



## Bootstrap Tahminini Kullanarak Pearson Korelasyon Katsayısının Önemiğinin Araştırılması

Tolga ZAMAN<sup>\*1</sup>, Kamil ALAKUŞ<sup>2</sup>

<sup>1</sup>*Çankırı Karatekin Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 18100, Çankırı, Türkiye*  
<sup>2</sup>*Odokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 55139, Samsun, Türkiye*

\*yazılan yazar e-posta: [tolgazaman@karatekin.edu.tr](mailto:tolgazaman@karatekin.edu.tr)

(Alınış / Received: 17.09.2018, Kabul / Accepted: 08.03.2019, Yayınlanma / Published: 31.05.2019)

**Özet:** Yeniden örnekleme yöntemleri parametrelerin etkili tahminlerini ve asimptotik dağılımlarını sunar. Bu çalışmada, basit doğrusal regresyon modeline bootstrap yöntemi uygulanarak Pearson korelasyon katsayısının anlamlılığının değerlendirilmesinde klasik ve çakı (birini dışarıda bırak işlemi) test istatistiklerine alternatif olarak bootstrap yönteminin kullanılması önerilmektedir. Bu prosedür Pearson korelasyon katsayısının anlamlılığını test etmek için etkili bir alternatif sağlar. Uygulamada bootstrap ve çakı yöntemlerine ait model parametreleri, standart hatalar, Pearson korelasyon katsayısı, yanlılık ve %95 güven aralığı gerçek bir veri yardımıyla elde edilerek sonuçlar yorumlanmıştır. Sonuçta, bootstrap yöntemi ile elde edilen test istatistiği, klasik ve çakı tabanlı test istatistiğine alternatif bir yöntem olarak önerilmiştir.

**Anahtar kelimeler:** Bootstrap, Çakı, Basit lineer regresyon, Korelasyon katsayısı, EKK tahmincileri, Yanlılık.

### Investigating the Significance of a Correlation Coefficient using Bootstrap Estimates

**Abstract:** Resampling methods offers effective estimates of parameters and its asymptotic distribution. In this study, it is recommended to use the bootstrap method as an alternative to the classical and knife (one exclusion procedure) test statistics in evaluating the significance of the Pearson correlation coefficient by applying the bootstrap method to the simple linear regression model. This procedure provides an effective alternative to test the significance of the Pearson correlation coefficient. In the application, the model parameters, standard errors, Pearson coefficients of correlation, bias and % 95 confidence intervals belonging to bootstrap and jackknife methods in estimated with the help of a real data and the obtained results are interpreted. As a result, the test statistic obtained by the bootstrap method is proposed as an alternative to the classical and jackknife test statistics.

**Keywords:** Bootstrap, Jackknife, Simple linear regression, Correlation of coefficient, OLS estimators, Bias.

### 1. Giriş

Modern bilgisayarlar ve yazılımlar sayesinde veriler grafiksel ve sayısal olarak oldukça kolay yorumlanabilmektedir. Bu sayede daha doğru ve bilgilendirici analiz yapmamıza olanak tanır. Yeniden örnekleme yöntemleri de bu yeniliğin bir parçasıdır. Yeniden örnekleme yöntemleri, standart hataları ve güven aralıkları hesaplayıp anlamlılık testleri yaparak belirsizlik miktarını belirlememizi sağlarlar. Uygulamalı istatistikte, parametrelerin tahmininden sonra, standart hata ve parametreler için güven aralıklarının oluşturulması ve tahmin edicinin doğruluğunun değerlendirilmesi her zaman çok önemli olmuştur [1].

Bootstrap ve çakı yöntemleri parametre tahmini ve ilgilenilen istatistiğin asimptotik dağılımı hakkında etkili sonuçlar vermektedirler.

Lineer regresyonda parametrelerin örnekleme dağılımını tahmin etmek için bootstrap ve çakı yöntemlerinin kullanılması ilk kez Efron [2] tarafından önerilmiş ve Freedman [3] ve Wu [4] tarafından geliştirilmiştir [5]. Son yıllarda, bilgisayar teknolojisindeki gelişmeler karşısında, yeniden örnekleme yöntemlerinin kullanılması ile daha etkin ve tutarlı parametre tahminleri elde edilmektedir.

Bu çalışmada, basit doğrusal regresyonda bootstrap, çakı ve en küçük kareler (EKK) yöntemleri ile pearson korelasyon katsayısının istatistiksel olarak anlamlılığı araştırılmıştır. Araştırmanın amacı, parametlerin bootstrap tahminlerini kullanarak pearson korelasyon katsayısının anlamlılığının değerlendirmesinde klasik ve çakı test istatistiklerine alternatif bir test istatistiğinin geliştirilmesidir.

## 2. Materyal ve Metot

Kullanılacak doğrusal regresyon modeli matris notasyonu ile

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \quad (1)$$

eşitliği ile ifade edilir. Bu eşitlikte

$\underline{Y}$ ,  $Y$  bağımlı değişkenine ait gözlemlerin  $n \times 1$ 'lik bir sütun vektörüdür.

$X$ , bilinen sabitlerin (açıklayıcı değişkenlerin)  $n \times p$ 'lik bir matrisidir.

$\underline{\beta}$ , bilinmeyen regresyon katsayılarının  $p \times 1$ 'lik bir sütun vektörüdür.

$\underline{\varepsilon}$ , tesadüfi hataların  $n \times 1$ 'lik bir sütun vektörüdür.

$\underline{\varepsilon}$  vektörünün bağımsız ve aynı dağılıma sahip olduğu,  $Var(\underline{\varepsilon}) = \sigma^2 I_n$  varsayılır. Burada  $I_n$ ,  $n \times n$ 'lik birim matris ve  $\sigma^2 > 0$  olan bir sabittir [24]. Burada kurulan bu regresyon modelinin parametre tahminleri EKK yöntemi ile;

$$\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{Y} \quad (2)$$

eşitliği ile elde edilir. EKK yöntemi ile tahmin edilen regresyon katsayılarının kovaryans matrisi ise

$$Var(\underline{\hat{\beta}}) = \hat{\sigma}^2(X'X)^{-1}_{p \times p} \quad (3)$$

eşitliği ile verilir. Tahmin edilen regresyon katsayılarının varyansları (3) eşitliğinin esas köşegen elemanları olup esas köşegen dışındaki değerler ise kovaryanslarıdır. Eşitlik (3)'ün köşegen elemanları EKK tahmin edicilerinin varyansları ve köşegen dışı elemanlar ise aynı parametrelerin kovaryanslarıdır.

