Ekonometri ve İstatistik Sayı:1 2005 15-05

İSTANBUL ÜNİVERSİTESİ
İKTİSAT FAKÜLTESİ
EKONOMETRİ VE İSTATİSTİK
DERGİSİ

ASAL BİLEŞENLER ANALİZİNE BOOTSTRAP YAKLAŞIMI

Dr. Aylin Aktükün*

Bu makale 15.12.2004 tarihinde alınmış, hakem kontrolü sonrasında yayını uygun bulunmuştur.

Absract :

In this paper, we apply bootstrap methods to principal components. We use a set of hypothetical data to illustrate how the bootstrap methods can be employed in constructing confidence intervals dealing with the principal components analysis. We write a Mathematica program to handle the all process.

Keywords : Principle components analysis; Bootstrap; Bootstrap quantiles; Bootstrap Confidence Intervals; Mathematica.

Özet :

Bu çalışmada, bootstrap yöntemlerin asal bileşenler analizine uygulanma sürecini sunduk. Hipotetik bir veri ile asal bileşenler analizinde başvurulan bazı güven aralıklarının bootstrap yöntemlerle nasıl gerçekleştirilebileceğini gösterdik. Makaledeki tüm bootstrap süreçleri Mathematica dilinde yazdığımız bir programla gerçekleştirdik.

Anahtar Kelimeler : Asal bileşenler analizi; Bootstrap; Bootstrap kantiller; Bootstrap Güven Aralıkları; Mathematica.

* **Adres :** İstanbul Üniversitesi, İktisat Fakültesi Ekonometri Bölümü, Tel: (0212) 440 00 00-11659
E-Mail : aylin@istanbul.edu.tr

ASAL BİLEŞENLER ANALİZİNE BOOTSTRAP YAKLAŞIMI

1. GİRİŞ

Bootstrap yöntemi literatüre ilk kez Efron 'ın 1979 yılındaki makalesi ile tanıtıldı. Teorik gelişme Freedman (1981) ve Wu (1986) ile devam etti. Daha sonraki gelişmelerden kitaplaştırılanlar ise tarihsel sırasıyla Beran ve Ducharme (1991), Hall (1992), Mammen (1992), Efron ve Tibshirani (1993), Davison ve Hinkley (1997), teorik bir çalışma olan Shao ve Tu (1995) sayılabilir.

Günümüzde bilgisayarların da gelişimine koşut olarak çok fazla sayıda araştırmaya konu olan Bootstrap yönteminde temel düşünce eldeki örneklemi, kitle olarak varsayıp buradan belirli sayıda tekrarlı örnekleme yaparak ilgilenilen tahmincinin suni bir örnekleme dağılımını yaratmaktır. Özellikle söz konusu tahmincinin örnekleme dağılımını asimptotik teori ile elde etmek zor ya da olanaksızsa bootstrap yöntemi yada genel anlamıyla tekrarlı örnekleme yöntemleri güçlü bir potansiyel oluşturmaktadırlar Aktükün (2003).

Çok değişkenli analizin boyut indirgemeci yöntemlerinden olan asal bileşenler analizinde bootstrap yaklaşımı, analizin eksik olan istatistiksel çıkarsama araçlarını tamamlamada önemli bir rol oynayabilir.

Bu çalışmanın amacı, bootstrap yönteminin asal bileşenler analizine uygulanma sürecini sunmak ve uygulama olanaklarını tartışmaktır.

Bilindiği gibi bootstrap süreci bilgisayar iterasyonlarına dayanmakta olup, istatistik paketler bu konuda çok fazla olanak sunmamaktadır. Bu nedenle asal bileşenler analizine bootstrap sürecin uygulanması için *Mathematica* dilinde bir program yazdık ve hipotetik bir veriye uyguladık.

