TARIM BİLİMLERİ DERGİSİ 2000, 6 (4), 57-60
DOI: 10.1501/Tarimbil_0000000996

Eşit ve Farklı Varyanslı Populasyonlardan Alınan Örneklerin Ayıklanmış (Trimmed)-t Testi ile Karşılaştırmasında Gerçekleşen I.Tip Hata ve Testin Gücü

Ensar BAŞPINAR[1] Fikret GÜRBÜZ[1]

Geliş Tarihi : 17.05.2000

Özet: Bu çalışmada, değişik ayıklama oranlarına tabi tutulmuş, eşit ve farklı varyanslı populasyonlardan alınan değişik örnek genişliklerine sahip örneklerin karşılaştırılmasında Ayıklanmış-t testinin kullanımı ele alınmıştır. Bu amaçla, 10000 simülasyon denemesi sonunda, populasyon varyansları bilinmediğinde veya homojen olmadıklarında, mümkün olduğunca eşit genişlikte örneklerle çalışılarak Ayıklanmış-t testinin güvenilir bir şekilde kullanılabileceği sonucuna varılmıştır.

Anahtar Kelimeler: Ayıklanmış-t testi, düzeltilmiş ortalama, ortalamaların karşılaştırması, I.Tip hata, testin gücü

The Realized Type I Error and Power of Test in Comparison of Samples Taken From Equal and Unequal Population Variances With Trimmed-t Test

Abstract: In this study, the efficiency of Trimmed-t Test was examined to compare the samples of varying sizes taken from populations with equal and unequal variances which were trimmed at different rates. The results of 10000 simulation test show that Trimmed-t Test would be used confidentially if the samples are equal in size and the population variance is not known and/or it is not homogenous.

Key Words: Trimmed-t test, winsorized mean, comparison of means, Type I error, power of test

Giriş

Bağımsız iki örnek ortalaması arasındaki farkın test edilmesi, oldukça fazla karşılaşılan bir konudur. Örneklerin normal dağılması ve eşit populasyon varyansına sahip olması ön şartları altında, Student'in t-Testi ile söz konusu örnek ortalamalarının karşılaştırılması yaygın olarak başvurulan yollardan biridir. Ancak, bu test istatistiğini kullanmada iki temel problemle yaygın olarak karşılaşılmaktadır. Bu problemlerin birincisi normal dağılım ön şartı, ikincisi de populasyon varyanslarının eşitliği ön şartıdır. Çünkü, pratikte özellikle psikoloji, ekonomi, tıp, diş hekimliği, eczacılık, gıda ve bazı ziraat konularında, bu ön şartlardan bazen birinin, bazen de her ikisinin sağlanması pek mümkün olmamaktadır. Böyle durumlarda, alternatif testlerden biri olan Ayıklanmış (trimmed)-t testinin kullanımı daha güvenilirdir. Benzer bir durum da dağılımın normal olması ancak, populasyon varyansının bilinmemesi halinde de geçerlidir (Yuen ve Dixon 1973,Wilcox 1995 , Wilcox at al. 1998, Anonymous 1998).

Ayıklanmış t istatistiği, örneklerde yapılan ayıklamaya (trimming) uygun olarak serbestlik derecesi düzeltilmiş Student's t-Dağılımı göstermektedir (Yuen 1974, Wilcox 1986).

Bu çalışmada, değişik ayıklama oranlarına tabi tutulmuş, eşit ve farklı varyanslı populasyonlardan alınan değişik örnek genişliklerine sahip örneklerin karşılaştırılmasında Ayıklanmış-t testinin kullanımı ele alınmıştır.