Ayrıca, doğrusal ilişkinin gücünü ölçen korelasyon katsayısının EKK tahmini de Pearson çarpım momenti tahmini ile verilmiştir. Bu eşitlik ise şu şekildedir;

$$\hat{\rho}_{x,y} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{(n \sum_{i=1}^n x_i^2 - [\sum_{i=1}^n x_i]^2)(n \sum_{i=1}^n y_i^2 - [\sum_{i=1}^n y_i]^2)}} \quad (4)$$

Ayrıca bu parametrenin istatistiksel olarak anlamlılığı aşağıdaki hipotez ve test istatistiği ile test edilir.

Hipotez;

$$H_0: \rho_{x,y} = 0$$

$$H_1: \rho_{x,y} \neq 0$$

Test İstatistiği;

$$\frac{\hat{\rho}_{x,y}\sqrt{n-2}}{\sqrt{1-\hat{\rho}_{x,y}^2}} \sim t_{(n-2)} \quad (5)$$

Eşitlik (5)'de verilen test istatistiği klasiktir. Çakı temelli test istatistiği Akpanta and Okorie (2015) [6] tarafından

$$\frac{\hat{\rho}_{x,y}(\zeta)}{\sqrt{\text{var}(\hat{\rho}_{x,y}(\zeta))}} \sim t_{n-2} \quad (6)$$

şeklinde önerilmiştir. Bu çalışmada ise yukarıdaki hipotezin bootstrap temelli test istatistiği

$$\frac{\hat{\rho}_{x,y}(b)}{\sqrt{\text{var}(\hat{\rho}_{x,y}(b))}} \sim t_{n-2} \quad (7)$$

şeklinde hesaplanacaktır. Önerilen bootstrap temelli test istatistiği ile klasik ve çakı temelli test istatistikleri karşılaştırılarak sonuçlar yorumlanacaktır.

## 2.1 Çakı yöntemi

Çakı yöntemi ilk olarak tahmin edicilerin yanlılığını azaltmak için Quenouille tarafından (1949;1956) ortaya atılmıştır [7-8]. Tukey (1958) [9] tarafından tahmin ve güven aralıklarını hesaplamak için geliştirilmiştir [10]. Çakı yöntemi her denemede bir gözlem değerinin dışarıda bırakılması mantığına dayanır ve her defasında geri kalan gözlemlerden  $\beta$  parametreleri için istatistikler hesaplanır. Daha sonra  $i$ 'nci eleman çifti örneklemden sırayla çıkarılır ve model parametreleri eksiltilmiş örneklemden tahmin edilir. Bunun sonucunda çıkan tahmin ediciler kısmi tahmin edici (sözde tahmin edici) olarak adlandırılır [11]. Sözde tahmin edicilerin ortalaması ana parametre değerinin yerine kullanılan Çakı tahmin edicisi olarak anılır [12]. Bu şekilde  $n$  tane gözlemden sadece  $n$  tane farklı örneklem oluşturulabilir. Parametrelerin istatistiksel olarak anlamlılığının test edilmesi ve güven aralıklarının oluşturulması için, parametrelerin standart hata tahminleri de sözde tahmin edicilerden hesaplanır [13]. Çakı yöntemine ait tahminlerin hesaplanmasında tablo 1.'deki formüller kullanılır.

**Tablo 1.** Çakı Tahminlerinde Yanlılık ve Varyans Formülleri

Çakı tahminleri	Varyans	Yanlılık
$\hat{\beta}_{(j)\zeta} = \frac{\sum_{i=1}^n \hat{\beta}_{(j)\zeta i}}{n}$	$\text{Var}(\hat{\beta}_{\zeta}) = \frac{\sum_{i=1}^n (\hat{\beta}_{(j)\zeta i} - \hat{\beta}_{\zeta})^2}{n(n-1)}$	Yanlılık = $\beta^{\text{ekk}} - \hat{\beta}_{(j)\zeta}$
$\hat{\rho}_{x,y}(\zeta) = \frac{\sum_{i=1}^n (\hat{\rho}_{x,y}(\zeta))}{n}$	$\text{Var}(\hat{\rho}_{x,y}(\zeta)) = \frac{\sum_{i=1}^n (1 - \hat{\rho}_{x,y}^2/n - d - 2)}{n-2}$	Yanlılık = $\hat{\rho}_{x,y} - \hat{\rho}_{x,y}(\zeta)$

Burada,  $\hat{\beta}_{(j)\zeta}$  ve  $\hat{\rho}_{x,y}(\zeta)$  ifadeleri (2) ve (4) numaralı eşitliklere bağlı olarak hesaplanan parametrelerin çakı tahminleridir.  $\text{Var}(\hat{\beta}_{\zeta})$  ve  $\text{Var}(\hat{\rho}_{x,y}(\zeta))$  ifadeleri ise bu tahminlerin varyanslarını göstermektedir. Basit doğrusal regresyon modelinde çakı yönteminin işleyiş algoritması üç adımdan oluşmaktadır. Bu adımlar takipte verildiği gibidir.

1)  $(y_i, x_i)$ ;  $i = 1, 2, \dots, n$  şeklindeki açıklayıcı ve bağımlı değişkenlerden  $(n)$  büyüklüğünde bir çift bağımsız örneklemin kullanılması.