2. ASAL BİLEŞENLER ANALİZİ

Boyut-indirgemeci çok değişkenli analiz tekniklerinden olan asal bileşenler analizine ilişkin temel bilgileri Anderson (1958), Morrison (1976) ve Jobson (1994) gibi hemen her çok değişkenli istatistik analiz kitabında bulabiliriz. Öte yandan analize bootstrap yaklaşımı konusunda literatürün pek yeterli olduğu söylenemez. Beran ve Srivastava (1985) özelinde asal bileşenler olmasa da kovaryans matrisinin bootstrap sürecini asimtotik analizle birlikte ele almış, Efron ve Tibshirani (1993), asal bileşenler analizinde ilgilenilen bazı istatistiklere ilişkin çıkarsama sürecinin bootstrap yöntemlerle gerçekleştirilme olanaklarını sunmuşlardır. Bu çalışmada Efron ve Tibshirani (1993) de tanımlanan süreçlere ek olarak Anderson (1963) 'ın varyanas-kovaryans matrisinin en küçük k tane öz değerinin birbirine eşitliği testine bootstrap sürecini uyguladık

Boyut-indirgemeci çok değişkenli analiz tekniklerinden olan asal bileşenler analizinde temel amaç X_1 , X_2 ,..., X_p rastlantı değişkenlerinin maksimum varyansı veren doğrusal bileşenlerini aramaktır.

X_1 , X_2 ,..., X_p 'nin bir doğrusal bileşeni

$$Y_i = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \ldots + \hat{\beta}_p X_{pi} \quad , \quad i = 1,2,\ldots,n \tag{1}$$

olarak ifade edilebilir. Matrisel olarak ise aynı eşitlik

$$\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} \tag{2}$$

şeklinde yazılabilir. (2) eşitliğindeki vektör ve matrisler açık şekilleriyle

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ . \\ . \\ . \\ \hat{\beta}_p \end{bmatrix} \quad , \quad \mathbf{X} = \begin{bmatrix} X_{11} & X_{21} & \ldots & X_{p1} \\ X_{12} & X_{22} & \ldots & X_{p2} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ X_{1n} & X_{2n} & \ldots & X_{pn} \end{bmatrix} \quad , \quad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix}$$

olacaktır.

Y 'nin örneklem varyansı, **S**, örneklem varyans-kovaryans matrisini göstermek üzere

$$\hat{\mathbf{Var}}(\mathbf{Y}) = \hat{\boldsymbol{\beta}}\,\mathbf{S}\,\hat{\boldsymbol{\beta}}$$

olarak tanımlanabilir. Amaç $\hat{\mathbf{Var}}(\mathbf{Y})$ 'yı maksimum yapacak $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$ katsayılarını bulmak olacaktır. Ancak $\hat{\beta}$ katsayıları üzerine bir kısıtlama yapılmadığı zaman Y 'nin varyansı keyfi olarak büyütülebilir, dolayısıyla belirli, tek bir çözümü elde etmek olanaksızlaşır. Öyleyse $\hat{\boldsymbol{\beta}}\,\boldsymbol{\beta} = 1$ kısıtının katılmasıyla elde edilen $\mathbf{u} = \hat{\boldsymbol{\beta}}\,\mathbf{S}\,\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\beta}}\,\boldsymbol{\beta} - \mathbf{1})$ fonksiyonunun $\hat{\boldsymbol{\beta}}$ 'ya göre kısmi türevini alıp

$$\frac{\partial \mathbf{u}}{\partial \hat{\boldsymbol{\beta}}} = \mathbf{2\,S}\hat{\boldsymbol{\beta}} - \mathbf{2}\,\hat{\boldsymbol{\lambda}}\,\hat{\boldsymbol{\beta}}$$

sıfıra eşitlemeliyiz.

$$\mathbf{2\,S}\hat{\boldsymbol{\beta}} - \mathbf{2}\,\hat{\boldsymbol{\lambda}}\,\hat{\boldsymbol{\beta}} = \mathbf{0}$$

$$(\mathbf{S} - \hat{\boldsymbol{\lambda}}\,\mathbf{I})\,\hat{\boldsymbol{\beta}} = \mathbf{0} \tag{3}$$

