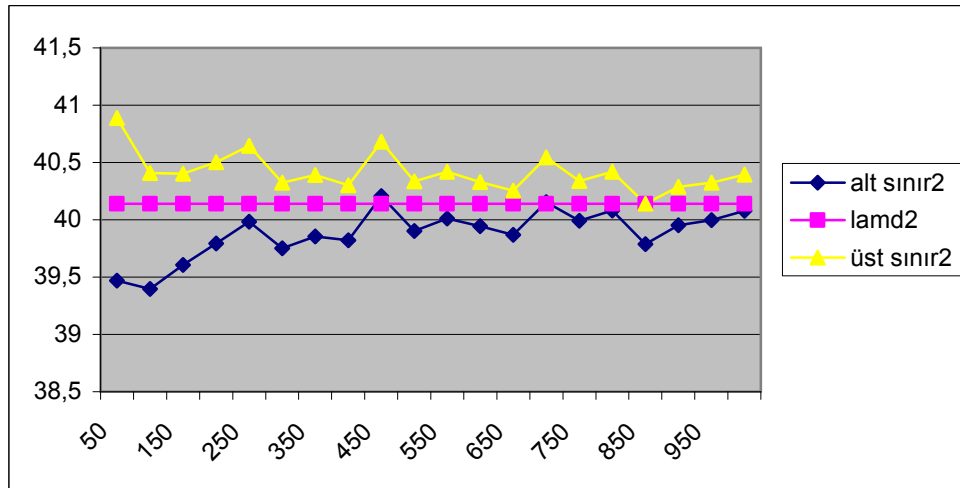
Tablo 5.13. Farklı örnek hacimleri (n) ve 500 deneme için diagonal olmayan matrisin özdeğerlerin güven aralıkları tahmini

n	alt sınır1	lamd1	üst sınır1	alt sınır2	lamd2	üst sınır2	alt sınır3	lamd3	üst sınır3
50	214.313	214.74	222.009	39.469	40.14	40.889	.112	0.116	.116
100	214.130	214.74	219.652	39.396	40.14	40.408	.112	0.116	.115
150	214.372	214.74	218.826	39.606	40.14	40.403	.114	0.116	.116
200	215.435	214.74	219.180	39.794	40.14	40.502	.115	0.116	.116
250	214.793	214.74	218.199	39.984	40.14	40.644	.114	0.116	.116
300	213.933	214.74	217.141	39.752	40.14	40.325	.114	0.116	.116
350	215.075	214.74	217.974	39.855	40.14	40.392	.115	0.116	.116
400	214.528	214.74	217.160	39.820	40.14	40.301	.115	0.116	.116
450	215.243	214.74	217.854	40.206	40.14	40.682	.115	0.116	.116
500	213.914	214.74	216.139	39.903	40.14	40.336	.115	0.116	.117
550	214.226	214.74	216.420	40.011	40.14	40.420	.115	0.116	.116
600	214.883	214.74	217.081	39.943	40.14	40.330	.115	0.116	.116
650	213.605	214.74	215.696	39.869	40.14	40.255	.115	0.116	.116
700	213.545	214.74	215.656	40.153	40.14	40.546	.115	0.116	.116
750	214.289	214.74	216.233	39.992	40.14	40.337	.115	0.116	.116
800	214.381	214.74	216.349	40.081	40.14	40.418	.115	0.116	.116
850	213.565	214.74	215.524	39.788	40.14	40.139	.115	0.116	.115
900	214.270	214.74	215.992	39.952	40.14	40.287	.115	0.116	.116
950	214.436	214.74	216.173	39.996	40.14	40.324	.115	0.116	.116
1000	213.836	214.74	215.460	40.079	40.14	40.395	.115	0.116	.116

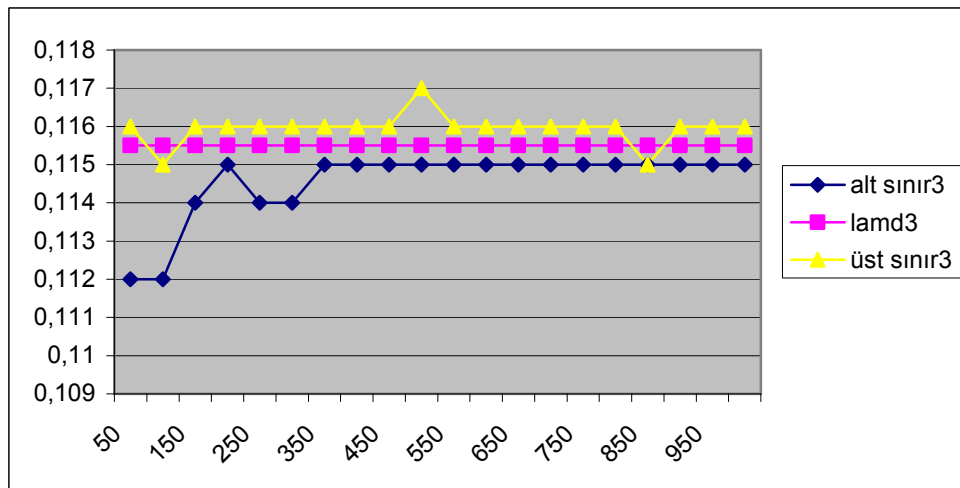
n ile ilişkisi; n büyüdükçe güven aralığı daha dar oluyor