2) Her iki değişkenden birinci gözlemin çıkarılması ve  $(n-1)$  gözlem kullanarak, en küçük kareler tahminlerinin  $(\hat{\beta}_{(0)\zeta_1}$  ve  $\hat{\beta}_{(1)\zeta_1})$  ve korelasyon katsayılarının  $(\hat{\rho}_{x,y(\zeta_1)})$  tahmin edilmesi ardından ikinci gözlemin çıkarılması ve ilk çıkarılan gözlemin geri konulması ile  $(n-1)$  gözlem kullanarak, en küçük kareler tahminlerinin  $(\hat{\beta}_{(0)\zeta_2}$  ve  $\hat{\beta}_{(1)\zeta_2})$  ve korelasyon

katsayılarının ( $\hat{\rho}_{x,y(\zeta_2)}$ ) tahmin edilmesi ve bu şekilde devam ederek (i)'inci gözlemin çıkarılması ve (i - 1)'inci çıkarılan gözlemin geri konulması ile (n - 1) gözlem kullanarak, en küçük kareler tahminlerinin ve korelasyon katsayısının ( $\hat{\beta}_{(0)\zeta_i}$  ve  $\hat{\beta}_{(1)\zeta_i}$ ,  $i = 3,4,5, \dots, n$ ) tahmin edilmesi ve bu işlemin çiftteki tüm gözlemler sırasıyla çıkarılana ve geri koyulana kadar tekrarlanması [6].

3) Parametlerin çakı tahminlerinin, yan miktarlarının ve standart hatalarının Tablo 1.'deki formüllerin kullanılarak hesaplanması.

## 2.2 Bootstrap yöntemi

Bootstrap ilk kez Efron (1979) [14] tarafından ileri sürülmüştür. Gerçek veri setinden yeniden örnekleme yöntemidir [15]. Yani metodun temel mantığı, herhangi büyüklükteki mevcut veri setindeki gözlemlerin şansa bağlı olarak yer değiştirerek yeniden örnekleme ile yeni veri setleri oluşturmaktır. Böylece mevcut veri setinden mümkün olduğunca fazla miktarda bilgi alınabilmektedir. Yöntemde ilk olarak original veri kümesinden yerine koyma yöntemi ile örnekler seçilerek bir bootstrap örnekleme oluşturulmaktadır. Ardından bu şekilde çok sayıda bootstrap örnekleme oluşturulabilir ve bu örneklemler yardımıyla ilgilenilen tahmin ediciye ait güven aralığı ve testin istatistiksel anlamlılığı hesaplanabilir [16]. Orijinal veri setinden elde edilen bootstrap örneklemleri uygulamaya bağlıdır. Aslında  $n$  genişliğindeki bir örnekten teorik olarak en fazla  $n^n$  sayıda bootstrap örneği oluşturmak mümkündür [17].

Bootstrap yöntemi kökeninde tahmin değerinin örnekleme dağılımını elde edip, bu dağılıma dayanarak kitle parametre değerinin belirsizliğini değerlendirmektir. Bootstrap yöntemini aşağıdaki şekilde özetleyebiliriz,

$n$  adet örneklemeden oluşan  $\beta = (\beta_1, \beta_2, \dots, \beta_n)$  orijinal veri kümesi kullanılarak iadeli ve tesadüfi olarak yapılan örnekleme ile yeni bir veri seti,  $\beta = (\beta_1^*, \beta_2^*, \dots, \beta_n^*)$  elde edilmektedir. Orijinal veri setinden örnekleme yardımıyla oluşturulan veri kümesi içinde  $\beta_i$  verilerinin birden fazla ya da hiç görülme olasılığı bulunmaktadır [18]. Yeniden örnekleme işlemi istenildiği kadar yinelenerek birbirinden farklı  $B$  adet bootstrap veri kümesi oluşturulabilir ve  $b = 1, \dots, B$  için oluşan bootstrap veri kümesi  $\beta_b^*$  ile gösterilebilir. İlgilenilen istatistik bu yeni veri kümeleri kullanılarak hesaplanmaktadır [19]. Bootstrap yöntemi standart sapma ve güven aralığı hesaplama gibi parametrik olmayan tahminleme problemlerinde kullanılan basit ve güvenilir bir yöntemdir.

Regresyon analizinde hata terimlerinin tekrarlanması ile uygulanan bu yöntem, 1979 yılında Bradley Efron tarafından ileri sürülmüş ve klasik en küçük kareler yönteminden daha etkili parametre tahminleri elde etmek amacı ile geliştirilmiştir. Bu yöntem hata teriminin yeniden örnekleme olarak bilinir. Bu bağlamda bootstrap yönteminin işleyiş algoritmasını aşağıdaki şekilde incelenebilir [20];

- 1) Kitleden tesadüfi olarak  $n$  sayıda bir örnek seçilir.
- 2) Seçilen bu örneğe EKK yöntemi uygulanır.
- 3) Bu modelden  $e_i$  değerleri hesaplanır.
- 4) Elde edilen  $e_i$  değerlerine  $1/n$  olasılığı verilerek her biri  $n$  hacminde  $B$  tane bootstrap hata alt örnekleme oluşturulur. Böylece deneysel dağılım fonksiyonu elde edilir.
- 5) Oluşan deneysel dağılım fonksiyonundan bootstrap hata değerlerinin ortalaması

$$\bar{\hat{\epsilon}}_i^* = \frac{\sum_{b=1}^B \hat{\epsilon}_{bi}}{B} \quad (8)$$

şeklinde hesaplanır.

6) Elde edilen  $\bar{\varepsilon}_i^*$  değerleri 2. adımda oluşturulan modeldeki  $e_i$ 'ler yerine konarak,

$$Y_i^* = \hat{\beta}X + \bar{\varepsilon}_i^* \quad (9)$$

şeklinde bootstrap  $Y_i^*$  değerleri hesaplanır.

7)  $Y_i^*$  ve  $X$ ' den hareketle  $\beta$ 'nin bootstrap tahmin edicisi, en küçük kareler yöntemi ile,

$$\hat{\beta}^* = (X'X)^{-1}X'Y_i^* \quad (10)$$

şeklinde hesaplanır [21]

8) Elde edilen  $Y_i^*$  değerleri ile açıklayıcı değişken(ler) yardımıyla bootstrap temelli korelasyon katsayısı hesaplanır.

