Nevşehir Bilim ve Teknoloji Dergisi Cilt 5(2) 74-85 2016
DOI: 10.17100/nevbiltek.284732
URL: http://dx.doi.org/10.17100/nevbiltek.284732

Bootstrap-t ve Yüzdelik Bootstrap Yöntemlerinde Tekrar Sayısı, Budama Yüzdesi ve Dağılımın Sonuçlara Etkisi

A. Fırat Özdemir [1], Gözde Navruz [1,*]

[1]*Dokuz Eylül Üniversitesi, Fen Fakültesi, İstatistik Bölümü, İzmir*

Öz

Bootstrap yöntemler, kestiricinin veya test istatistiğinin dağılımının bilinmediği durumlarda çıkarsama yapılmasını sağlayan ve eldeki rassal örneklemden tekrarlı olarak yapılan seçimler ile yeni örneklemler türetme ilkesine dayanan yöntemlerdir. Bootstrap yöntemlerde; tekrar sayısı, budanmış ortalama içeren bir yöntem ile birlikte kullanılırlarsa budama yüzdesi ve kitle dağılımının yöntemin performansını nasıl etkilediği tartışılmakta olan konulardır [1-7]. Bu çalışmada; tek örneklem hipotez testi yapmak amacıyla Tukey-McLaughlin testinin [8] bootstrap-t ile birlikte kullanımı ve yüzdelik bootstrap, iki örneklem hipotez testi yapmak amacıyla ise Yuen testinin [9] bootstrap-t ile birlikte kullanımı ve yüzdelik bootstrap yöntemi kullanılmıştır. Bahsedilen yöntemlerin performansları; farklı tekrar sayıları, budama yüzdeleri ve kitle dağılımları kullanılarak gerçekleşen 1. Tip hata değerlerine göre karşılaştırılmıştır. Karşılaştırma bir simülasyon çalışması ve ayrıca iki gerçek veri seti ile yapılmıştır. Kitle budanmış ortalaması için tek ve iki örneklemde hipotez testi yöntemi, budama yüzdesi ve tekrar sayısı için öneriler geliştirilmiştir.

Anahtar Kelimeler: Bootstrap, Tukey-McLaughlin testi, Yuen testi

The Effect of Number of Bootstrap Samples, Trimming Proportion and Distribution to the Results in Bootstrap-t and Percentile Bootstrap Methods

Abstract

Bootstrap methods are procedures which enable to make inference when the distribution of estimator or test statistics is unknown, and based on the principle of generating new samples by using the original random sample with replacement. In bootstrap methods, how number of bootstrap samples, trimming proportion if they are used with a method that involves trimmed mean and population distribution affect the performance of the methods are issues that have been discussed [1-7]. In this study; with the aim of performing one sample hypothesis testing use of Tukey-McLaughlin test [8] with bootstrap-t and percentile bootstrap, and with the aim of performing two samples hypothesis testing use of Yuen test [9] with bootstrap-t and percentile bootstrap are used. The performances of these methods are compared in terms of actual type 1 error rates by using different number of bootstrap samples, trimming proportions and population distributions. The comparison is done with a simulation study by using theoretical distributions and two real data sets. Suggestions for the method to be used, trimming proportion and number of bootstrap samples are developed.

Keywords: Bootstrap, Tukey-McLaughlin test, Yuen test

[*]e-mail: gnavruz@gmail.com

1. Giriş

Bootstrap yöntemler ilk olarak Efron'un 1979 yılındaki çalışmasında Quenouille-Tukey jackknife yöntemine alternatif olarak önerilmiş yeniden örnekleme yöntemleridir [10]. Örneklemin çekildiği kitlenin dağılımının bilinmediği durumlarda, kestiricinin veya test istatistiğinin örnekleme dağılımı, elde edilmiş olan örnekleme dayanılarak bulunur. Yani mevcut örneklem kitle gibi düşünülerek bu örneklemden yerine koyarak yapılacak çekilişlerle yeni örneklemler türetilir. Bootstrap yöntemler ile kitlenin dağılımının bilinmesine veya kitle dağılımı ile ilgili herhangi bir varsayım yapılmasına gerek olmadan çıkarsama yapmak mümkündür. Efron 1979'daki çalışmasında bir istatistiğe ait bootstrap dağılımın bulunmasının; doğrudan teorik hesaplama, Monte Carlo yaklaşımı ya da Taylor seri açılımı olmak üzere üç farklı yolla mümkün olduğunu belirtmiştir [10].

Bootstrap yöntemlerin kullanımı, yoğun bilgisayar hesaplamalarına ihtiyaç duyulması nedeniyle, 1990'lı yıllardan itibaren bilgisayarların gelişimi ile birlikte artmıştır [4]. Sonraki yıllarda bootstrap kullanılarak standart hata bulunması, yanlılık gibi performans kriterlerinin hesaplanması, güven aralığı hesaplamaları, korelasyon katsayısı hesaplaması, regresyon analizi ve zaman serisi analizi gibi pek çok konuda uygulamalar yapılmıştır [2].