Materyal ve Yöntem

Çalışmanın materyalini, simülasyon tekniği ile üretilen tesadüf sayıları oluşturmuştur. Üretilen tesadüf sayılarından standart normal tesadüf değişkenleri üretilmiş ve bunlar yardımıyla sırasıyla N(0,1), N(0,5), N(0,10) ve N(0,15) parametreli dört adet populasyon elde edilmiştir. Bu populasyonlardan; N(0,1)-N(0,1), N(0,1)-N(0,5), N(0,1)-N(0,10) ve N(0,1)-N(0,15) olmak üzere ikili populasyon kombinasyonları oluşturulmuş ve bu populasyon kombinasyonlarının her birinden sırasıyla $n_1=n_2=20$, $n_1=n_2=40$, $n_1=n_2=60$, $n_1=20$ ve $n_2=40$, $n_1=40$ ve $n_2=20$, $n_1=20$ ve $n_2=60$, $n_1=60$ ve $n_2=20$, $n_1=40$ ve $n_2=60$, $n_1=60$ ve $n_2=40$ örnek genişliğinde örnekler alınarak %0, %5, %10 ve %20 ayıklama yapılıp, Ayıklanmış-t istatistiği hesaplanmıştır. Bu işlem her populasyon kombinasyonu-örnek genişliği-ayıklama oranı kombinasyonu için 10000 defa yapılıp gerçekleşen I. Tip hata olasılıkları ve populasyon ortalamaları arasındaki farklar $\delta=1.0$, $\delta=2.0$ için uygulanan Ayıklanmış-t testinin güç değerleri ampirik olarak elde edilmiştir. Her ayıklama oranı, alınan örneklerin her iki tarafına da uygulanmıştır (mesela %5'lik ayıklama oranı için, örneklerin en küçük

[1] Ankara Üniv. Ziraat Fak. Biyometri ve Genetik Anabilim Dalı - Ankara

%5'lik ve en büyük %5'lik değerleri olmak üzere toplam %10'luk bir kısmına uygulanmıştır).

$x_1, x_2, \ldots, x_n$ bir populasyondan alınmış n deney ünitesinin sıralanmış değerleri olmak üzere, k defa ayıklanmış (trimmed) ve düzeltilmiş (Winsorized) örnek ortalamaları sırasıyla;

$$\bar{x}_{k,a} = \frac{1}{n-2k}\left(x_{k+1} + x_{k+2} + \ldots + x_{n-k}\right) \text{ ve}$$

$$\bar{x}_{k,d} = \frac{1}{n}\left\{(k+1)x_{k+1} + x_{k+2} + \ldots + x_{n-k-1} + (k+1)x_{n-k}\right\}$$

ifadelerine göre hesaplanırlar. k defa düzeltilmiş sapma kareler toplamı ve varyans ise;

$$SKT_{kd} = (k+1)\left(x_{k+1} - \bar{x}_{kd}\right)^2 + \left(x_{k+2} - \bar{x}_{kd}\right)^2 + \ldots + \left(x_{n-k-1} - \bar{x}_{kd}\right)^2 + (k+1)\left(x_{n-k} - \bar{x}_{kd}\right)^2$$

h=n-2k olmak üzere, $S^2_{kd} = \frac{SKT_{kd}}{h-1}$ ifadelerine göre hesaplandıktan sonra, bağımsız iki ayıklanmış örnek ortalaması arasındaki fark için Ayıklanmış-t istatistiği;

$$t = \frac{\left(\bar{x}_{ka1} - \bar{x}_{ka2}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{S^2_{kd1}}{h_1} + \frac{S^2_{kd2}}{h_2}}}$$

şeklinde hesaplanır. Bu şekilde hesaplanan Ayıklanmış-t istatistiği yaklaşık f serbestlik dereceli t-Dağılımı göstermektedir. f serbestlik derecesi;

$$c = \frac{S^2_{kd1}/h_1}{S^2_{kd1}/h_1 + S^2_{kd2}/h_2} \quad \text{olmak üzere} \quad \frac{1}{f} = \frac{c^2}{h_1 - 1} + \frac{(1-c)^2}{h_2 - 1}$$

ifadesinden hesaplanır. Ayıklanmış-t istatistiğinin kesin dağılımını analitik olarak elde etmek oldukça zordur. Ancak simülasyon teknikleri ile dağılım hakkında fikir edinilebilmektedir (Winer 1971, Yuen 1974, Wilcox 1997).