Özetle bootstrap yöntemine ait tahminlerin hesaplanmasında Tablo 2'deki formüller kullanılır.

**Tablo 2.** Bootstrap Tahminlerinde Yanlılık ve Varyans Formülleri

Bootstrap tahminleri	Varyans	Yanlılık
$\bar{\beta}_B^* = \frac{\sum_{b=1}^B \hat{\beta}_b^*}{B}$	$Var(\bar{\beta}_B^*) = \frac{\sum_{b=1}^B (\hat{\beta}_b^* - \bar{\beta}_B^*)^2}{B-1}$	Yanlılık = $\hat{\beta}^{ekk} - \hat{\beta}_b$
$\hat{\rho}_{x,y}(b) = \frac{\sum_{b=1}^B (\hat{\rho}_{x,y}(bi))}{B}$	$V(\hat{\rho}_{x,y}(b)) = \frac{\sum_{b=1}^B (1 - r_{x,y}^2/n - 2)}{B}$	Yanlılık = $\hat{\rho}_{x,y} - \hat{\rho}_{x,y}(b)$

Burada,  $\bar{\beta}_B^*$  ve  $\hat{\rho}_{x,y}(b)$  ifadeleri (2) ve (4) numaralı eşitliklere bağlı olarak hesaplanan parametrelerin bootstrap tahminleridir.  $Var(\bar{\beta}_B^*)$  ve  $V(\hat{\rho}_{x,y}(b))$  ifadeleri ise bu tahminlerin varyanslarını göstermektedir.

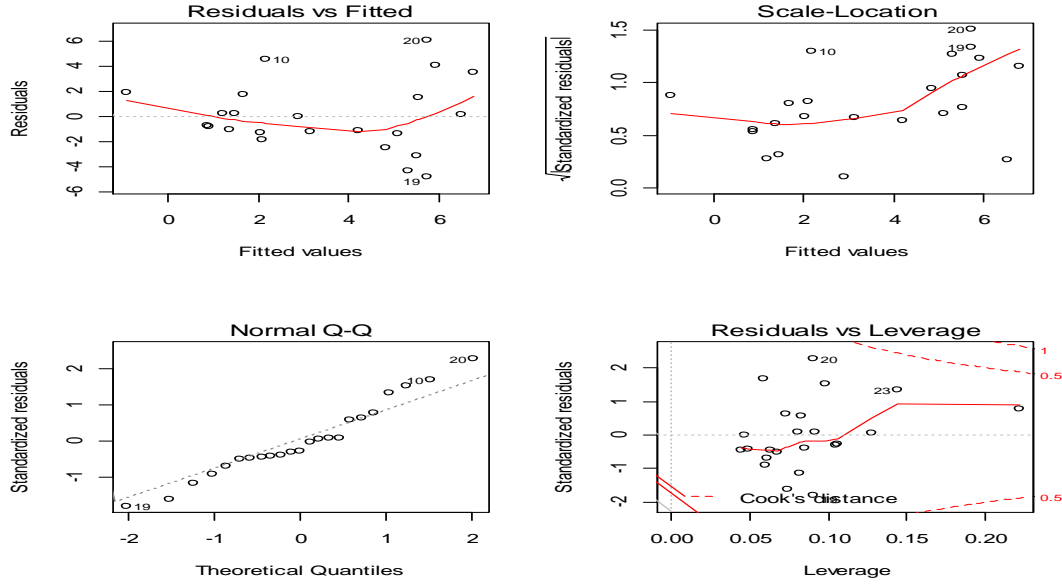
### 3. Uygulama

Çalışmada kullanılan veri seti [www.statstutor.ac.uk/resources/uploaded/sparmans.pdf](http://www.statstutor.ac.uk/resources/uploaded/sparmans.pdf) sitesindeki beşinci sayfadan alınmış olup Tablo 3.'de verildiği gibidir [22].

**Tablo 3.** Uranyum Verisi.

Sıra No	Uranyum Konsantrasyonu	Toplam Çözünmüş Katı	Sıra No	Uranyum Konsantrasyonu	Toplam Çözünmüş Katı
1	302.38	0.97	13	956.06	3.1
2	531.16	0.13	14	1034.55	2.41
3	533.99	0.09	15	1069.82	3.77
4	573.14	1.41	16	1095.42	0.98
5	593.7	0.37	17	1122.58	2.46
6	605.51	1.72	18	1124.17	7.09
7	633.25	3.4	19	1149.38	0.96
8	678.1	0.8	20	1149.6	11.8
9	686.51	0.26	21	1172.84	9.97
10	696.96	6.76	22	1247.95	6.7
11	788.36	2.87	23	1282.95	10.27
12	818.93	1.93	24		

Bu veriye ait basit doğrusal regresyon analizi R paket program dili kullanılarak çözümlenmiştir. Öncelikle tahmin modeli kullanılarak elde edilen artık teriminin serpmeye diyagramı Şekil 1. 'de verilmiştir.



Şekil 1. Tahmin modeli kullanılarak elde edilen artık terimine ait grafikler

Şekil 1. incelendiğinde veride 10 ve 20. değerlerin sapan değerler olduğu görülmektedir. Veride sapan değer(ler) varken parametrik olmayan regresyon yöntemlerine başvurmak yerine basit doğrusal regresyon analizinde yeniden örnekleme yöntemleri birer düzeltme terimi olarak kullanılabilir. Bu veri setine ait basit doğrusal regresyon sonuçları Tablo 4.'de verilmiştir.

Tablo 4. Basit Doğrusal Regresyon Sonuçları.