Bootstrap yöntemler hem parametrik hem de parametrik olmayan yeniden örnekleme ile kullanılabilmektedir. Parametrik olmayan bootstrap yöntemlerde örneklemin çekildiği kitlenin dağılımı ile ilgili herhangi bir varsayım yapılmazken, parametrik bootstrap yönteminde kitle dağılımı bilinmektedir. Yani parametrik bootstrap kullanılan yöntemlerde, bootstrap örneklemleri dağılımı bilinen kitleden çekilir. Parametrik olmayan bootstrapta ise yöntemin işleyişi şu şekilde özetlenebilir: $\{X_1, X_2, ..., X_n\}$ dağılımı bilinmeyen bir kitleden gelen, n genişliğinde rassal bir örneklem olsun. Bu örneklemden yerine koyarak n genişliğinde B tane rassal örneklem çekilir. θ parametresinin tahmin edilmesiyle ilgileniliyorsa, elde edilen B tane bootstrap örneklemin her birinden $\hat{\theta}^*$ istatistiği hesaplanır. $\hat{\theta}_1^*, ..., \hat{\theta}_B^*$ değerlerinin ortalamasının ve standart hatasının bulunmasıyla örnekleme dağılımı kestirilmiş olur.

$$\hat{\theta}_{(.)}^* = \frac{\sum_{b=1}^{B} \hat{\theta}_b^*}{B} \;,\; \hat{se}_{\theta}^* = \left(\frac{\sum_{b=1}^{B} (\hat{\theta}_b^* - \hat{\theta}_{(.)}^*)^2}{B-1} \right)^{1/2} ,[3] \tag{1}$$

Bootstrap güven aralıklarının hesaplanmasında literatürde pek çok yöntem tanıtılmıştır. Bütün bootstrap yöntemlerin ardında yatan temel fikir örnekleme dağılımını kestirmek için elde edilmiş örneklemi kullanmak, güven aralıklarını elde etmek ve hipotez testi yapmaktır [6]. En yaygın olarak karşılaşılan bootstrap güven aralıklarına örnek olarak normal yaklaşım ile bootstrap güven aralığı, yüzdelik bootstrap güven aralığı, bootstrap-t güven aralığı, yanlılığı düzeltilmiş bootstrap güven aralığı, yanlılığı düzeltilmiş ve hızlandırılmış bootstrap güven aralığı verilebilir [1]. Bu çalışmada yüzdelik bootstrap ve bootstrap-t yöntemlerine detaylı olarak yer verilmiştir.

Bootstrap yöntemlerde tekrar sayısı (B), örneklemin çekildiği kitlenin dağılımı, kullanılan bootstrap güven aralığı yöntemi ve kullanılan kestiriciye bağlı olarak sonuçların nasıl etkilendiği tartışılmakta olan konulardır [1-7]. Budanmış ortalama içeren bir yöntem kullanıldığında budama yüzdesinin (γ) etkisi de araştırılmaktadır. Bu çalışmada bootstrap-t ve yüzdelik bootstrap yöntemlerinin performansları; gerçekleşen 1. Tip hata ölçütüne göre, farklı tekrar sayıları (B), budama yüzdeleri (γ) ve kitle dağılımları kullanılarak karşılaştırılmıştır. Karşılaştırma bir simülasyon çalışması ile yapılmış, budama yüzdesi ve tekrar sayısı için öneriler sunulmuştur.

2. Bootstrap Yöntemler

$\{X_1,...,X_n\}$ olasılık dağılımı bilinmeyen bir kitleden elde edilmiş n genişliğinde rassal bir örneklem olsun. Ayrıca ilgilenilen kitle parametresi θ ve bu parametrenin örneklemden elde edilmiş kestirimi $\hat{\theta}$ olduğunda, θ parametresine ait güven aralığı bootstrap örneklemlerinden yararlanılarak bulunur. $\{X_1,...,X_n\}$ örnekleminden yerine koyarak çekilen ve $\{X_1^*,...,X_n^*\}$ şeklinde gösterilen rassal örneklemler bootstrap örneklemleridir. Bootstrap güven aralıklarının elde edilmesinde en sık kullanılan yöntemler bootstrap-t ve yüzdelik bootstrap yöntemleridir.

2.1. Bootstrap-t Yöntemi

Bootstrap-t yöntemi aynı zamanda yüzdelik-t yöntemi olarak da adlandırılmaktadır. θ parametresinin bootstrap örneklemlerden elde edilen kestirimleri $\hat{\theta}_i^*$, i=1,…,B ile gösterildiğinde, bu yöntemi uygulamak için her $\hat{\theta}_i^*$ değerinin standart hatasının bilinmesi gereklidir. Bootstrap-t yöntemi, kestiricinin standart hatası bilindiğinde uygulaması kolay bir yöntemdir. Ancak kestiricinin standart hatası bilinmediğinde veya hesaplaması zahmetli olduğunda her bir bootstrap örneklem için yeniden bootstrap uygulanması ve böylece standart hatanın tahmin edilmesi gerekmektedir. Double (çift) bootstrap olarak adlandırılan bu yöntem zaman alıcı olabilmektedir. Bootstrap-t yönteminin adımları aşağıdaki gibidir:

1) Her bir bootstrap örnekleminden t^* değeri hesaplanır.

$$t_b^* = \frac{\hat{\theta}_b^* - \hat{\theta}}{\hat{se}_b^*} \quad ,b=1,\dots,B \tag{2}$$

$\hat{se}_b^*$, $\hat{\theta}$ ’nın standart hatasının kestirimini göstermektedir.

2) Elde edilen B tane t^* değeri küçükten büyüğe sıralanır. $t_{(1)}^* \le ... \le t_{(B)}^*$

3) $\ell = \dfrac{\alpha B}{2}$ ve $u = B - \ell$ değerleri hesaplanır.

4) θ parametresi için $\%100(1-\alpha)$ güven düzeyinde güven aralığı, u. sıradaki ve ℓ. sıradaki t^* değerlerinin kullanılmasıyla $(\hat{\theta} - t_{(u)}^*\hat{se}_b^*, \hat{\theta} - t_{(l)}^*\hat{se}_b^*)$ şeklinde elde edilir.

2.2. Yüzdelik Bootstrap Yöntemi

Bootstrap-t yönteminin aksine, yüzdelik bootstrap yöntemi ile güven aralığı elde etmek için $\hat{\theta}_i^*$ değerlerinin standart hatasına ihtiyaç yoktur. Güven aralığının alt ve üst limitleri $\hat{\theta}$ ’nın bootstrap dağılımı ile belirlenir. Yöntemin uygulanışı aşağıdaki gibidir:

1) Her bir bootstrap örnekleminden $\hat{\theta}_b^*$ değeri hesaplanır (b=1,…,B).