Çalışmada, tesadüf sayılarının üretilmesinde ve gerekli hesaplamalarda, "Microsoft FORTRAN Power Station"'da Fortran-90 programlama dilinde yazılan programlardan yararlanılmıştır.

Bulgular ve Tartışma

Çalışmada ele alınan populasyon ortalamaları, varyansları, örnek genişlikleri ve populasyon ortalamaları arasındaki farklara (δ=0.0, δ=1.0, δ=2.0) göre 10 000 simülasyon denemesi sonunda elde edilen I. Tip hata olasılıkları Çizelge 1'de, uygulanan Ayıklanmış-t testinin güç değerleri ise Çizelge 2 ve Çizelge 3'de topluca verilmiştir.

Çizelge 1'de, değişik genişlikteki örneklerin alındığı populasyon ortalamaları arasındaki fark, δ=0.0 ($\mu_1=\mu_2$) olacak şekilde düşünülmüş ve 10 000 simülasyon sonunda gerçekleşen I. Tip Hata olasılıkları değişik ayıklama oranları için ampirik olarak elde edilmiştir. Örneklerin alınmış oldukları populasyon ortalamaları eşit olmakla birlikte, varyansları arasında, 1,5,10 ve 15 kat fark olmasına rağmen örnek genişlikleri ne olursa olsun (ele alınan örnek genişliği kombinasyonları için) bütün ayıklama oranlarında I. Tip hata olasılıklarının deneme başında kararlaştırılan α=%5'ten aşırı bir sapma göstermedikleri Çizelge 1'den anlaşılmaktadır. Bu durumda, populasyon varyansları arasında 15 kat fark olduğu veya populasyon varyansları bilinmediği durumlarda, hiç ayıklama yapılmasa bile, Ayıklanmış-t testinin 1-α=0.95 güven katsayısı ile kullanılabileceği söylenebilir.

Çizelge 2'de, değişik genişlikteki örneklerin alındığı populasyon ortalamaları arasındaki fark, δ=1.0 (μ_1=0 ve μ_2=1.0) olacak şekilde düşünülmüş ve 10 000 simülasyon sonunda uygulanan Ayıklanmış-t testinin güç değerleri değişik ayıklama oranları için ampirik olarak elde edilmiştir. Bu Çizelgede, populasyon varyansları eşit olduğunda, örnek genişlikleri ve bunların kombinasyonu ne olursa olsun testin gücünün her ayıklama oranında güvenilir güç sınırları içinde (%80'den büyük) kaldığı görülmektedir. Ancak, populasyon varyansları arasında 5 kat fark olması halinde testin gücünün sadece $n_1=n_2$=60 örnek genişlikleri kombinasyonunda güvenilir güç sınırları içinde kalmakta, bunun dışındaki bütün varyans ve örnek genişliği kombinasyonlarında güvenilir sınırlar dışında değer almaktadır. Bu durumda, populasyon ortalamaları arasında 1.0 standart sapma ve varyansları arasında da 5 kat fark olduğu veya populasyon varyansları bilinmediği durumlarda, hiç ayıklama yapılmasa bile, Ayıklanmış-t testinin kullanılabileceği söylenebilir.