Değişken	Tahmin Değerleri		t-testi	Önem Olasılığı
	B	Std. Hata		
Sabit	-3.302	1.906	-1.733	0.098
Uran.Kon. ( $X_1$ )	0.008	0.002	3.742	0.001
$R^2 = 0.40$ ; $\hat{\sigma} = 2.793$ ; $r = 0.632$				
Modelin önemlilik testi, $F=14$ olup; önem olasılığı ise 0.001'dir				

Tablo 4. incelendiğinde model için tahmin edilen eğim katsayısı istatistiksel olarak önemli olduğu görülmektedir ( $p < 0.05$ ). Bu değişkenler yardımıyla kurulan basit doğrusal regresyon modeli de istatistiksel olarak anlamlı bulunmuştur ( $p = 0.001 < 0.05$ ). Model için belirtme katsayısı,  $R^2 = 0.40$  olarak hesaplanmıştır. Yani modelde bulunan açıklayıcı değişkenin modelin varyansını açıklama oranı 0.40'dir. Modelin standart hatası 2.793 ve değişkenler arasındaki korelasyon ise 0.632 olarak hesaplanmıştır. Eğim parametresinin işaretinden de anlaşılacağı üzere değişkenler arasında aynı yönde bir ilişki olduğu söylenebilir.

Şimdi ise basit doğrusal regresyon analizinde model parametrelerinin çakı tahminlerini inceleyelim. Tablo 4.'de, Çakı yöntemi ile sırasıyla birer gözlem dışarıda bırakılarak uygulanan regresyon analizi sonucu elde edilen beta katsayıları, korelasyon katsayısı ve modelin standart hatası görülmektedir. Sırasıyla birer gözlem silindiğinde;

$$\hat{Y}_{-1} = \hat{\beta}_{0(-1)} + \hat{\beta}_{1(-1)}X = -4.07469 + 0.008641X; \hat{\rho}_{x,y(-1)} = 0.634311; \hat{\sigma}_{(-1)} = 2.821427$$

$$\hat{Y}_{-2} = \hat{\beta}_{0(-2)} + \hat{\beta}_{1(-2)}X = -3.13011 + 0.007711X; \hat{\rho}_{x,y(-2)} = 0.612631; \hat{\sigma}_{(-2)} = 2.856460$$

...

$$\hat{Y}_{-23} = \hat{\beta}_{0(-23)} + \hat{\beta}_{1(-23)}X = -2.64440 + 0.006902X; \hat{\rho}_{x,y(-23)} = 0.578346; \hat{\sigma}_{(-23)} = 2.735828$$

sonuçları elde edilir. Elde edilen bu değerler Tablo 5.'de özetlenmiştir.

**Tablo 5.** Parametrelerin Çakı tahmini ile elde edilen istatistikler

İşlemden Çıkarılan Değişken İkili	$\hat{\beta}_{0(-n)}$	$\hat{\beta}_{1(-n)}$	$\hat{\sigma}_{(-n)}$	$\hat{\rho}_{x,y(-n)}$	$V(\hat{\rho}_{x,y}(j))$
Hiç Biri	-3.30211	0.007868	2.792949	0.632455	0.0300
1.çıkart	-4.07469	0.008641	2.821427	0.634311	0.029882
2.çıkart	-3.13011	0.007711	2.856460	0.612631	0.031234
3.çıkart	-3.11725	0.0077	2.855518	0.612415	0.031247
4.çıkart	-3.34335	0.007905	2.861528	0.624565	0.030496
5.çıkart	-3.11085	0.007702	2.852374	0.617214	0.030952
6.çıkart	-3.34961	0.007909	2.861290	0.626897	0.03035
7.çıkart	-3.59128	0.00811	2.833915	0.641613	0.029417
8.çıkart	-3.12587	0.007731	2.847703	0.623734	0.030548
9.çıkart	-3.04780	0.007673	2.830257	0.623553	0.030559
10.çıkart	-3.90872	0.008326	2.660232	0.677808	0.027029
11.çıkart	-3.29950	0.007867	2.861912	0.631763	0.030044
12.çıkart	-3.21968	0.007837	2.848467	0.632497	0.029997
13.çıkart	-3.30456	0.007931	2.850370	0.635995	0.029776
14.çıkart	-3.40671	0.00812	2.806581	0.648383	0.02898
15.çıkart	-3.38546	0.008038	2.844900	0.637914	0.029653
16.çıkart	-3.63125	0.008486	2.678568	0.679298	0.026928
17.çıkart	-3.58151	0.008361	2.770761	0.659795	0.028234
18.çıkart	-3.16001	0.007619	2.839065	0.615364	0.031066
19.çıkart	-3.81003	0.008722	2.633297	0.692309	0.026035
20.çıkart	-2.65802	0.006787	2.484954	0.620419	0.030754
21.çıkart	-2.81737	0.007081	2.698943	0.603382	0.031797
22.çıkart	-3.27178	0.007823	2.861587	0.612751	0.031227
23.çıkart	-2.64440	0.006902	2.735828	0.578346	0.033276
<b>Ortalama</b>	-3.30391	0.00787	2.791128	0.632302	0.029977
<b>Std. Hata</b>	0.074667	0.000102	0.020476	0.005415	0.036102

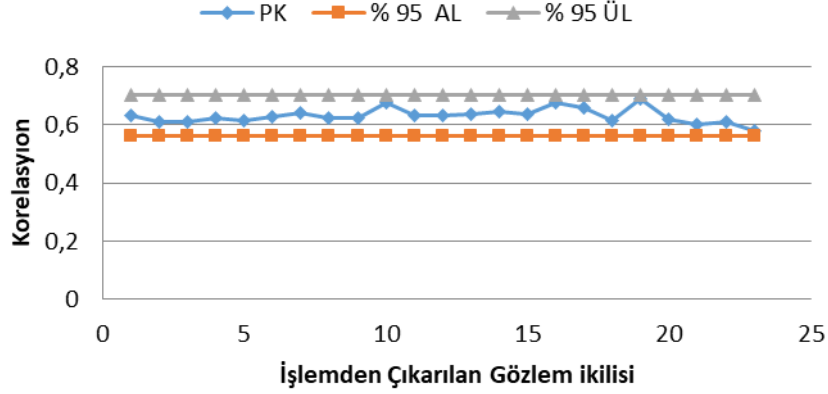
Parametrelerin EKK ve Çakı tahminlerine ait özet bilgiler Tablo 6.'deki gibi verilmiştir.