a) Farklı örnek hacimleri için λ_1 'in güven aralığı



b) Farklı örnek hacimleri için λ_2 'nin güven aralığı



c) Farklı örnek hacimleri için λ_3 'ün güven aralığı

Şekil 5.11. Farklı örnek hacimleri için özdeğerlerin güven aralıkları

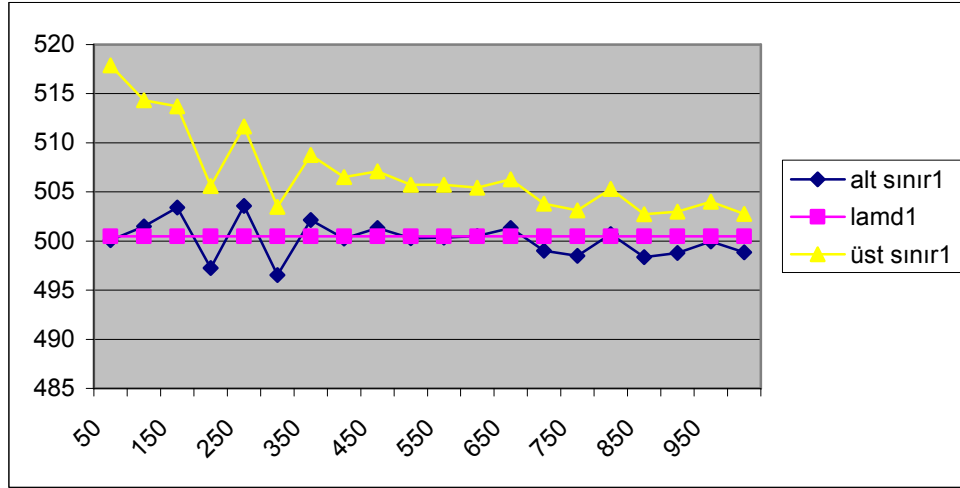
5.6. Uygulama 6

$$S_3 = \begin{bmatrix} 200 & -10 & 30 \\ -10 & 500 & 10 \\ 30 & 10 & 5 \end{bmatrix} \quad \lambda_1=500.49 \quad \lambda_2=204.27 \quad \lambda_3=0.23$$

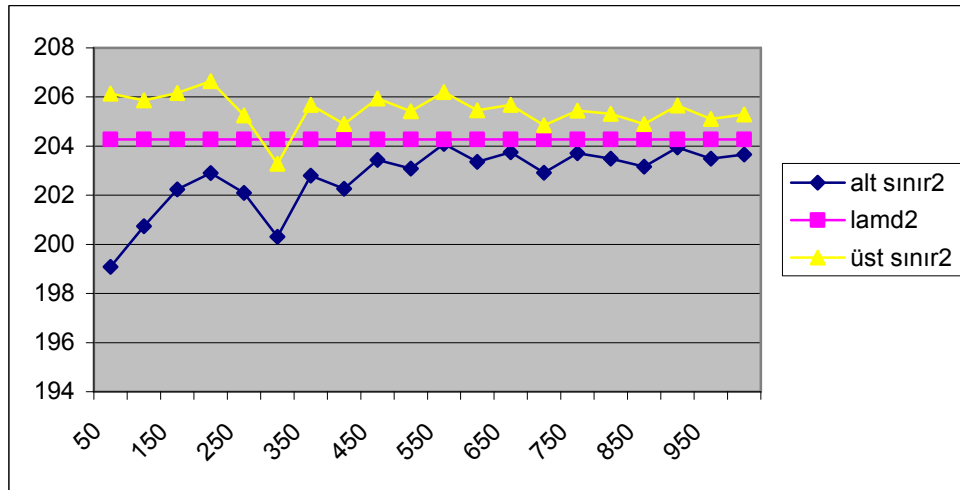
Tablo 5.14. Farklı örnek hacimleri (n) ve 500 deneme için diagonal olmayan matrisin özdeğerlerin güven aralıkları tahmini