2) Elde edilen B tane $\hat{\theta}^*$ değeri küçükten büyüğe sıralanır. $\hat{\theta}_{(1)}^* \le ... \le \hat{\theta}_{(B)}^*$

3) $\ell = \frac{\alpha B}{2}$ ve $u = B - \ell$ değerleri hesaplanır.

4) θ parametresi için $\%100(1-\alpha)$ güven düzeyinde güven aralığı, (ℓ+1). sıradaki ve u. sıradaki $\hat{\theta}^*$ sayesinde $(\hat{\theta}^*_{(l+1)}, \hat{\theta}^*_{(u)})$ şeklinde bulunur.

2.3. Tek Örneklem için Bootstrap Yöntemler

2.3.1. Tukey-McLaughlin Testi

Kitle ortalaması μ için güven aralığı hesaplarken veya hipotez testi yaparken $T = \frac{\sqrt{n}(\bar{x}-\mu)}{s}$ test istatistiğinin n-1 serbestlik derecesi ile Student's-*t* dağılımına sahip olduğu varsayılır. Bu varsayım, örneklem normal dağılıma uyan bir kitleden çekildiğinde sağlansa da, kitle çarpık olduğunda ve uç değerler içerdiğinde Tukey-McLaughlin testinin kullanılması önerilmektedir [8].

Tukey ve McLaughlin'e göre $T_t = \frac{\bar{X}_t - \mu_0}{s_w / (1-2\gamma)\sqrt{n}}$ test istatistiği n-2g-1 serbestlik derecesi ile Student's-*t* dağılıma uymaktadır. Burada $H_0 : \mu_t = \mu_0$ hipotezi test edilmektedir ve kitle budanmış ortalaması μ_t 'nin örneklemden elde edilmiş kestirimi $\bar{X}_t$ ile belirtilmiştir.

Örneklem budanmış ortalaması $\bar{X}_t$ 'nin hesaplanması için öncelikle budama yüzdesi γ belirlenir ($0 \le \gamma \le 0.5$). $\{X_1,...,X_n\}$ rassal örneklemi $X_{(1)} \le ... \le X_{(n)}$ şeklinde küçükten büyüğe sıralanır. $g = [\gamma n]$ değerinin hesaplanmasında $[\gamma n]$, γn çarpımından küçük en yakın tam sayıya yuvarlanmasını ifade etmektedir. Buna göre örneklem budanmış ortalaması,

$$\bar{X}_t = \frac{X_{(g+1)} + ... + X_{(n-g)}}{n-2g} \tag{3}$$

şeklinde elde edilir [7]. Ayrıca s_w, örneklem Winsorized standart sapmasını göstermektedir ve bulunuşu şu şekildedir:

$W_i = \begin{cases} X_{(g+1)}, \text{if } X_i \le X_{(g+1)} \\ X_i, \quad \text{if } X_{(g+1)} < X_i < X_{(n-g)} \\ X_{(n-g)}, \text{if } X_i \ge X_{(n-g)} \end{cases}$ olmak üzere örneklem Winsorized ortalaması $\bar{X}_w = \frac{1}{n}\sum W_i$ bulunur. $s_w^2 = \frac{1}{n-1}\sum (W_i - \bar{X}_w)^2$ formülü ile örneklem Winsorized varyansı elde edilir.

Buna göre, kitle budanmış ortalaması μ_t için $\%100(1-\alpha)$ güven aralığı $\bar{X}_t \pm t_{1-\alpha/2} \frac{s_w}{(1-2\gamma)\sqrt{n}}$ şeklinde bulunur. Burada $t_{1-\alpha/2}$, Student's-*t* dağılımının n-2g-1 serbestlik dereceli $1-\alpha/2$ kantilini göstermektedir. $|T_t| > t_{1-\alpha}$ eşitsizliği sağlandığında H_0 hipotezi reddedilir [7].

2.3.2. Bootstrap-t ile Tukey-McLaughlin Testi

Bu çalışmada Tukey-McLaughlin testi bootstrap yöntemi ile birlikte kullanılmıştır. Uygulanan yöntemin adımları şu şekildedir:

1) $\{X_1,...,X_n\}$ rassal örneklemi kullanılarak örneklem budanmış ortalaması $\bar{X}_t$ hesaplanır.

2) Oluşturulan $\{X_1^*, X_2^*,...,X_n^*\}$ bootstrap örneklemlerinin her birinden $T_t^* = \dfrac{\left|\bar{X}_t^* - \bar{X}_t\right|}{s_w^* \big/ (1-2\gamma)\sqrt{n}}$ test istatistiği hesaplanır.

3) Bir önceki adımın B kez tekrarlanması ile elde edilen $T_{t1}^*,...,T_{tB}^*$ istatistikleri, $T_{t(1)}^* \le ... \le T_{t(B)}^*$ şeklinde küçükten büyüğe sıralanır.

4) $c = (1-\alpha)B$ değerinin hesaplanmasıyla, c. sıradaki $T_{t(c)}^*$ istatistiği bulunur.

5) T_t orijinal örneklemden elde edilen test istatistiği olmak üzere, $T_t < -T_{t(c)}^*$ veya $T_t > T_{t(c)}^*$ ise H_0 hipotezi reddedilir [7].