Tablo 6. incelendiğinde hem regresyon katsayılarının hem de korelasyon katsayısının Çakı tahmin edicisinin küçük yanlılığa sahip olduğu görülmektedir. Ayrıca parametrelerin çakı tahmin edicileri, EKK tahmin edicilerine göre daha küçük standart hatalara sahiptir.

**Tablo 6.** Parametrelerin EKK ve Çakı Tahminleri Arasındaki Karşılaştırma

Tahminler	EKK	Çakı	Yan
$SE(\hat{\beta}_0)$	1.906	0.074667	
$\hat{\beta}_1$	0.008	0.00787	0.00013
$SE(\hat{\beta}_1)$	0.002	0.000102	-
$\rho_{x,y}$	0.632455	0.632302	-0.0003
$SE(\rho_{x,y})$	-	0.036102	-

## Uranyum Verisi Çalışması



Şekil 2. Pearson Korelasyonu Çakı Tahminlerinin İzleme Sinyali Grafiği

Şekil 2.'de de görüldüğü gibi Pearson korelasyonun çakı tahmini 0.6323 olarak hesaplanmış ve % 95 simetrik güven aralığı alt limiti, 0.5615 ve üst limiti ise 0.7031 olarak tahmin edilmiştir. Yine aynı izleme sinyali grafiğinden de anlaşıldığı gibi hiçbir gözlem çifti % 95'lik güven sınırları dışına çıkmamıştır.

Tablo 7. Regresyon Modelinden Elde Edilen Artık Değerler

Sıra No	$e_i$	$e_i/23$
1	1.89283	0.082297
2	-0.74732	-0.03249
3	-0.80959	-0.0352
4	0.20236	0.008798
5	-0.99942	-0.04345
6	0.25766	0.011203
7	1.71938	0.074756
8	-1.23352	-0.05363
9	-1.83969	-0.07999
10	4.57808	0.199047
11	-0.0311	-0.00135
12	-1.21164	-0.05268
13	-1.12065	-0.04872
14	-2.42825	-0.10558
15	-1.34577	-0.05851
16	-4.3372	-0.18857
17	-3.07091	-0.13352
18	1.54658	0.067243
19	-4.78178	-0.2079
20	6.05648	0.263325
21	4.04362	0.17581
22	0.18262	0.00794
23	3.47722	0.151183

Şimdi ise basit doğrusal regresyon analizinde parametrelerinin bootstrap tahminlerini inceleyelim. Bootstrap yönteminin işleyiş algoritmasına göre; Tablo 7. incelenecek olursa, öncelikle  $e_i$  artık değerleri klasik EKK yöntemi ile hesaplanmıştır. Daha sonra elde edilen her bir  $e_i$  değerine  $1/23$  olasılığı verilmiştir. Elde edilen 23 tane  $e_i/23$  değerine bootstrap yöntemi uygulanmıştır. Burada bootstrap deneme sayısı 200 alınmış ve bu işlem 20 kez tekrar edilmiştir. Her tekrar sonucunda elde edilen bootstrap hata



terimleri ortalaması ( $\bar{\varepsilon}_i^*$ ) ve bootstrap bağımlı değişkeni ( $Y_i^*$ ) aşağıdaki gibi özetlenmiştir.

**Tablo 8.** Bootstrap artık değerleri ve hesaplanan tahmin değerleri

Sıra No	Birinci denemede ki Bootstrap sonuçları		İkinci denemede ki Bootstrap sonuçları		...	20'inci denemede ki Bootstrap sonuçları	
	$\bar{\varepsilon}_i^*$	$Y_i^*$	$\bar{\varepsilon}_i^*$	$Y_i^*$		$\bar{\varepsilon}_i^*$	$Y_i^*$
1	0.03053631	-0.89244685	-0.0116242	-0.93460732	...	-0.012734	-0.93571712
2	-0.02114253	0.85591535	0.01965676	0.89671464	...	0.0062673	0.88332518
3	-0.00175739	0.89756693	0.00366615	0.90299047	...	-0.0071079	0.89221638
4	-0.01225586	1.19510066	0.00266208	1.2100186	...	0.02254687	1.22990339
5	-0.03392375	1.33519885	0.00232354	1.37144614	...	0.00196126	1.37108386
6	0.00955913	1.47160281	0.01156107	1.47360475	...	-0.0042297	1.45781398
7	0.0052937	1.6855957	-0.0002646	1.68003737	...	-0.0032337	1.67706834
8	-0.00950879	2.02367301	-0.0028409	2.03034088	...	-0.0015451	2.03163671
9	0.0027447	2.10209638	-0.003285	2.09606673	...	-0.0170042	2.08234749
10	0.00129074	2.18286302	-0.0133136	2.16825868	...	-0.0060796	2.17549268
11	0.00455374	2.90526122	0.02811097	2.92881845	...	0.00297909	2.90368657
12	-0.02037152	3.12086072	0.00384496	3.1450772	...	-0.0131816	3.12805068
13	0.02686037	4.24703145	-0.0189276	4.2012435	...	0.00249864	4.22266972
14	-0.00206329	4.83566711	-0.0071653	4.83056513	...	-0.0166604	4.82107
15	-0.00804832	5.10718644	-0.0072518	5.10798293	...	-0.0022212	5.11301355
16	0.01766489	5.33432045	-0.0074991	5.30915646	...	0.00311925	5.31977481
17	0.00793344	5.53828388	-0.0001933	5.53015718	...	0.01824789	5.54859833
18	-0.02188076	5.5209798	-0.0144533	5.52840726	...	-0.0152284	5.52763221
19	-0.00201365	5.73919919	-0.0030446	5.73816825	...	0.0086785	5.74989134
20	0.01373441	5.75667821	0.00606853	5.74901233	...	-0.003313	5.73963084
21	-0.00128466	5.92451146	0.00287994	5.92867606	...	-0.0211991	5.90459705
22	0.02587154	6.54263314	0.00312333	6.51988493	...	-0.0025408	6.51422078
23	0.01072546	6.80286706	0.01205299	6.80419459	...	-0.0016419	6.79049971