Çizelge 3'de, değişik genişlikteki örneklerin alındığı populasyon ortalamaları arasındaki fark, δ=2.0 (μ_1=0 ve μ_2=2.0) olacak şekilde düşünülmüş ve 10 000 simülasyon sonunda uygulanan Ayıklanmış-t testinin güç değerleri değişik ayıklama oranları için ampirik olarak elde edilmiştir. Bu Çizelgede, populasyon varyansları eşit olduğunda, örnek genişlikleri ve bunların kombinasyonu ne olursa olsun testin gücünün her ayıklama oranında güvenilir güç sınırları içinde (%80'den büyük) kaldığı görülmektedir. Ancak, populasyon varyansları arasında 5 kat fark olması halinde testin gücünün $n_1=n_2$=20 örnek genişlikleri kombinasyonunda bile bütün ayıklama oranlarında güvenilir güç sınırları içinde kalmaktadır. Örnek genişliği kombinasyonları, $n_1=n_2$=40 ve $n_1=n_2$=60 olduğunda bütün varyans farklarında ve her ayıklama oranında testin gücünün %80'in üzerinde olduğu görülmektedir. Örnek genişliklerinin farklı olduğu durumlarda, n_1=20 ve n_2=40 (toplam 60) olduğunda populasyon varyansları arasında 15 kat fark olduğunda bile, ayıklama oranı arttıkça testin gücünde de tedrici bir artışın olduğu ve testin gücünün de %85'in üzerinde olduğu görülmektedir. Ancak, örnek genişliği n_1=40 ve n_2=20 kombinasyonunda olduğunda testin güvenilir seviyesine, ayıklama oranı %20 ve populasyon varyansları arasındaki farkın da en fazla 10 kat olduğu durumdan itibaren %78.4 değeri ile ulaştığı ve bundan daha düşük ayıklama oranı ve daha yüksek varyans kombinasyonlarında testin gücünde önemli düşüşler

Çizelge 1. Normal dağılım gösteren iki populasyonda δ=0.0 için elde edilen ampirik I. tip hatalar (%)

Gözlem sayıları		Varyanslar	Her iki taraftan ayıklama oranları			
n_1	n_2	$\sigma^2_1 : \sigma^2_2$	%0	%5	%10	%20
20	20	1 : 1	4.8	5.0	4.8	5.3
		1 : 5	4.9	5.3	4.9	5.5
		1 : 10	5.2	5.2	5.0	5.1
		1 : 15	4.8	5.0	5.1	5.1
40	40	1 : 1	4.9	4.9	4.7	4.9
		1 : 5	5.5	5.1	5.1	5.0
		1 : 10	5.0	5.3	5.3	5.1
		1 : 15	5.2	5.0	4.9	5.3
60	60	1 : 1	4.9	4.9	5.2	4.8
		1 : 5	4.6	4.9	4.7	4.9
		1 : 10	5.1	4.9	5.2	5.1
		1 : 15	5.1	5.3	5.2	5.4
20	40	1 : 1	4.9	5.0	5.0	5.2
		1 : 5	4.9	5.3	4.9	5.1
		1 : 10	5.2	4.7	5.2	5.6
		1 : 15	4.9	5.1	5.2	5.3
40	20	1 : 1	5.2	5.3	4.7	5.4
		1 : 5	5.3	4.9	5.7	5.7
		1 : 10	5.1	5.2	5.0	5.5
		1 : 15	5.0	5.0	5.7	5.6
		1 : 20	5.1	5.1	5.7	5.4
20	60	1 : 1	5.3	5.1	5.0	5.5
		1 : 5	5.1	5.2	4.9	4.9
		1 : 10	5.0	5.2	4.7	5.4
		1 : 15	4.6	5.3	5.4	5.4
60	20	1 : 1	5.0	5.4	5.2	5.5
		1 : 5	5.3	5.0	4.8	5.4
		1 : 10	4.8	4.9	5.2	5.3
		1 : 15	5.0	4.8	4.9	5.4
40	60	1 : 1	4.8	4.8	5.1	5.2
		1 : 5	5.3	5.4	4.4	4.9
		1 : 10	5.1	4.9	5.2	5.4
		1 : 15	5.1	5.1	4.8	5.5
60	40	1 : 1	5.3	5.2	5.0	5.2
		1 : 5	4.8	4.9	5.2	5.3
		1 : 10	5.1	4.9	4.9	5.1
		1 : 15	5.1	4.8	5.1	5.3