(3) eşitliği bir homojen denklem sistemini simgeler. Dolayısıyla denklem sisteminin sıfır-dışı çözümlerinin olması $|\mathbf{S} - \hat{\boldsymbol{\lambda}}\,\mathbf{I}|$ determinantının sıfır olması ile mümkündür :

$$|\mathbf{S} - \hat{\boldsymbol{\lambda}}\mathbf{I}| = \mathbf{0} \tag{4}$$

(4) eşitliği aynı zamanda $\hat{\boldsymbol{\lambda}}$ 'nın değişken olarak yer aldığı bir polinomu simgeler. Eşitlikten ayrıca $\hat{\boldsymbol{\lambda}}$ 'nın **S** matrisinin öz değerler vektörünü simgelediği açıktır. Eşitlik (3) soldan $\hat{\boldsymbol{\beta}}$ vektörü ile çarpılırsa

$$\hat{\boldsymbol{\beta}}\ (\mathbf{S} - \hat{\boldsymbol{\lambda}}\mathbf{I})\ \hat{\boldsymbol{\beta}} = \mathbf{0}$$

$$\hat{\boldsymbol{\beta}}\ \mathbf{S}\ \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\lambda}} \tag{5}$$

eşitliği elde edilir. Amacımız $\hat{\boldsymbol{\beta}}\mathbf{S}\hat{\boldsymbol{\beta}}$ 'yı maksimum yapmak olduğuna göre, bu amaca $\hat{\boldsymbol{\lambda}}$ öz değerler vektöründeki maksimum öz değeri seçerek ulaşabiliriz. Bu öz değer λ_i olsun. p tane öz değer ve dolayısıyla p tane doğrusal bileşen (asal bileşen) Y_i olacaktır. Bu durumda (1) eşitliğindeki $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$ katsayıları da karşılık gelen öz değere ilişkin öz vektörün katsayıları olacaktır. Öz değerlerin özelliğinden

$$\text{İz}\,(S) = \sum_{i=1}^{p} \hat{\lambda}_i$$

ve dolayısıyla

$$\text{İz}\,(S) = \sum_{i=1}^{p} \text{Var}\hat{(}Y_i)$$

yazabiliriz. Öyleyse asal bileşenin, diğer bileşenlere izafi önemi ise

$$\hat{\theta} = \frac{\text{Var}\,(Y_i)}{\sum_{i=1}^{n} \text{Var}\hat{(}Y_i)} = \frac{\hat{\lambda}_i}{\text{İz}\,(S)} = \frac{\hat{\lambda}_i}{\sum_{i=1}^{p} \lambda_i} \tag{6}$$

olarak tanımlanabilir. Aktükün (2002).

Asal bileşenler analizinde daha çok ilgilenilen test istatistikleri varyanslara karşılık gelen öz değerler, ile Eşitlik (6) 'da tanımlanan $\hat{\theta}$ 'dır. Ancak bu istatistiklerin belirli bir örneklemden hesaplanacak nokta tahminleri ile güven aralıkları oluşturmak ve hipotez testi yapmak için klasik teoride, X_i 'lerin bağımsız olmaları ve normal dağılımdan gelmeleri varsayımları yapılır. Bu varsayımların karşılanmaması durumunda ise test istatistiklerinin kullanılması olanaksızlaşır ve Bootstrap yaklaşımı önem kazanır.

Kısım 3 'te, süreçler için 15 gözlem ve 4 değişkenden oluşan hipotetik bir veri kullanıldı. Klasik teoriyle karşılaştırmalar yapabilmek için verideki gözlemler çok değişkenli normal dağılımdan elde edilmiş olup, Ek 1 'de verilmiştir.

3. VARYANS-KOVARYANS MATRİSİNİN ÖZ DEĞERLERİNİN ve ORANLARININ BOOTSTRAP GÜVEN ARALIKLARI

Asal bileşenler analizinde varyans-kovaryans matrisinin öz değerleri lineer bileşimlerin varyanslarını simgeledikleri için bunlara ilişkin güven aralıkları ve hipotez testlerine özellikle önem verilir. Anderson (1963), çok büyük örneklemler için ($n > \infty$) aşağıda tanımlanan L rastlantı değişkeninin standart normal dağılım izlediğini göstermiştir:

$$L = \frac{\hat{\lambda}_i - \lambda_i}{\lambda_i \sqrt{2/(n-1)}} \sim N(0,1)$$

Buradan λ_i için % 100(1 - α) güven aralığı

$$\frac{\hat{\lambda}_i}{1 + z_{\alpha/2}\sqrt{2/(n-1)}} \le \lambda_i \le \frac{\hat{\lambda}_i}{1 - z_{\alpha/2}\sqrt{2/(n-1)}} \tag{7}$$

olacaktır. Burada λ_i , kitle varyans-kovaryans matrisinin bir öz değerini, $\hat{\lambda}_i$ bu öz değerin nokta tahminini n örneklem büyüklüğünü $z_{\alpha/2}$ standart normal dağılım kantilini göstermektedir. Ancak bu güven aralığı asimtotik olduğu için küçük örneklemlerde her zaman iyi sonuç vermeyebilir. Bu anlamda örneklem büyüklüğünden etkilenmeyen Bootstrap yaklaşımı önem kazanır.

Aşağıda Ek 1 'deki verinin varyans kovaryans matrisini ve öz değerlerini görüyoruz.

$$S = \begin{pmatrix} 5.956 & -0.9658 & 0.2935 & -0.3020 \\ -0.9658 & 2.534 & 0.8619 & -2.482 \\ 0.2935 & 0.8619 & 24.54 & 8.131 \\ -0.3020 & -2.482 & 8.131 & 45.29 \end{pmatrix}, \ \lambda 1=48.1875 , \ \lambda 2=21.8736 , \ \lambda 3=6.20675 , \ \lambda 4=2.04776$$

Her bir öz değer için (7) 'den hesaplanan güven aralıkları ise aşağıda verilmiştir:

Asimtotik Güven Aralıkları

	Yüzdeler	Asimtotik Kantiller
λ1		
	0.05	29.7143
	0.15	34.6241
	0.25	38.3985
	0.35	42.0617
48.1875	0.45	46.0026
	0.55	50.5903
	0.65	56.4017
	0.75	64.6754
	0.85	79.2212
	0.95	127.378
λ2		
	0.05	13.4881
	0.15	15.7168
	0.25	17.4301
	0.35	19.0929
21.8736	0.45	20.8818
	0.55	22.9643
	0.65	25.6022
	0.75	29.3579
	0.85	35.9606
	0.95	57.8201
λ3		
	0.05	3.82732
	0.15	4.45972
	0.25	4.94588
	0.35	5.41773
6.20675	0.45	5.92533
	0.55	6.51625
	0.65	7.26478
	0.75	8.33047
	0.85	10.204
	0.95	16.4068
λ4		
	0.05	1.26272
	0.15	1.47137
	0.25	1.63176
	0.35	1.78744
2.04776	0.45	1.95491
	0.55	2.14986
	0.65	2.39682
	0.75	2.74842
	0.85	3.36655
	0.95	5.41299

Yukarıda verilen güven aralıklarını bootstrap karşılıkları için, Efron ve Tibshirani (1993) 'nin tanımladığı süreci uyguladık. Orijinal veriden iadeli olarak B = 1000 tane örneklem çekildi. Her bir örneklemin varyans-kovaryans matrisi ve her bir matrisin öz değerleri hesaplandı. Her bir öz değer için 1000 öz değerin ortalaması (λi_{boot}) bootstrap kitle öz değerlerinin bootstrap tahmini olarak düşünüldü. 1000 öz değerin kantilleri ise aşağıda verilen güven aralıkların oluşturdu. Orijinal veriden elde edilen öz değerlerin bootstrap tahminlere oldukça yakın olduğunu görmekteyiz. Asimtotik güven aralıklarının ise bootstrap güven aralıklarından daha geniş olduğu görülmektedir. Bu durumun örneklemin yeterince büyük olmamasından kaynaklandığın söylemek yanlış olmaz.