Tablo 8.'de, 20 defa 200'erlik bootstrap denemelerine ait elde edilen bootstrap hata terimleri ortalaması ( $\bar{\varepsilon}_i^*$ ) ve bootstrap bağımlı değişkeni ( $Y_i^*$ ) görülmektedir. Burada bootstrap hata terimleri ortalaması eşitlik (2.8) yardımıyla hesaplanmıştır. Örneğin,

1. denemede;

1. Bootstrap artık değer ortalaması (0.03053631) 200 bootstrap örneğindeki 1. değerlerin ortalaması alınarak hesaplanmıştır. 23. Bootstrap artık değer ortalaması (0.01072546) ise elde edilen 200 bootstrap örneğindeki 23. değerlerin ortalaması alınarak hesaplanmıştır.

Benzer şekilde 20. denemede,

1. Bootstrap artık değer ortalaması (-0.012734) 200 bootstrap örneğindeki 1. değerlerin ortalaması alınarak hesaplanmıştır. 23. Bootstrap artık değer ortalaması (-0.0016419) ise elde edilen 200 bootstrap örneğindeki 23. değerlerin ortalaması alınarak hesaplanmıştır.

Elde edilen 23 boyutlu 20 tane bootstrap artık değer ortalamaları ( $\bar{\varepsilon}_i^*$ ) kullanılarak bootstrap yöntemine ait bağımlı değişken aşağıdaki şekilde hesaplanır.

$Y_i^*$ : EKK' dan edilen regresyon modeli +  $\bar{\varepsilon}_i^*$  şeklinde ifade edilir. Yani,

$$Y_i^* = -3.302 + 0.008x_i + \bar{\varepsilon}_i^*$$

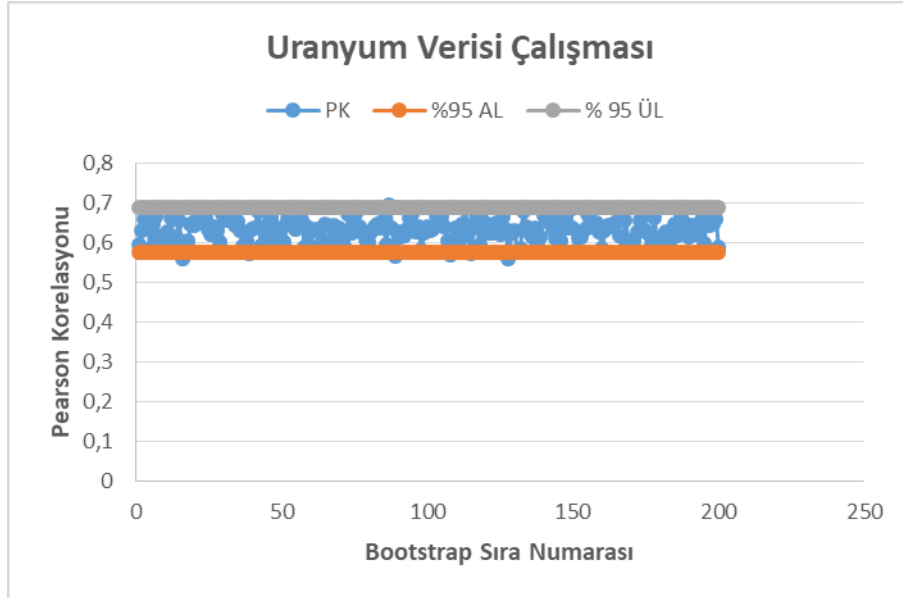
'dir. Burada elde edilen  $Y_i^*$  değeri hata teriminin yeniden örneklemesine dayanan bootstrap tahminidir. Elde edilen bootstrap bağımlı değişken ( $Y_i^*$ ) değeri ile açıklayıcı değişken yardımıyla aşağıdaki regresyon sonuçları elde edilmiştir.

EKK ve Boot EKK tahminlerine ait özet bilgiler Tablo 9.'daki gibi verilmiştir.

**Tablo 9.** Parametrelerin EKK ve Bootstrap Tahminleri Arasındaki Karşılaştırma.

Tahminler	EKK	Bootstrap	Yan
$SE(\hat{\beta}_0)$	1.906	0.002	
$\hat{\beta}_1$	0.008	0.00786	0.00014
$SE(\hat{\beta}_1)$	0.002	0.000002	-
$\rho_{x,y}$	0.63245	0.63235	-0.0001
$SE(\rho_{x,y})$		0.029073	

Tablo 9. incelendiğinde hem regresyon katsayılarının hem de korelasyon katsayısının bootstrap tahmin edicisinin EKK tahmin edicisine göre daha küçük yanlılığa sahip olduğu görülmektedir. Ayrıca bootstrap tahmin edicileri, EKK tahmin edicilerine göre daha küçük standart hatalara sahiptir. Aşağıda pearson korelasyonunun bootstrap tahminlerine ait izleme sinyali grafiği verilmiştir.



**Şekil 3.** Pearson Korelasyonun Bootstrap Tahminlerinin İzleme Sinyali Grafiği.

Şekil 3.'de de görüldüğü gibi Pearson korelasyonun bootstrap tahmini 0.63235 olarak hesaplanmış ve % 95 simetrik güven aralığı alt limiti, 0.5680 ve üst limiti 0.7111 olarak tahmin edilmiştir.

Bu aşamada, hesaplanan pearson korelasyon katsayısının önemliliğini 3 yönteme göre inceleyelim.