n	alt sınır1	lamd1	üst sınır1	alt sınır2	lamd2	üst sınır2	alt sınır3	lamd3	üst sınır3
50	500.076	500.49	517.891	199.081	204.27	206.138	.219	0.23	.227
100	501.483	500.49	514.345	200.736	204.27	205.868	.226	0.23	.231
150	503.397	500.49	513.707	202.243	204.27	206.166	.226	0.23	.230
200	497.242	500.49	505.591	202.906	204.27	206.641	.225	0.23	.229
250	503.560	500.49	511.666	202.101	204.27	205.256	.227	0.23	.231
300	496.524	500.49	503.485	200.305	204.27	203.286	.226	0.23	.230
350	502.144	500.49	508.753	202.802	204.27	205.685	.228	0.23	.231
400	500.258	500.49	506.511	202.265	204.27	204.895	.228	0.23	.231
450	501.312	500.49	507.088	203.438	204.27	205.944	.228	0.23	.231
500	500.273	500.49	505.739	203.086	204.27	205.425	.228	0.23	.230
550	500.336	500.49	505.740	204.085	204.27	206.200	.229	0.23	.232
600	500.523	500.49	505.436	203.355	204.27	205.461	.228	0.23	.230
650	501.315	500.49	506.269	203.747	204.27	205.680	.228	0.23	.230
700	499.022	500.49	503.812	202.913	204.27	204.846	.227	0.23	.230
750	498.484	500.49	503.105	203.708	204.27	205.445	.228	0.23	.231
800	500.723	500.49	505.298	203.490	204.27	205.312	.229	0.23	.231
850	498.358	500.49	502.737	203.163	204.27	204.904	.228	0.23	.230
900	498.777	500.49	503.000	203.943	204.27	205.657	.229	0.23	.231
950	499.952	500.49	503.999	203.494	204.27	205.103	.228	0.23	.230
1000	498.855	500.49	502.753	203.661	204.27	205.284	.228	0.23	.230

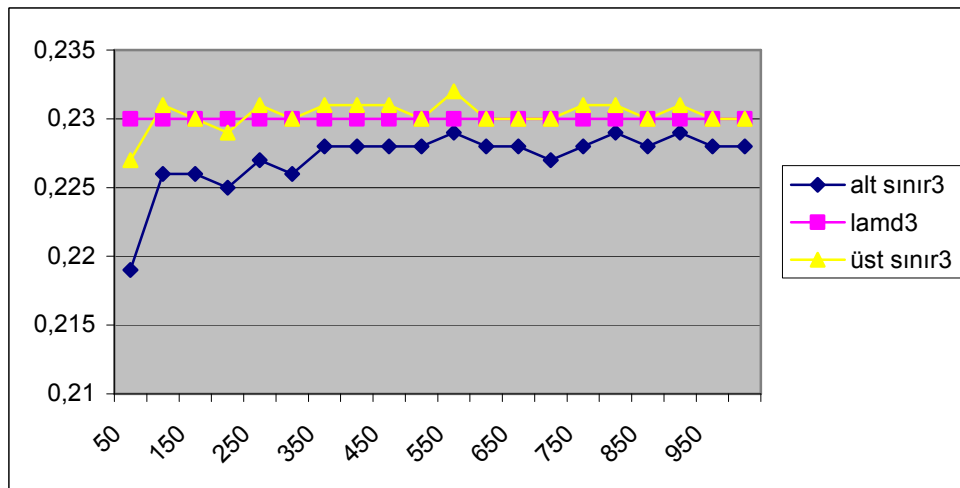
n ile ilişkisi; n büyüdükçe güven aralığı daha dar oluyor



a) Farklı örnek hacimleri için λ_1 'in güven aralığı



b) Farklı örnek hacimleri için λ_2 'nin güven aralığı



c) Farklı örnek hacimleri için λ_3 'ün güven aralığı

Şekil 5.12. Farklı örnek hacimleri için özdeğerlerin güven aralıkları

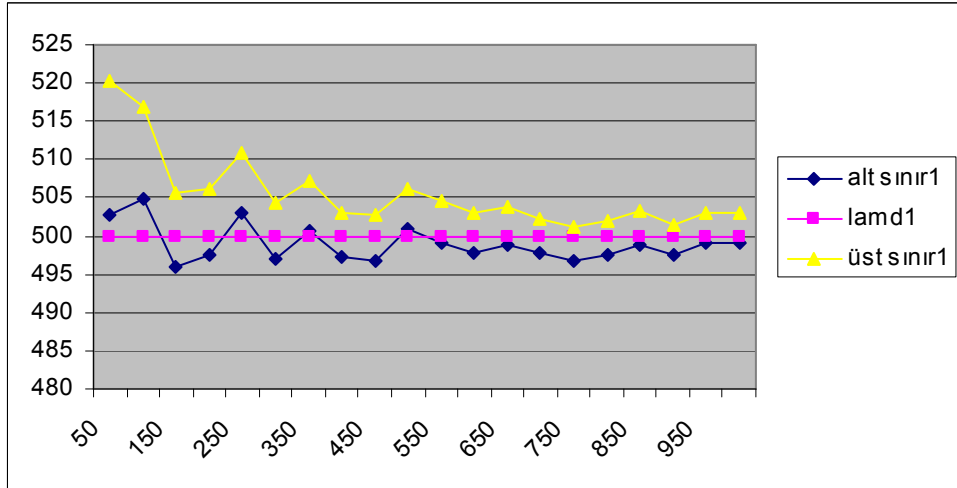
5.7. Uygulama 7

$S = \text{diag}(500, 50, 10, 5, 1)$ $k=500$ için.