Bootstrap Güven Aralıkları

$\lambda 1_{boot}$	Yüzdeler	Bootstrap Kantiller
	0.05	32.3409
	0.15	38.3046
	0.25	42.1568
	0.35	45.4512
51.3153	0.45	48.5666
	0.55	52.379
	0.65	55.508
	0.75	59.828
	0.85	64.9596
	0.95	73.4571
$\lambda 2_{boot}$		
	0.05	11.6622
	0.15	14.3664
	0.25	16.027
	0.35	17.6228
20.2676	0.45	19.2221
	0.55	20.8491
	0.65	22.3307
	0.75	24.1395
	0.85	26.2876
	0.95	30.0632
$\lambda 3_{boot}$		
	0.05	2.10629
	0.15	3.37521
	0.25	4.10639
	0.35	4.58874
5.46056	0.45	5.21157
	0.55	5.78494
	0.65	6.29567
	0.75	6.874
	0.85	7.61334
	0.95	8.65801
$\lambda 4_{boot}$		
	0.05	0.567834
	0.15	0.882356
	0.25	1.07023
	0.35	1.27747
1.49595	0.45	1.43761
	0.55	1.5937
	0.65	1.74414
	0.75	1.91094
	0.85	2.09563
	0.95	2.35796

Asal Bileşenler analizinde öz değerler kadar eşitlik (6) 'daki $\hat{\theta}$ istatistiğine de önem verilir. Her bir öz değerin toplam varyansa katkısını diğer bir deyişle izafi önemimi gösteren bu istatistiğin sayısal değerine bakarak kaç tane asal bileşenle yetinebileceğimize karar verebiliriz. Bu karar için klasik teoride yaygın olarak asimtotik bir test olan Bartlett (1950) 'in küresellik testinin ve bazı grafik yöntemlerin (örn. bkz. Jobson (1994)) kullanıldığını da belirtelim.

Aşağıda verilen, her bir öz değerin toplam öz değerlere oranlarına bakıldığında büyüklük sıralamasına göre birinci ve ikinci öz değerlerin toplamdaki payının 0,89 'dan büyük olduğu dolayısıyla dört değişkenin 2 asal bileşenle ifade edilebileceğini söyleyebiliriz :

$$\hat{\theta}_1 = \frac{\lambda 1}{\lambda 1 + \lambda 2 + \lambda 3 + \lambda 4} = 0.615299$$

$$\hat{\theta}_2 = \frac{\lambda 2}{\lambda 1 + \lambda 2 + \lambda 3 + \lambda 4} = 0.2793$$

$$\hat{\theta}_3 = \frac{\lambda 3}{\lambda 1 + \lambda 2 + \lambda 3 + \lambda 4} = 0.0792531$$

$$\hat{\theta}_4 = \frac{\lambda 4}{\lambda 1 + \lambda 2 + \lambda 3 + \lambda 4} = 0.0261475$$

Öyleyse ilk iki öz değere karşılık gelen öz vektörlerin belirlediği

$$Y_1 = 0{,}00348x_1 + 0{,}0452x_2 - 0{,}323x_3 - 0{,}945x_4$$

ve

$$Y_2 = -0{,}0185x_1 - 0{,}0821x_2 - 0{,}944x_3 + 0{,}319x_4$$

lineer bileşimleri toplam varyansın % 89 'undan fazlasını açıklamaktadır.

(6) eşitliğinde tanımlanan istatistiğe bootstrap uygularken, ilgilenilen $\hat{\theta}$ istatistiğinin değeri her bir bootstrap örneklemi için hesaplanıp, böylelikle elde edilen değerlerin ortalaması ($\hat{\theta}_{boot}$) bulunabilir. Betimlenen bu süreç B = 1000 kez uygulanmış, $\hat{\theta}_1$ ve $\hat{\theta}_2$ için aşağıdaki değerler elde edilmiştir :

$\hat{\theta}_1$ için

$\hat{\theta}_{boot}$	Yüzdeler	BootStrap Kantiller
	0.05	0.528301
	0.15	0.576246
	0.25	0.599673
	0.35	0.620687
0.651049	0.45	0.64253
	0.55	0.661757
	0.65	0.6799
	0.75	0.702222
	0.85	0.72647
	0.95	0.768307