2.3.3. Yüzdelik Bootstrap ile Tek Örneklem Hipotez Testi

Test edilmesi amaçlanan hipotez yine $H_0 : \mu_t = \mu_0$ hipotezidir. Yöntemin adımları aşağıdaki gibidir:

1) $\{X_1,...,X_n\}$ rassal örnekleminden $\{X_1^*, X_2^*,...,X_n^*\}$ bootstrap örneklemi elde edilir ve elde edilen bootstrap örnekleminden $\bar{X}_t^*$ örneklem budanmış ortalaması hesaplanır.

2) İlk adımın B kez tekrarlanmasıyla bulunan $\bar{X}_{t1}^*, \bar{X}_{t2}^*,...,\bar{X}_{tB}^*$ değerleri, $\bar{X}_{t(1)}^* \le \bar{X}_{t(2)}^* \le ... \le \bar{X}_{t(B)}^*$ olacak şekilde küçükten büyüğe sıralanır.

3) $\ell = \dfrac{\alpha B}{2}$ ve $u = B - \ell$ değerlerinin bulunmasıyla (ℓ+1). ve u. sıralardaki $\bar{X}_t^*$ istatistikleri kullanılarak %100$(1-\alpha)$ güven aralığı $(\bar{X}_{t(l+1)}^*, \bar{X}_{t(u)}^*)$ şeklinde bulunur.

Hesaplanan güven aralığı μ_0 ‘ı içermiyorsa, H_0 hipotezi reddedilir [7].

2.4. İki Örneklem için Bootstrap Yöntemler

2.4.1. Yuen Testi

İki bağımsız grubun kitle budanmış ortalamalarının eşitliği hipotezi $H_0 : \mu_{t1} = \mu_{t2}$ ‘nin test edilmesi ve buna bağlı olarak $\mu_{t1} - \mu_{t2}$ farkına ait %100$(1-\alpha)$ güven aralığının bulunması amaçlanmaktadır. Yuen testi, budanmış ortalama ile kullanılan, normal dağılım ve homojen varyanslılık varsayımlarının bozulmasına karşı dayanıklı bir yöntemdir [9].

Yuen testinin uygulanması için öncelikle budama yüzdesi γ belirlenir ($0 \le \gamma \le 0.5$). $g_j = \left[\gamma n_j\right]$ ile j. grup (j=1,2) için her kuyruktan budanacak gözlem sayıları bulunur. $h_j = n_j - 2g_j$, j. grupta budama sonrası kalan gözlem sayısını belirtmektedir. s_{wj}^2 ‘nin j. grubun örneklem Winsorized varyansını gösterdiği d_j formülü ile örneklem budanmış ortalaması $\bar{X}_{tj}$ ‘nin varyansının tahmini gösterilmiştir.

$$d_j = \frac{(n_j - 1)s_{w_j}^2}{h_j(h_j - 1)} \qquad (4)$$

Yuen test istatistiği $T_y = \dfrac{\bar{X}_{t1} - \bar{X}_{t2}}{\sqrt{d_1 + d_2}}$ şeklindedir ve bu istatistik $\upsilon_j = \dfrac{(d_1 + d_2)^2}{\dfrac{d_1^2}{h_1 - 1} + \dfrac{d_2^2}{h_2 - 1}}$ serbestlik derecesi ile Student's-*t* dağılımına uyar. $|T_y| > t_{\upsilon_y}$ ise H_0 hipotezi reddedilir. Burada t_{υ_y} değeri, Student's-*t* dağılımının υ_y serbestlik dereceli $1-\alpha/2$ kantilidir. Ayrıca $\mu_{t1} - \mu_{t2}$ farkına ait $\%100(1-\alpha)$ güven aralığı $(\bar{X}_{t1} - \bar{X}_{t2}) \pm t_{\upsilon_y} \sqrt{d_1 + d_2}$ şeklindedir.

2.4.2. Bootstrap-t ile Yuen Testi

Bu çalışmada bootstrap-t yöntemi, iki bağımsız grubun kitle budanmış ortalamalarını karşılaştırmak amacıyla Yuen testi ile birlikte kullanılmıştır. Uygulanan yöntemin adımları şu şekildedir:

1) $\{X_{1j}, X_{2j}, ..., X_{nj}\}$ rassal örneklemleri kullanılarak örneklem budanmış ortalamaları $\bar{X}_{tj}$ hesaplanır (j=1,2).

2) Oluşturulan $\{X_{1j}^*, X_{2j}^*, ..., X_{nj}^*\}$ bootstrap örneklemlerinin her birinden $\bar{X}_{tj}^*$, d_j^* değerleri ve bu değerleri kullanarak $T_y^* = \dfrac{(\bar{X}_{t1}^* - \bar{X}_{t2}^*) - (\bar{X}_{t1} - \bar{X}_{t2})}{\sqrt{d_1^* + d_2^*}}$ test istatistiği hesaplanır.

3)Bir önceki adımın B kez tekrarlanması ile elde edilen $T_{y1}^*, ..., T_{yB}^*$ istatistikleri, $T_{y(1)}^* \le ... \le T_{y(B)}^*$ şeklinde küçükten büyüğe sıralanır.

4) $\ell = \dfrac{\alpha B}{2}$ ve $u = B - \ell$ değerlerinin bulunmasıyla (ℓ+1). ve u. sıralardaki $T_{y(u)}^*$ ve $T_{y(\ell+1)}^*$ istatistikleri bulunur.

5) $\mu_{t1} - \mu_{t2}$ için $\%100(1-\alpha)$ güven aralığı $(\bar{X}_{t1} - \bar{X}_{t2} - T_{y(u)}^* \sqrt{d_1 + d_2}, \bar{X}_{t1} - \bar{X}_{t2} - T_{y(l+1)}^* \sqrt{d_1 + d_2})$ şeklindedir. Bulunan güven aralığı 0'ı içermiyorsa H_0 hipotezi reddedilir [7].