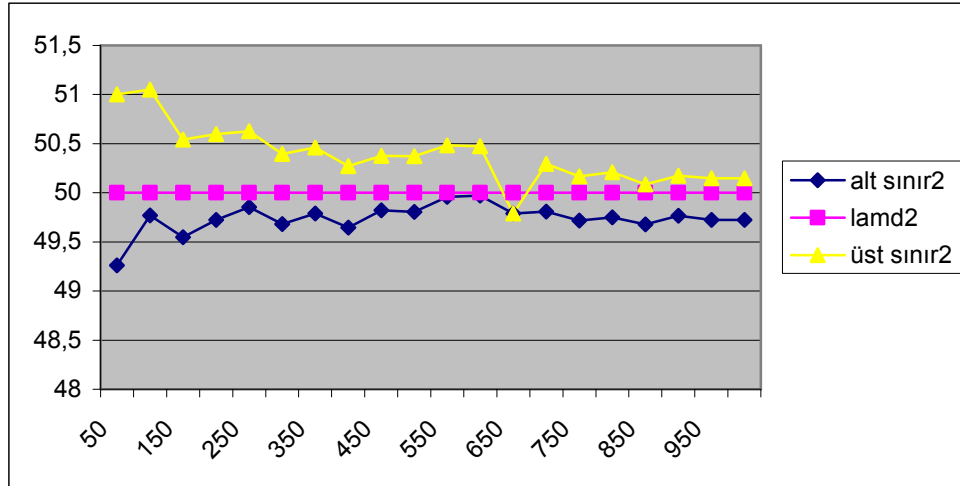
Tablo 5.15. Farklı örnek hacimleri (n) ve 500 deneme için en küçük ve en büyük karakteristik kök arasındaki farkın büyük olduğu matrisin özdeğerlerin güven aralıkları tahmini

n	alt sınır1	üst sınır1	alt sınır2	üst sınır2	alt sınır3	üst sınır3	alt sınır4	üst sınır4	alt sınır5	üst sınır5
50	502.640	520.255	49.260	51.000	9.784	10.121	4.671	4.834	.911	.946
100	504.845	516.980	49.771	51.049	9.908	10.160	4.803	4.919	.954	.978
150	495.864	505.628	49.549	50.542	9.939	10.149	4.868	4.976	.968	.988
200	497.499	506.093	49.724	50.598	9.953	10.122	4.896	4.986	.977	.994
250	503.048	510.881	49.855	50.627	9.935	10.091	4.903	4.980	.978	.993
300	497.084	504.308	49.681	50.394	9.916	10.054	4.935	5.009	.977	.991
350	500.626	507.279	49.789	50.459	9.914	10.048	4.935	5.004	.987	1.002
400	497.224	503.064	49.647	50.270	9.941	10.069	4.966	5.031	.990	1.003
450	496.799	502.636	49.821	50.375	9.967	10.084	4.966	5.027	.980	.992
500	500.804	506.116	49.807	50.374	9.915	10.029	4.953	5.010	.989	.999
550	499.142	504.597	49.958	50.483	9.961	10.067	4.957	5.009	.986	.996
600	497.822	502.991	49.972	50.475	9.972	10.075	4.938	4.991	.991	1.001
650	498.864	503.818	49.789	49.789	9.930	10.022	4.961	5.008	.990	1.000
700	497.755	502.354	49.809	50.295	9.900	9.998	4.964	5.008	.992	1.001
750	496.738	501.252	49.719	50.168	9.965	10.053	4.953	4.996	.992	1.001
800	497.430	501.849	49.751	50.210	9.985	10.070	4.945	4.992	.993	1.002
850	498.806	503.197	49.677	50.086	9.977	10.056	4.976	5.017	.992	1.000
900	497.437	501.564	49.765	50.174	9.988	10.076	4.961	5.003	.993	1.002
950	499.131	503.151	49.724	50.147	9.967	10.048	4.979	5.020	.994	1.002
1000	499.131	503.151	49.724	50.147	9.967	10.048	4.979	5.020	.994	1.002

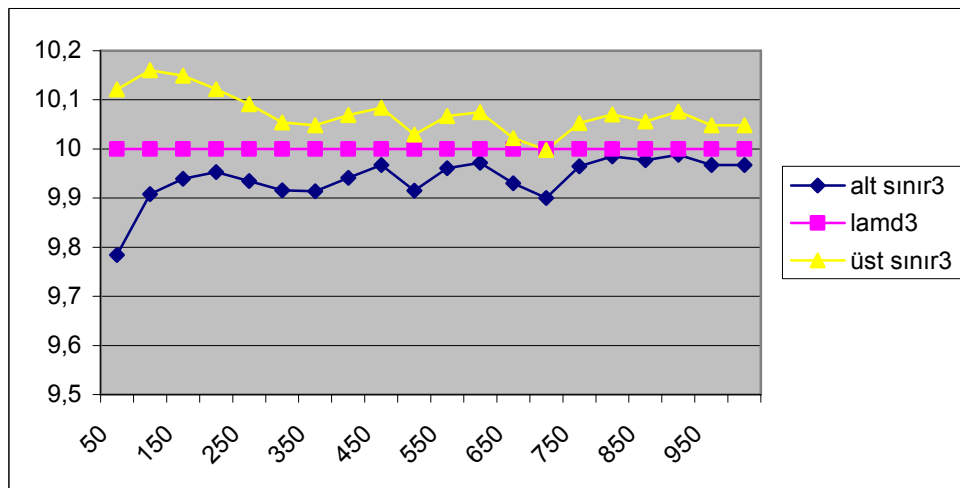
n ile ilişkisi; n büyüdükçe güven aralığı daha dar oluyor en küçük ve en büyük karakteristik kökler arasındaki fark büyüdükçe güven aralığı daralır.