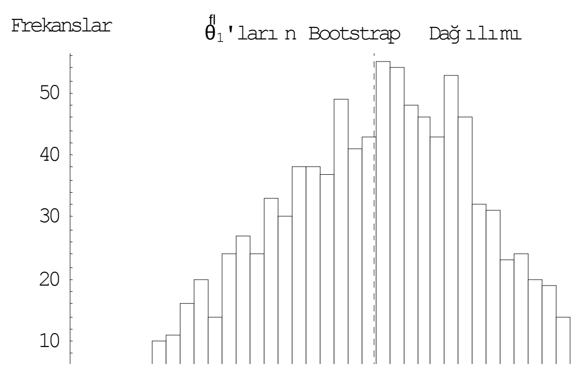
$\hat{\theta}_2$ için

$\hat{\theta}_{boot}$	Yüzdeler	BootStrap Kantiller
	0.05	0.149777
	0.15	0.18668
	0.25	0.207486
	0.35	0.226544
0.259573	0.45	0.245151
	0.55	0.265293
	0.65	0.285755
	0.75	0.307191
	0.85	0.336207
	0.95	0.389565

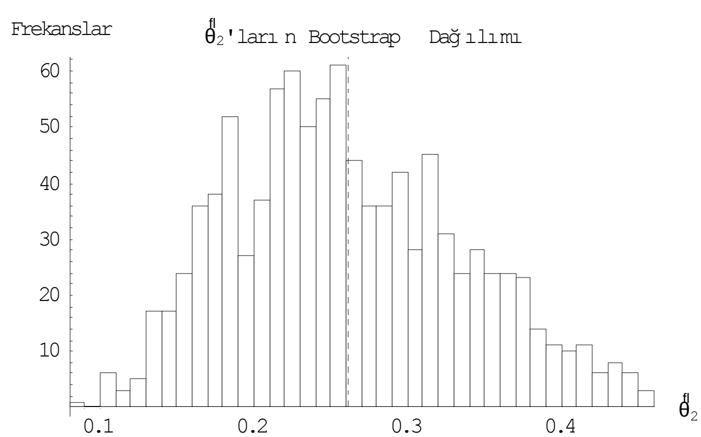
Bootstrap tahminlerinin orijinal örneklemden hesaplanan $\hat{\theta}_1 = 0{,}6510$ ve $\hat{\theta}_2 = 0{,}2595$ yakın olduğu görülmektedir. Bu aşamada $\hat{\theta}_1 = 0{,}6510$ ve $\hat{\theta}_2 = 0{,}2595$ değerleri yardımıyla θ_1 ve θ_2 için güven aralığı oluşturmak istenebilir. Ancak klasik teoride küçük örneklemler için $\hat{\theta}_i$'nın standart hatasına ilişkin analitik bir formülün olmaması, bu sürecin uygulanmasına engeldir. Orijinal örnekleme, bootstrap süreci uygulandıktan sonra $\hat{\theta}$ 'ya yada daha genel olarak ifade edersek test istatistiğine ilişkin ampirik bir dağılım elde edileceği için ilgilenilen θ_1 ve θ_2 'ye ilişkin güven aralıkları ampirik dağılımın kantilleri diğer adıyla bootstrap kantilleri hesaplanarak oluşturulabilir. Yukarıdaki tablodan θ_1 için soldan ve sağdan % 5 'i dışarıda bırakan değerler 0,528 ve 0,768 'dir. Öyleyse ampirik olarak [0,528 - 0,768] aralığı % 90 güven aralığı olarak kabul edilebilir. Aynı aralık θ_2 için [0,149 - 0,389] olacaktır.

Şekil 1 'de bootstrap örneklemlerinden elde edilen değerlerin oluşturduğu histogramlar yer almaktadır Kesikli doğru, dağılımların ortalamasını yani $\hat{\theta}_{boot}$ değerini göstermektedir

$\hat{\theta}_1$ için



$\hat{\theta}_2$ için



Şekil 1: $\hat{\theta}_1$ ve $\hat{\theta}_2$ 'nın Bootstrap Dağılımları

4. SONUÇ

Belirli bir istatistiğin standart hatasını veren bir formülün olmadığı yada asimtotik bir formülün olduğu durumlarda küçük örneklemler için güvenilir istatistiksel çıkarsama yapma da bootstrap yaklaşımı benimsenebilir. Daha çok betimsel bir teknik olan ve bu nedenle çıkarsama araçları çok da yeterli olmayan asal bileşenler analizine bootstrap yaklaşımı bu anlamda önem kazanabilir. Bu çalışmada asal bileşenler sürecinde ilgilenilen bazı istatistiklere bootstrap sürecinin uygulanma yöntemlerini tanıtmayı amaç edindik. İstatistik yazılımlar bu konuda yeterli olanakları sunmadıkları için söz konusu süreçleri *Mathematica* dilinde yazdığımız bir programla gerçekleştirdik. Kodlar Ek 2 'de sunulmuştur.