2.4.3. Yüzdelik Bootstrap ile İki Örneklem Hipotez Testi

Test edilmesi amaçlanan hipotez yine iki bağımsız grubun kitle budanmış ortalamalarının eşitliği hipotezidir ($H_0 : \mu_{t1} = \mu_{t2}$). Yüzdelik bootstrap ile iki örneklem hipotez testinin adımları şu şekildedir:

1) $\{X_1, X_2, ..., X_n\}$ ve $\{Y_1, Y_2, ..., Y_n\}$ örneklemlerinden $\{X_1^*, X_2^*, ..., X_n^*\}$ ve $\{Y_1^*, Y_2^*, ..., Y_n^*\}$ bootstrap örneklemleri elde edilir.

2)Elde edilen bootstrap örneklemlerinden $\bar{X}_t^*$ ve $\bar{Y}_t^*$ örneklem budanmış ortalamalarının hesaplanmasıyla $D^* = \bar{X}_t^* - \bar{Y}_t^*$ farkı elde edilir.

3)İlk iki adımın B kez tekrarlanmasıyla bulunan $D_1^*, D_2^*, ..., D_B^*$ farkları, $D_{(1)}^* \le D_{(2)}^* \le ... \le D_{(B)}^*$ şeklinde küçükten büyüğe sıralanır.

4) $\ell = \dfrac{\alpha B}{2}$ ve $u = B - \ell$ değerlerinin bulunmasıyla (ℓ+1). ve u. sıralardaki D^* değerleri kullanılarak $\%100(1-\alpha)$ güven düzeyinde $\mu_{t1} - \mu_{t2}$ farkına ait güven aralığı $(D_{(l+1)}^*, D_{(u)}^*)$ şeklinde bulunur. Bu aralık 0'ı içermiyorsa H_0 hipotezi reddedilir [7].

3. Simülasyon Çalışması

Çalışmanın bu bölümünde bootstrap-t ve yüzdelik bootstrap yöntemlerinin performansları; farklı tekrar sayıları, budama yüzdeleri ve kitle dağılımları kullanarak gerçekleşen 1. Tip hata değerlerine göre karşılaştırılmıştır.

3.1. Simülasyon Çalışmasının Tasarımı

Bu çalışmada bootstrap-t ve yüzdelik bootstrap yöntemlerini karşılaştırmak amacıyla teorik dağılım olarak standart normal dağılım ve g&h dağılımı kullanılmıştır. g&h dağılımı, çarpıklığın kontrol edilebileceği g ve basıklığın kontrol edilebileceği h parametreleri ile normal dağılımdan uzaklaşmayı derecelendirerek gözleme imkanı veren ve g=h=0 için standart normal dağılıma eşdeğer bir dağılımdır. Z standart normal dağılıma sahip bir rassal değişken olmak üzere, verilen dönüşümlerle g&h dağılımından veri üretilir.

$$g \neq 0 \text{ için } X = \frac{\left(\exp(gZ)-1\right)\exp\left(\frac{hZ^2}{2}\right)}{g} \qquad (5)$$

$$g=0 \text{ için } X = Z\exp\left(\frac{hZ^2}{2}\right) \text{ [11].} \qquad (6)$$

Bu çalışmada g=0.5 ve h=0.2 parametrelerine sahip g&h dağılımından veri türetilmiştir. Ayrıca kullanılan g&h dağılımına ait çarpıklık ve basıklık değerleri Tablo 1'de gösterilmiştir.

Tablo 1. Kullanılan g&h dağılımına ait çarpıklık ve basıklık değerleri

g	h	$\hat{\kappa}_1$	$\hat{\kappa}_2$
0	0	0	3
0.5	0.2	10.554	200.736

$\gamma = 0.1$ ve $\gamma = 0.2$ olmak üzere iki farklı budama yüzdesi ve B=600, B=1000, B=2000 olmak üzere üç farklı bootstrap tekrar sayısı belirlenmiştir. Örneklem genişlikleri her durum için n=20'dir.

Yürütülen simülasyon çalışması ile hem tek örneklem hipotez testi hem de iki örneklem hipotez testi ile gerçekleşen anlam düzeyleri incelenmiştir. Tek örneklem için yüzdelik bootstrap yöntemi ve Bootstrap-t Tukey-McLaughlin Testi uygulanmıştır. İki örneklem için ise yine yüzdelik bootstrap yöntemi ve Bootstrap-t Yuen Testi uygulanmıştır.

Simülasyon çalışması 10000 tekrar yapılarak R yazılımı ile gerçekleştirilmiştir. Belirlenen anlam düzeyi $\alpha = 0.05$ 'dir.

3.2. Simülasyon Çalışmasının Sonuçları

Bu bölümde yapılan simülasyon çalışmasının sonucunda gerçekleşen 1. Tip hata değerleri verilmiştir. Sonuçları değerlendirmek amacıyla Bradley'in tutucu dayanıklılık kriteri esas alınmıştır [12]. Bradley'e göre, $\alpha = 0.05$ nominal anlam düzeyi olarak belirlendiğinde, gerçekleşen 1. Tip hata değerleri 0.045 ve 0.055 arasında kalıyorsa, bu durum belirlenen anlam düzeyinin korunduğunu göstermektedir. Sonuçların verildiği tablolarda (0.045, 0.055) aralığı dışında kalan gerçekleşen 1. Tip hata değerleri vurgulanmıştır.