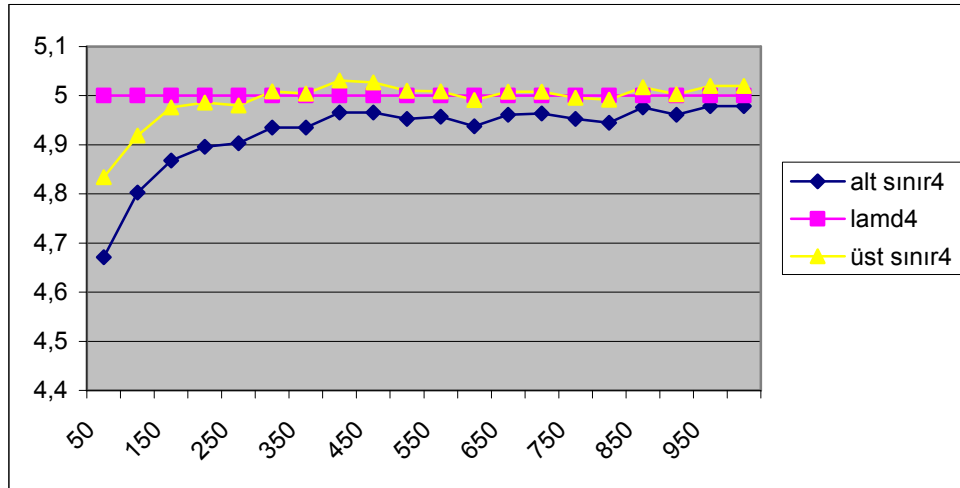
a) Farklı örnek hacimleri için λ_1 'in güven aralığı



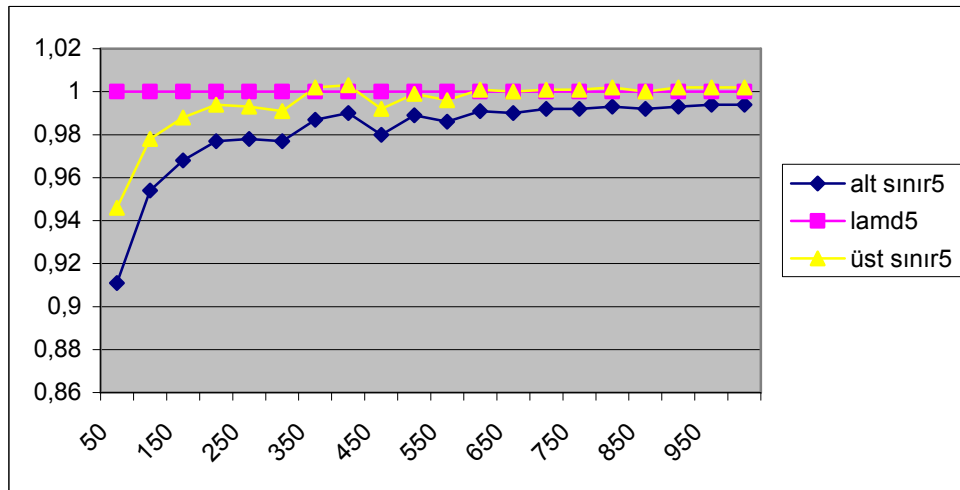
b) Farklı örnek hacimleri için λ_2 'nin güven aralığı