KAYNAKÇA

- Aktükün, L.A. (2002), **Bootstrap Yöntemler ve Doğrusal Regresyon Modellerine Uygulanması**, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.

- Anderson, T.W. (1958), **An Introduction to Multivariate Analysis**, Wiley, NewYork.

- Anderson, T.W. (1963), "Asymptotic theory for principal components", **Annals of Mathematical Statistics**, 34, 122-148.

- Bartlett, M.S. (1950), "Test of significance in factor analysis", **British Journal of Psychology**, 3, 77-78.

- Beran, R., Srivastava, M.S., (1985), "Bootstrap tests and confidence regions for functions of covariance matrix", **Ann. Statist.**, 13, 95-115.

- Beran, R., Ducharme, G.R. (1991), **Asymptotic Theory for Bootstrap Methods in Statistics**, Les Publications Centre de Recherches Mathematiques, Université de Montréal, Montréal.

- Davison, A.C., Hinkley, D.V. (1997), **Bootstrap Methods and Their Application**, Cambridge, Cambridge University Press.

- Efron, B. (1979), "Bootstrap Methods: Another look at the Jackknife", **Ann. Statist.**, Vol.7, s.1-26.

- Efron, B., Tibshirani, R.J. (1993), **An Introduction to The Bootstrap**, USA, Chapman&Hall.

- Freedman, D.A. (1981), "Bootstrapping regression models.", **Ann. Statist.**, Vol.9, s.1218-1228.

- Hall, P. (1992), **The Bootstrap and Edgeworth Expension**, New York, Springer.

- Jobson, J.D. (1994), **Applied Multivariate Data Analysis : Volume II: Categorical and Multivariate Methods**, Springer, NewYork.

- Mammen, E. (1992), **When Does Bootstrap Work? Asymptotic Results and Simulations**, Volume 77 of Lecture Notes in Statistics, New York, Springer.

- Morrison, D.F. (1976), **Multivariate Statistical Methods**, Second Ed., McGraw-Hill, NewYork.

- Shao, J., Tu, D. (1995), **The Jackknife and Bootstrap**, New York, Springer.

- Wu, F.J. (1986), "Jackknife, bootstrap and other resampling methods in regression analysis (With discussions)", **Ann. Statist.**, Vol.14, s.1261-1350.

Ek 1. Çok Değişkenli Normal Dağılımdan Türetilen Hipotetik Veri

x_1	x_2	x_3	x_4
14.2324	54.8199	61.6357	62.6974
15.4832	53.9587	53.6667	46.2785
17.9279	56.7678	56.4821	48.1623
13.3259	55.2607	60.0429	60.6884
13.3767	54.9219	52.6191	52.3652
10.8281	58.2343	49.5043	43.9428
13.7561	53.4531	56.1293	59.3454
14.9728	51.9761	53.3387	59.9688
16.5149	56.6812	59.5466	54.7969
18.434	52.4669	53.8989	49.3047
21.0525	55.0778	59.2011	57.9376
16.9422	53.3292	46.3162	53.6431
14.7467	55.1901	53.0857	69.844
16.1851	55.2698	47.8543	60.8118
13.5081	54.9659	64.248	57.1117