Tablo 2'de tek örneklem için standart normal dağılım ile gerçekleşen anlam düzeyleri verilmiştir. Bootstrap-t yöntemi ile her iki budama yüzdesi için de tekrar sayısının 1000 ya da 2000 olması, 600 tekrara göre gerçekleşen anlam düzeylerinde önemli bir farklılık yaratmamıştır. Ayrıca her iki budama yüzdesi için de 600 tekrar ile sonuçların (0.045, 0.055) aralığında kaldığı görülmektedir. İşlem süresinin kısa olması nedeniyle 600 tekrar kullanılması önerilmektedir. Bu yöntem için $\gamma = 0.1$ budama yüzdesi ile gerçekleşen anlam düzeyleri nominal düzeye daha yakındır. Yüzdelik bootstrap için ise tekrar sayısının 2000 olması ile (0.045, 0.055) aralığına düşen sonuçlar elde edilmiştir. Ayrıca bu yöntem ile $\gamma = 0.2$ budama yüzdesinin kullanılması uygundur.

Tablo 2. Tek örneklem ve normal dağılım için gerçekleşen anlam düzeyleri

			B		
			600	**1000**	**2000**
Budama Yüzdesi	$\gamma = 0.1$	**Bootstrap-t**	0.0473	0.0473	0.0460
		Yüzdelik bootstrap	0.0575	0.0564	0.0557
	$\gamma = 0.2$	**Bootstrap-t**	0.0451	0.0443	0.0432
		Yüzdelik bootstrap	0.0572	0.0553	0.0548

Tablo 3'de iki örneklem için standart normal dağılım ile gerçekleşen anlam düzeyleri verilmiştir. Tüm sonuçlar Bradley'in [12] belirlediği (0.045,0.055) aralığının içindedir. İşlem süresinin daha kısa olması sebebiyle hem bootstrap-t yöntemi için, hem de yüzdelik bootstrap yöntemi için tekrar sayısının B=600 alınması önerilmektedir. Ayrıca bootstrap-t yöntemi ve yüzdelik bootstrap yöntemleri için sırasıyla $\gamma = 0.1$ ve $\gamma = 0.2$ budama yüzdelerinin kullanılması nominal düzeye daha yakın sonuçlar elde edilmesini sağlamıştır.

Tablo 3. İki örneklem ve normal dağılım için gerçekleşen anlam düzeyleri

			B		
			600	**1000**	**2000**
Budama Yüzdesi	$\gamma = 0.1$	**Bootstrap-t**	0.0513	0.0491	0.0510
		Yüzdelik bootstrap	0.0524	0.0523	0.0524
	$\gamma = 0.2$	**Bootstrap-t**	0.0509	0.0529	0.0529
		Yüzdelik bootstrap	0.0503	0.0502	0.0502

Tablo 4'de tek örneklem için g=0.5 ve h=0.2 dağılımı ile gerçekleşen anlam düzeyleri verilmiştir. Bootstrap-t yöntemi ile gerçekleşen anlam düzeyleri her durumda 0.045 ile 0.055 arasında kalmıştır. Bu yöntemin iki budama yüzdesi ile uygulanmasından alınan sonuçlar hemen hemen aynı olduğu için, her ikisinin de kullanılması uygundur. Tekrar sayısı olarak da işlem süresinin kısalığı ve nominal değer daha yakın sonuçlar getirmesi bakımından B=600 tercih edilebilir. Yüzdelik bootstrap yöntemi için ise, B=2000 tekrar sayısı ve $\gamma = 0.1$ budama yüzdesi tercih edilebilir.

Tablo 4. Tek örneklem ve g=0.5 ve h=0.2 dağılımı için gerçekleşen anlam düzeyleri

			B		
			600	**1000**	**2000**
Budama Yüzdesi	$\gamma = 0.1$	**Bootstrap-t**	0.0495	0.0488	0.0493
		Yüzdelik bootstrap	0.0545	0.0543	0.0540
	$\gamma = 0.2$	**Bootstrap-t**	0.0496	0.0487	0.0494
		Yüzdelik bootstrap	0.0582	0.0562	0.0539

Tablo 5'de iki örneklem için g=0.5 ve h=0.2 dağılımı ile gerçekleşen anlam düzeyleri verilmiştir. Bootstrap-t yöntemi için her iki budama yüzdesinde de tekrar sayısı gerçekleşen anlam düzeyinde önemli bir farklılık yaratmamış, dolayısıyla işlem süresinin kısalığı sebebiyle B=600 tercihi önerilmektedir. $\gamma = 0.2$ budama yüzdesi ile gerçekleşen anlam düzeylerinin nominal düzeye daha yakın olduğu görülmüştür. Yüzdelik bootstrap yönteminde ise, sonuçlar Bradley'in belirlediği aralığa düşmese de $\gamma = 0.2$ budama yüzdesi ve B=600 tekrar sayısı ile nominal düzeye daha yakın sonuçlar alınmıştır.

Tablo 5. İki örneklem ve g=0.5 ve h=0.2 dağılımı için gerçekleşen anlam düzeyleri

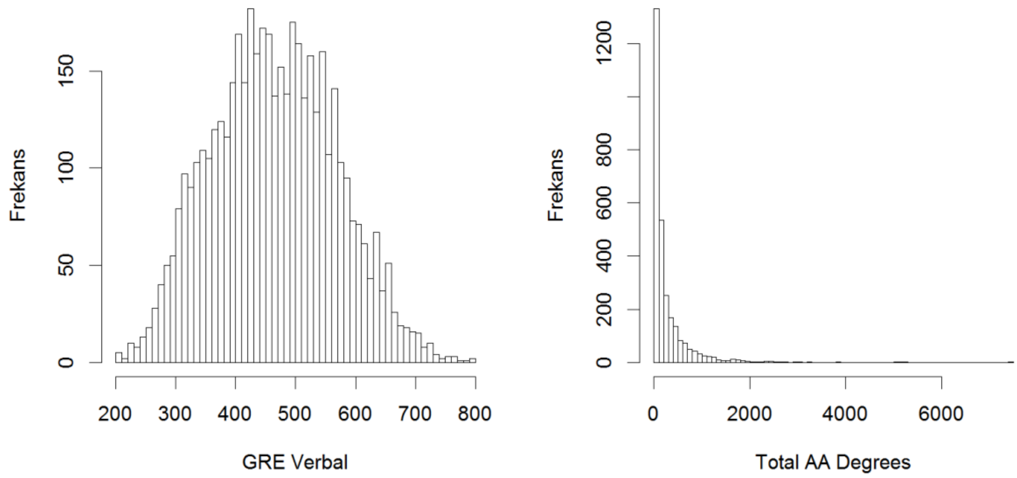
			B		
			600	**1000**	**2000**
Budama Yüzdesi	$\gamma = 0.1$	**Bootstrap-t**	0.0425	0.0418	0.0425
		Yüzdelik bootstrap	0.0437	0.0431	0.0436
	$\gamma = 0.2$	**Bootstrap-t**	0.0458	0.0443	0.0448
		Yüzdelik bootstrap	0.0443	0.0438	0.0440