c) Farklı örnek hacimleri için λ_3 'ün güven aralığı



d) Farklı örnek hacimleri için λ_4 'ün güven aralığı



e) Farklı örnek hacimleri için λ_5 'in güven aralığı

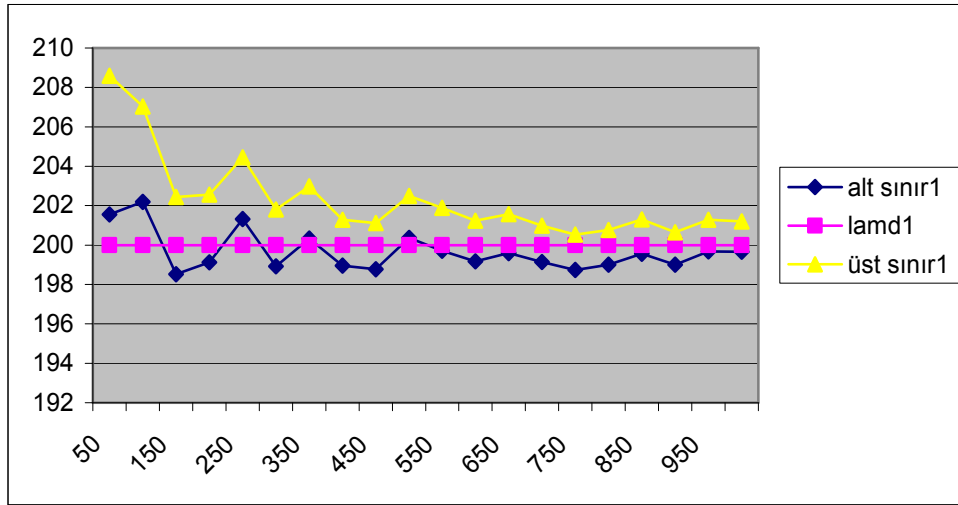
Şekil 5.13. Farklı örnek hacimleri için en küçük ve en büyük karakteristik kök arasındaki farkın büyük olduğu matrisin özdeğerlerinin güven aralıkları

5.8. Uygulama 8

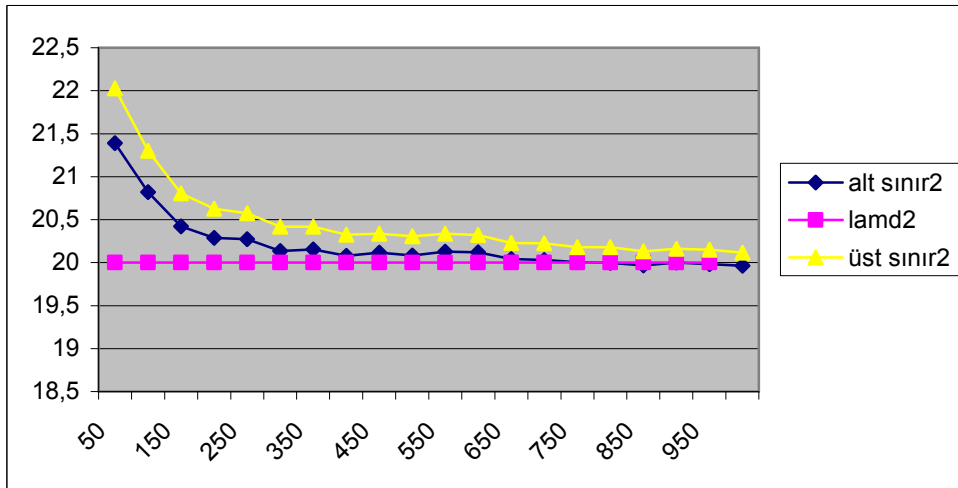
$S = \text{diag}(200, 20, 15, 10, 5)$ $k=500$ için.

Tablo 5.16. Farklı örnek hacimleri (n) ve 500 deneme için en küçük ve en büyük karakteristik kök arasındaki farkın küçük olduğu matrisin özdeğerlerin güven aralıkları tahmini