Korelasyon Katsayısının Önemliliği için Test İstatistiği,

Hipotez:

$$H_0: \rho_{x,y} = 0$$

$$H_1: \rho_{x,y} \neq 0$$

Test İstatistiği:

$$t_{Klasik} = \frac{\hat{\rho}_{x,y} * \sqrt{n-2}}{\sqrt{1 - \hat{\rho}_{x,y}^2}} = \frac{0.632455 * \sqrt{21}}{\sqrt{1 - (0.632455)^2}} = 3.74165$$

$$t_{\text{Çakı}} = \frac{\hat{\rho}_{x,y}(\zeta)}{\sqrt{\text{var}(\hat{\rho}_{x,y}(\zeta))}} = \frac{0.632302}{0.036102} = 17.5143$$

$$t_{\text{Bootstrap}} = \frac{\hat{\rho}_{x,y}(b)}{\sqrt{\text{var}(\hat{\rho}_{x,y}(b))}} = \frac{0.63235}{0.029073} = 21.7504$$

Karar:  $\alpha = 0.05$  önem seviyesinde kritik değer ise,

$t_{n-2, \frac{\alpha}{2}} = t_{21, 0.025} = 2.080$  olup,  $|t_{Klasik}| > t_{n-2, \frac{\alpha}{2}}$ ,  $|t_{\text{Çakı}}| > t_{n-2, \frac{\alpha}{2}}$  ve  $|t_{\text{Bootstrap}}| > t_{n-2, \frac{\alpha}{2}}$   $H_0$  hipotezi reddedilir.

Klasik ve çakı tabanlı test istatistik değerleri önerilen bootstrap tabanlı test istatistik değerinden daha küçüktür. Bu, EKK ve çakı tahminlerinin varyanslarının daha büyük olmasının bir sonucudur. Dolayısıyla önerilen pearson korelasyonunun bootstrap tahmini ile elde edilen test istatistiği, çakı ve klasik tabanlı test istatistik değerlerine nazaran daha etkili olduğu söylenebilir.

#### 4. Sonuç ve Yorum

Bu çalışmada, bootstrap yönteminin doğrusal regresyon katsayılarının ve korelasyon katsayısının tahmininde EKK ve çakı yöntemlerinden daha etkili olduğu gözlenmiştir. Korelasyon katsayısının önemini test etmek için kullanılan klasik test istatistiği, EKK tahmin edicilerinin büyük standart hatalarının etkisi ve dolayısıyla yokluk hipotezinin yanlışlıkla kabul edilmesine neden olabilir (2.Tip hata). Bu bağlamda, çalışma kapsamında korelasyon katsayısının önemini test etmek için önerilen bootstrap tabanlı test istatistiğinin bu yanlışlığın giderilmesi hususunda fayda sağlayacağı açıkça görülmüştür. Sonraki çalışmalarda bootstrap yönteminin diğer korelasyon katsayılarına ilavesi ile elde edilecek olan korelasyon katsayılarının karşılaştırılma aşamaları incelenecektir.

#### Kaynakça

- [1] M.R. Chernick, *Bootstrap Methods a Guide for Practitioners and Researchers*. 2nd ed; John Wiley & Sons Inc, New Jersey, 2008
- [2] B. Efron, "Bootstrap Method; Another Look at Jackknife," *Annals of Statistics*, 7, 1-26, 1979.
- [3] D.A. Freedman, "Bootstrapping Regression Models," *Annals of Statistics*, 1 (6), 1218-1228, 1981.
- [4] C. F. J. Wu, "Jackknife, Bootstrap and other Resampling Methods in Regression Analysis," *Annals of Statistics*, 14 (4), 1261-1290, 1986.
- [5] Z. Y. Algama, K. B. Rasheed, "Re-sampling in Linear Regression Model Using Jackknife and Bootstrap," *Iraqi Journal of Statistical Science*, 18, 59-73, 2010.
- [6] A. Akpanta, I. Okorie, "Investigating the Significance of a Correlation Coefficient using Jackknife Estimates," *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, 22 (2), 2015.
- [7] M. H. Quenouille, "Approximate tests of correlation in time series," *Journal of the Royal Statistical Society*, 11, 18-44, 1949.
- [8] M. H. Quenouille, "Notes on bias in estimation," *Biometrika*, 61, 353-360, 1956.
- [9] J. W. Tukey, "Bias and confidence in not-quite large sample," *Ann. Math. Stat.*, 29, 614, 1958.
- [10] S. Lohr, *Sampling: design and analysis*. Nelson Education, 2009.

- [11] H. Friedl, E. Stampfer, "Jackknife Resampling", *Encyclopaedia of Econometrics*. 2, 1089-1098, 2002.
- [12] S. Sahinler, D. Topuz, "Bootstrap and Jackknife Resampling Algorithms for Estimation of Regression Parameters," *Journal of Applied Quantitative Methods*, 2 (2), 188-199, 2007.
- [13] H. Abdi, J.L Williams, *Jackknife*. In Neil Salkind (Ed.), *Encyclopaedia of Research Design*. Thousand Oaks, CA: Sage, 2010.
- [14] B. Efron, "Bootstrap Method; another Look at Jackknife," *Annals of Statistics*, 7 (1), 26, 1979.
- [15] M. R. Chernick, *Bootstrap methods: A Guide for Practitioners and Researchers*. John Wiley & Sons, 619, 2008
- [16] N. Hamajima, "Methods for Statistical Inferences," *Biotherapy-Tokyo*, 13, 739-744, 1999.
- [17] R. Stine, *Modern Methods of Data Analysis*, by John Fox, Scotland: 325- 373, 1990.
- [18] N. Barker, "A Practical Introduction to The Bootstrap Using The Sas System". In SAS Conference Proceedings: Phuse 2005: Heidelberg, Germany SAS
- [19] D. Okutan, *Bootstrap Yönteminin Regresyon Analizinde Kullanımı ve Diğer Yöntemlerle Karşılaştırılması*. Ege Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, İzmir, 2009.
- [20] J. Fox, *Applied Regression Analysis, Linear Models and Related Methods*. Sage Publications Inc., 1997.
- [21] D. Topuz, *Regresyonda Yeniden Örnekleme Yöntemlerinin Karşılaştırmalı Olarak İncelenmesi*. Yüksek Lisans Tezi, Niğde, 2002.
- [22] Spearman's Correlation. Available: <http://www.statstutor.ac.uk/resources/uploaded/spearman.pdf>.