3.3. Gerçek Veri Örneği

Çalışmanın bu bölümünde, tek örneklem ve iki örneklem hipotez testleri için kullanılan yöntemler iki farklı gerçek veri seti ile uygulanmıştır. Gerçek veriler http://www.freewebs.com/tedstats/ adresinden alınan farklı karakterdeki veri setlerdir[13]. Total AA Degrees ve GRE Verbal adlı veri setlerine ait tanımlayıcı istatistikler ve bu verilerin histogramları sırasıyla Tablo 6'da ve Şekil 1'de verilmiştir. Şekil 1'de de görüldüğü gibi GRE Verbal simetrik dağılıma sahip bir veri setiyken, Total AA Degrees sağdan çarpık bir dağılıma sahiptir. Kullanılan yöntemler, örneklem genişlikleri, budama yüzdeleri ve tekrar sayıları ilk benzetim çalışması ile aynı alınmıştır.

Tablo 6. Kullanılan gerçek verilere ait istatistikler

Veri Seti	**N**	**Ortalama**	**Medyan**	**Standart Sapma**	**IQR**	**Çarpıklık**	**Basıklık**	$\mu_{tr=0.1}$	$\mu_{tr=0.2}$
Total AA Degrees	2859	268.98	114.0	451.96	281	4.98	47.31	172.82	140.83
GRE Verbal	4637	468.81	470	103.23	150	0.13	2.54	467.29	467.54



Şekil 1. Kullanılan gerçek verilerin histogramları

Tek örneklem için Total AA Degrees veri seti kullanılarak hesaplanan anlam düzeyleri Tablo 7'de verilmiştir. Bootstrap-t yöntemi ile elde edilen sonuçlar B=600 tekrar sayısının kullanılmasını desteklemektedir. Ayrıca $\gamma = 0.2$ budama yüzdesi ile nominal değere daha yakın sonuçlar elde edildiği görülmüştür. Yüzdelik bootstrap yöntemi uygulandığında ise B=2000 tekrar sayısı ile nominal değerin daha iyi korunduğu ve $\gamma = 0.2$ budama yüzdesinin tercih edilebilir olduğu görülmüştür. Total AA Degrees veri seti ile tek örneklem hipotez testi sonuçlarına göre yüzdelik bootstrap yönteminin kullanılması daha uygundur.

Tablo 7. Tek örneklem ve Total AA Degrees veri seti için gerçekleşen anlam düzeyleri

			B		
			600	**1000**	**2000**
Budama Yüzdesi	$\gamma = 0.1$	**Bootstrap-t**	0.0640	0.0651	0.0647
		Yüzdelik bootstrap	0.0588	0.0564	0.0552
	$\gamma = 0.2$	**Bootstrap-t**	0.0610	0.0616	0.0614
		Yüzdelik bootstrap	0.0582	0.0554	0.0548

İki örneklem ve Total AA Degrees veri seti ile gerçekleşen anlam düzeyleri Tablo 8'de verilmiştir. Bootstrap-t ve yüzdelik bootstrap yöntemleriyle gerçekleşen anlam düzeyleri çoğunlukla tutucu eğilim sergilemektedir. Bootstrap-t yöntemi ile elde edilen sonuçlar B=600 tekrar sayısını desteklerken, $\gamma = 0.2$ budama yüzdesi ile nominal değere daha yakın sonuçlar elde edilmiştir. Yüzdelik bootstrap için ise B=2000 tekrar sayısı ve $\gamma = 0.1$ budama yüzdesi önerilebilir. Tüm tekrar sayıları ve budama yüzdelerine bakıldığında, yüzdelik bootstrap yönteminin tercih edilmesi önerilebilir.

Tablo 8. İki örneklem ve Total AA Degrees veri seti için gerçekleşen anlam düzeyleri

			B		
			600	**1000**	**2000**
Budama Yüzdesi	$\gamma = 0.1$	**Bootstrap-t**	0.0296	0.0308	0.0294
		Yüzdelik bootstrap	0.0451	0.0457	0.0459
	$\gamma = 0.2$	**Bootstrap-t**	0.0299	0.0301	0.0302
		Yüzdelik bootstrap	0.0438	0.0444	0.0443

Tek örneklem ve GRE Verbal veri seti ile gerçekleşen anlam düzeyleri Tablo 9'da verilmiştir. Bootstrap-t yöntemi kullanıldığında B=600 tekrar sayısı ile gerçekleşen anlam düzeyleri 0.045 ile 0.055 arasındadır. Ayrıca $\gamma = 0.1$ budama yüzdesi kullanıldığında sonuçlar nominal düzeye daha yakındır. Yüzdelik bootstrap yöntemi kullanıldığında ise B=2000 tekrar sayısı ve $\gamma = 0.2$ budama yüzdesi ile daha iyi sonuçlar elde edilmiştir. Sonuçlara genel olarak bakıldığında bootstrap-t yönteminin kullanılması önerilebilir.