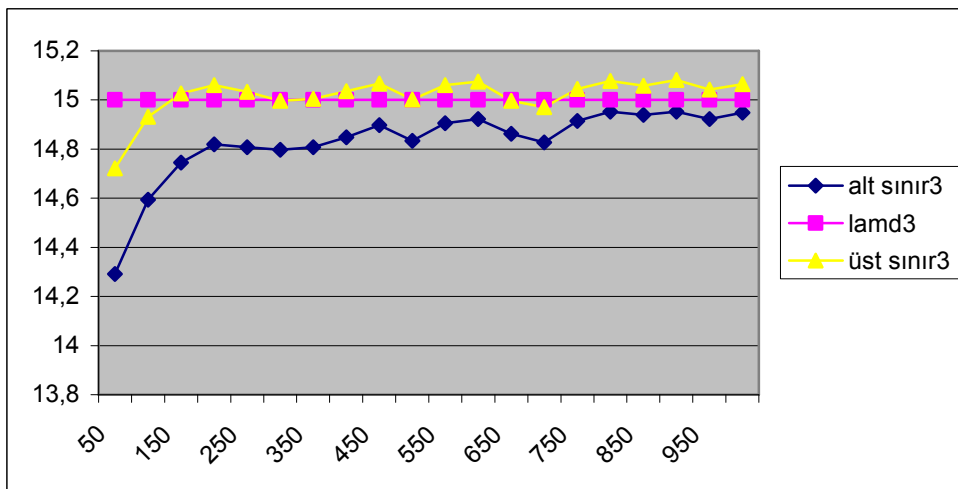
n	alt sınır1	üst sınır1	alt sınır2	üst sınır2	alt sınır3	üst sınır3	alt sınır4	üst sınır4	alt sınır5	üst sınır5
50	201.556	208.600	21.391	22.027	14.291	14.721	9.076	9.369	4.395	4.559
100	202.186	207.041	20.822	21.300	14.594	14.932	9.475	9.693	4.694	4.813
150	198.527	202.435	20.422	20.804	14.745	15.026	9.639	9.848	4.790	4.889
200	199.121	202.559	20.287	20.628	14.819	15.060	9.737	9.914	4.851	4.933
250	201.320	204.453	20.274	20.576	14.807	15.033	9.762	9.762	4.861	4.937
300	198.917	201.807	20.136	20.418	14.797	14.997	9.834	9.979	4.864	4.935
350	200.324	202.985	20.154	20.417	14.808	15.005	9.835	9.970	4.917	4.988
400	198.953	201.289	20.080	20.324	14.848	15.036	9.897	10.026	4.935	4.998
450	198.775	201.110	20.116	20.335	14.897	15.067	9.906	10.025	4.887	4.946
500	200.371	202.496	20.083	20.305	14.834	15.002	9.882	9.995	4.929	4.981
550	199.704	201.886	20.129	20.337	14.905	15.060	9.891	9.996	4.917	4.968
600	199.173	201.240	20.121	20.322	14.922	15.075	9.858	9.961	4.944	4.996
650	199.584	201.566	20.041	20.228	14.862	14.997	9.906	9.998	4.941	4.989
700	199.137	200.976	20.030	20.223	14.827	14.971	9.911	10.000	4.951	4.998
750	198.730	200.536	20.003	20.180	14.914	15.045	9.889	9.975	4.950	4.996
800	199.004	200.771	19.999	20.181	14.952	15.078	9.877	9.970	4.957	5.001
850	199.552	201.308	19.968	20.130	14.939	15.058	9.939	10.019	4.951	4.994
900	199.002	200.653	20.001	20.162	14.952	15.082	9.911	9.994	4.960	5.002
950	199.677	201.285	19.983	20.151	14.922	15.042	9.945	10.027	4.965	5.003
1000	199.662	201.195	19.964	20.117	14.948	15.064	9.924	10.003	4.961	5.000



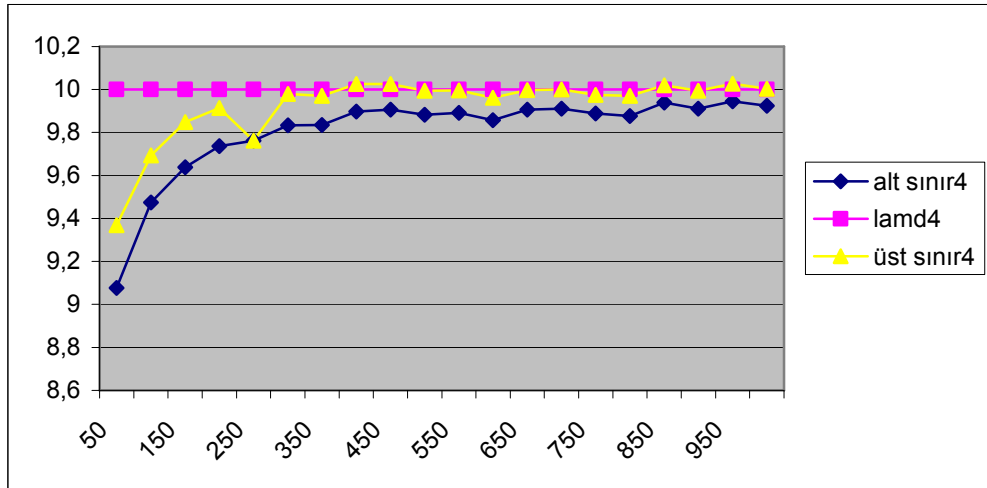
a) Farklı örnek hacimleri için λ_1 'in güven aralığı



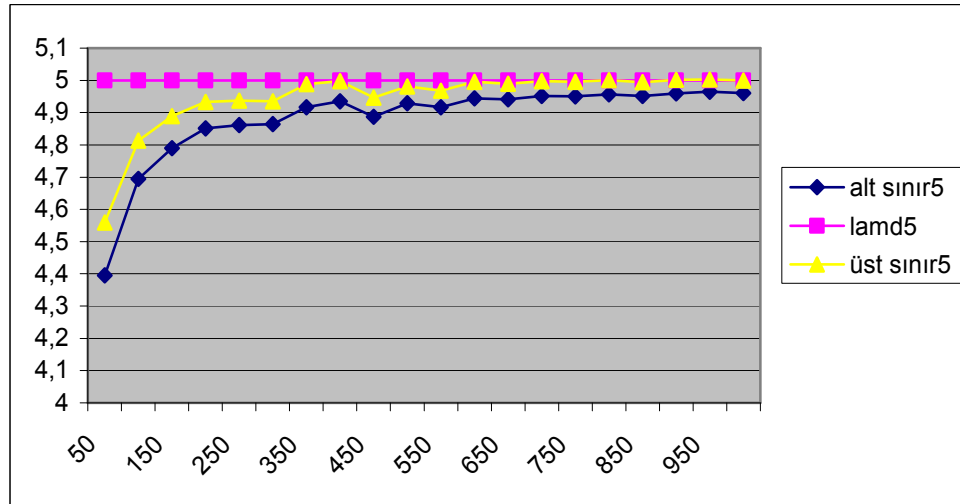
b) Farklı örnek hacimleri için λ_2 'nin güven aralığı



c) Farklı örnek hacimleri için λ_3 'ün güven aralığı



d) Farklı örnek hacimleri için λ_4 'ün güven aralığı



e) Farklı örnek hacimleri için λ_5 'in güven aralığı

Şekil 5.14. Farklı örnek hacimleri için en küçük ve en büyük karakteristik kök arasındaki farkın küçük olduğu matrisin özdeğerlerinin güven aralıkları