Ek2. Mathematica Kodları

$\hat{\theta}_i$ ' nın Bootstrap Tahminleri, Güven Aralıkları ve Histogramlar

```
Needs[Statistics`MultiDescriptiveStatistics`]
Needs[Statistics`DescriptiveStatistics`]
Needs[Statistics`ConfidenceIntervals`]
Needs[Graphics`Graphics`]
Needs[DiscreteMath`Combinatorica`]
S = N[N[CovarianceMatrix[data]] * (Length[Transpose[data][[1]]] - 1) /
                                 Length[Transpose[data][[1]]], 4];
özdeğerler = N[Eigenvalues[S], 3]
özvektörler = N[Eigenvectors[S], 3];
p1 = Prepend[özdeğerler, özdeğerler];
p2 = Prepend[özvektörler, özvektörler];
Print[TableForm[{p1, p2}]]; B = 1000;
popboot = Flatten[Join[Table[data, {i, 1, Length[data]}]], 1];
Do[eboot[i] = RandomKSubset[popboot, Length[data]], {i, 1, B}];
eboos = Table[eboot[i], {i, 1, B}];
kovboot = N[Table[CovarianceMatrix[eboos[[i]]], {i, 1, B}]];
eigboot = N[Table[Eigenvalues[kovboot[[i]]], {i, 1, B}]];
k = Length[Transpose[data]];
birler = Table[1, {i, 1, k}];
total = {birler}.Transpose[eigboot];
özdeğerler = Table[eigboot[[i]] / total[[1, i]], {i, 1, B}];
ilközdeğerler = Transpose[özdeğerler][[1]];
ortalamanınstandarthatası1 = StandardDeviation[ilközdeğerler];
mean1 = Mean[ilközdeğerler];
q95 = Quantile[ilközdeğerler, 0.95]; q90 = Quantile[ilközdeğerler, 0.90];
q84 = Quantile[ilközdeğerler, 0.84]; q50 = Quantile[ilközdeğerler, 0.50];
q05 = Quantile[ilközdeğerler, 0.05];
kantiller = Table[{m, Quantile[ilközdeğerler, m]}, {m, .05, .95, .1}];
kantiller = Table[{Quantile[ilközdeğerler, m]}, {m, .05, .95, .1}];
Print[TableForm[{{θ̂boot, Yüzdeler, BootStrap Kantiller}, {mean1, yüzdeler, kantiller}}]];
yüzdeler = Table[{m}, {m, .05, .95, .1}];
Print[TableForm[{{Yüzdeler, BootStrap Kantiller}, {yüzdeler, kantiller}}]];
ListPlot[ilközdeğerler, AxesLabel -> {B, "θ̂1"}]
hist100 = Histogram[ilközdeğerler, BarStyle → {RGBColor[1, 1, 1]},
        Epilog → {Dashing[{.01}], Line[{{mean1, 0}, {mean1, B}}]},
  AxesLabel → {"\^θ1", Frekanslar}, PlotLabel → "θ̂1'ların Bootstrap Dağılımı, ImageSize → {350, 350}]
ikinciözdeğerler = Transpose[özdeğerler][[2]];
ortalamanınstandarthatası3 = StandardDeviation[ikinciözdeğerler];
mean2 = Mean[ikinciözdeğerler];
q95 = Quantile[ikinciözdeğerler, 0.95]; q90 = Quantile[ikinciözdeğerler, 0.90];
q84 = Quantile[ikinciözdeğerler, 0.84]; q50 = Quantile[ikinciözdeğerler, 0.50];
q05 = Quantile[ikinciözdeğerler, 0.05];
kantiller = Table[{m, Quantile[ikinciözdeğerler, m]}, {m, .05, .95, .1}];
kantiller = Table[{Quantile[ikinciözdeğerler, m]}, {m, .05, .95, .1}];
yüzdeler = Table[{m}, {m, .05, .95, .1}];
Print[TableForm[{{θ̂boot, Yüzdeler, BootStrap Kantiller}, {mean2, yüzdeler, kantiller}}]];
ListPlot[ilközdeğerler, AxesLabel -> {B, "θ̂2"}]
hist100 = Histogram[ikinciözdeğerler, BarStyle → {RGBColor[1, 1, 1]},
        Epilog → {Dashing[{.01}], Line[{{mean2, 0}, {mean2, B}}]},
  AxesLabel → {"\^θ2", Frekanslar}, PlotLabel → "\^θ2'ların Bootstrap Dağılımı", ImageSize → {350, 350}]
```