Tablo 9. Tek örneklem ve Greverb veri seti için gerçekleşen anlam düzeyleri

			B		
			600	**1000**	**2000**
Budama Yüzdesi	$\gamma = 0.1$	**Bootstrap-t**	0.0456	0.0447	0.0446
		Yüzdelik bootstrap	0.0578	0.0573	0.0556
	$\gamma = 0.2$	**Bootstrap-t**	0.0454	0.0442	0.0442
		Yüzdelik bootstrap	0.0571	0.0573	0.0548

İki örneklem ve GRE Verbal veri seti ile gerçekleşen anlam düzeyleri Tablo 10'da verilmiştir. Elde edilen tüm sonuçlar nominal düzeyi korumuştur. Bootstrap-t yöntemi kullanıldığında budama yüzdesinin değişmesi sonuçlarda önemli bir değişiklik yaratmamıştır. Tekrar sayısı olarak ise B=600 tercih edilebilir. Yüzdelik bootstrap yöntemi kullanıldığında ise $\gamma = 0.1$ budama yüzdesi ile nominal değere daha yakın sonuçlar elde edilmiştir. Tekrar sayısı olarak ise nominal değere en yakın sonuçları vermesi sebebiyle B=1000 önerilebilir. Tüm tekrar sayıları ve budama yüzdelerine göre yüzdelik bootstrap yöntemini tercih etmenin daha uygun olduğu söylenebilir.

Tablo 10. İki örneklem ve Greverb veri seti için gerçekleşen anlam düzeyleri

			B		
			600	**1000**	**2000**
Budama Yüzdesi	$\gamma = 0.1$	**Bootstrap-t**	0.0475	0.0469	0.0474
		Yüzdelik bootstrap	0.0508	0.0500	0.0494
	$\gamma = 0.2$	**Bootstrap-t**	0.0485	0.0472	0.0474
		Yüzdelik bootstrap	0.0468	0.0493	0.0467

4. Sonuçlar

Bu çalışmada bootstrap-t ve yüzdelik bootstrap yöntemlerinin performansları farklı budama yüzdeleri, tekrar sayıları ve kitle dağılımları altında karşılaştırılmıştır. Tek örneklem ve iki örneklem hipotez testlerinde hangi yöntemin daha tercih edilebilir olduğunun, ayrıca uygun budama yüzdesi ve tekrar sayısı önerilerinin belirtilmesi amaçlanmıştır.

Bootstrap yöntemler budanmış ortalama içeren bir yöntem ile birlikte kullanılırlarsa, uygun budama yüzdesi kitle dağılımının ne kadar asimetrik ve ağır kuyruklu olduğu göz önünde bulundurularak belirlenmelidir. Kitle dağılımının asimetri derecesi arttıkça, budama yüzdesi daha yüksek bir kestirici tercih edilmelidir.

Tekrar sayısının artması bazı durumlarda nominal düzeye yakınlığı iyileştirilebilir, ancak eğer gerçekleşen anlam düzeylerinde önemli bir değişiklik olmuyorsa işlem süresi düşünülerek çok büyük olmayan tekrar sayıları daha uygundur. Nominal anlam düzeyine yakınlık açısından bootstrap-t yöntemi için 600 tekrar, yüzdelik bootstrap için 2000 tekrar önerilmektedir. Gerçek verilerde, yöntemlerin nominal anlam düzeyini koruyamama dereceleri artmıştır ancak önerilen tekrar sayıları ile bu olumsuz etkinin sınırlandığı gözlenmiştir.

Bir kitleye ait budanmış ortalama için yapılan testlerde bootstrap-t Tukey-McLaughlin testi nominal anlam düzeyini koruma açısından yüzdelik bootstrap yöntemine göre daha başarılıdır ve önerilmektedir. İki kitle budanmış ortalamasının karşılaştırılması için ise hem bootstrap-t Yuen testi hem de yüzdelik bootstrap yöntemi önerilmektedir.

5. Kaynaklar

[1] Carpenter J., Bithell J., "Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians" ***Statistics in Medicine***, 19, 1141–1164, 2000

[2] Davison A. C., Hinkley D.V., "Bootstrap methods and their application" *Cambridge University Press*,1997

[3] Efron B., Tibshirani R. , “Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy” ***Statistical Science***, 1, 54 – 77, 1986

[4] Efron B., Tibshirani R., “An Introduction to the Bootstrap” *Chapman and Hall*, London,1993

[5] Hall, P., “On the number of bootstrap simulations required to construct a confidence interval” ***Annals of Statistics***, *14*, 1431–1452, 1986

[6] Wilcox R.R., “Fundamentals of Modern Statistical Methods” Second Edition, *Springer-Verlag* , 2010

[7] Wilcox R.R., “Introduction to Robust Estimation and Hypothesis Testing” Third Edition, *Elsevier Acedemic Press*, 2012

[8] Tukey J. W., McLaughlin D. H., “Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/ Winsorization 1” ***Sankhyā: The Indian Journal of Statistics, Series A***, 331-352, 1963

[9] Yuen K. K., “The two-sample trimmed *t* for unequal population variances” ***Biometrika***, *61*, 165–170, 1974

[10] Efron B., “Bootstrap methods: another look at the jackknife” ***The Annals of Statistics***, 7(1), 1-26,1979

[11] Hoaglin D.C., “Summarizing shape numerically: The g and h distributions, in Exploring data tables, trends, and shapes” Edited by D. Hoaglin, F. Mosteller, J. Tukey , *Wiley* , *New York* ,461-513, 1985

[12] Bradley J.V., “Robustness?” ***British Journal of Mathematical and Statistical Psychology***, 31, 144-152,1978

[13] Ted Micceri's Web Site. http://www.freewebs.com/tedstats.