SONUÇ

Bu çalışmada kovaryans matrisin karakteristik kökleri için güven aralıkları ve farklı şartlar altında gerçek güven aralıkları incelendi. Monte Carlo simülasyonları, en küçük ve en büyük karakteristik kökler birbirinden çok farklı olduğunda N_c' aralığının gerçek güven seviyesinin daha doğru olduğunu göstermektedir. Örneklem kovaryans matrisinin en küçük ve en büyük karakteristik kökleri için korelasyon katsayısı aynı sonucu destekler; bu fark büyük olduğunda n^{-1} 'in katsayısı küçüktür. Fakat asıl önemli olan p 'nin n 'e yakın olduğunda ve farkın küçük olabileceği durumlarda bu aralığın belirlenmesi ve kullanımınıdır. Ayrıca en küçük ve en büyük karakteristik kökler arasındaki fark küçük olduğunda N_c' ve N_c nin genişliği A_c 'nin genişliğinden daha büyük olabilir.

KAYNAKLAR

1. Alpar, R.,(2003), Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş, Nobel Yayın No:452, Ankara.
2. Anderson, G.A., (1965), “Asymptotic Expansion for The Distribution of The Latent Roots of The Estimated Covariance Matrix”.
3. Anderson, T.W.,(2003), An Introduction to Multivariate Statistical Analysis. Stanford. California.
4. Er, F., Sönmez, H., (2001), İstatistik Formülleri ve Olasılık Dağılımları, Anadolu Üniversitesi, Fen Fakültesi Yayınları, No:21, Eskişehir.
5. Erfon, B., Gong, G., (1983). “Aleisure Look at The Bootsrap, The Jackknife and Cross-Validation “ The American Statistician, Jasa, Vol. 37, No.1, s36-48.
6. Graybill, F.A., (1998), Theory And Application of The Linear Model, North Scituate, Massachusetts.
7. Hair, J.F., Anderson, R.E., Tahtam, R.L., Black, W.C., (1995), Multivariate Data Analysis, United States of America.
8. James, A.T., (1959), “The Distribution of The Covariance Matrix”, Yale University.
9. Johnson, R.A., Wichern, D.W., (1999), Applied Multivariate Statistical Analysis, New Jersey.

10. Kollo, T., Rosen, D., (2005), *Advanced Multivariate Statistics with Matrics*.
11. Miller, I., Miller, M., (2001), *John E. Freund'dan Matematiksel İstatistik*, Çevik Matbaacılık, İstanbul.
12. Morrison, D.F., (1990), *Multivariate Statistical Methods*, United States of America.
13. Muirhead, R.J., Chikuse, Y., (1975), "Asymptotic Expansion for The Joint and Marginal Distribution of The Latent Roots of The Covariance Matrix", Vol.3. No.4. pp.1011-1017.
14. Orhunbilge, N., (2000), *Tanımsal İstatistik Olasılık ve Olasılık Dağılımları*, Avcıol Basım Yayın, İstanbul.
15. Sakaori, F., Yamada, T., Kawamura, A., Sugiyama, T., (2007), "A New Confidence Interval for All Characteristic Roots of A Covariance Matrix ", 22:pp.121-131.
16. Sazak, H.S., Okutan, D., (2009), "Bootstrap Yönteminin Regresyon Analizinde Kullanımı Ve Diğer Yöntemlerle Karşılaştırılması", 6. İstatistik Kongresi, Antalya.
17. Semiz, M., (2007), *İstatistik Laboratuvarı ve SPSS-MP*, Ofset Matbaası, Konya.
18. Sugiyama, T., (1970), "Joint Distribution of The Extreme Roots of A Covariance Matrix ", Vol.41, No.2, pp.655-657.
19. Tatlıdil, H., (1996), *Uygulamalı Çok Değişkenli İstatistiksel Analiz*. Akademi Matbaası, Ankara.

20.Topuz, D., (2002), “Regresyonda Yeniden Örnekleme Yöntemlerinin Karşılaştırmalı Olarak İncelenmesi”, Niğde Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Niğde.

21.Yakupoğlu, Ç., Atıl, H., (2003), " Bootstrap Metodu Ve Uygulanışı Üzerine Bir Çalışma”, Ege Üniversitesi Ziraat Fakültesi Dergisi, İzmir.