Çizelge 2. Normal dağılım gösteren iki populasyonda δ=1.0 için elde edilen ampirik testin gücü (%)

Gözlem sayıları		Varyanslar	Her iki taraftan ayıklama oranları			
n_1	n_2	$\sigma^2_1 : \sigma^2_2$	%0	%5	%10	%20
20	20	1 : 1	87.6	87.5	86.8	84.8
		1 : 5	42.0	42.5	44.0	45.3
		1 : 10	25.0	26.2	27.1	28.4
		1 : 15	18.5	20.1	20.6	22.1
40	40	1 : 1	99.2	99.3	99.2	99.1
		1 : 5	72.3	72.9	73.2	73.7
		1 : 10	46.7	47.2	47.9	48.7
		1 : 15	34.8	34.8	35.2	35.8
60	60	1 : 1	100.0	100.0	100.0	100.0
		1 : 5	87.6	88.6	88.0	87.8
		1 : 10	63.3	64.4	64.1	64.9
		1 : 15	48.3	48.5	48.6	50.2
20	40	1 : 1	94.6	93.8	94.3	91.9
		1 : 5	65.5	66.0	65.2	66.1
		1 : 10	43.8	43.8	43.8	44.8
		1 : 15	32.8	32.9	33.8	34.8
40	20	1 : 1	94.3	95.0	94.7	94.7
		1 : 5	44.3	45.4	47.2	49.4
		1 : 10	26.8	26.7	27.6	29.7
		1 : 15	18.9	19.9	19.7	22.9
20	60	1 : 1	96.3	96.3	95.5	93.6
		1 : 5	77.3	76.9	76.8	76.8
		1 : 10	56.5	56.7	56.8	57.0
		1 : 15	43.6	44.7	44.2	44.3
60	20	1 : 1	96.4	96.7	96.8	96.6
		1 : 5	46.0	47.1	48.5	49.9
		1 : 10	26.4	26.4	28.2	30.0
		1 : 15	18.5	20.1	20.4	21.6
40	60	1 : 1	99.9	99.8	99.8	99.7
		1 : 5	85.5	85.4	86.1	85.5
		1 : 10	61.0	62.7	62.4	62.5
		1 : 15	46.4	46.8	47.5	48.3
60	40	1 : 1	99.8	99.8	99.8	99.8
		1 : 5	73.4	75.1	74.5	76.6
		1 : 10	48.5	47.7	47.9	50.2
		1 : 15	33.9	35.4	35.8	36.5

olmaktadır. Örnek genişliği kombinasyonları n_1=20 ve n_2=60, n_1=60 ve n_2=20 için de durum aynen n_1=20 ve n_2=40 ve n_1=40 ve n_2=20 kombinasyonu gibidir. n_1=40 ve n_2=60 ile n_1=60 ve n_2=40 kombinasyonunda ise bütün populasyon farklarında ve ayıklama oranlarında testin gücü güvenilir sınırlar arasındadır.

Çizelge 3 için genel olarak, örnek genişlikleri eşit olmak kaydıyla toplam 40 adet deney ünitesi ile Ayıklanmış-t testinin istenilen güç değerlerine ulaştığı, örnek genişlikleri farklı olduğunda; mümkün olduğunca büyük varyanslı populasyondan alınan örnek genişliğinin yüksek tutularak en az toplam 60 adet deney ünitesi ile Ayıklanmış-t testinin güvenilir sonuçlar verdiğini söylemek mümkündür.

Çizelge 3. Normal dağılım gösteren iki populasyonda δ=2.0 için elde edilen ampirik testin gücü (%)

Gözlem sayıları		Varyanslar	Her iki taraftan ayıklama oranları			
n_1	n_2	$\sigma^2_1 : \sigma^2_2$	%0	%5	%10	%20
20	20	1 : 1	100.0	100.0	100.0	100.0
		1 : 5	94.4	93.5	94.7	94.7
		1 : 10	74.4	74.0	75.4	76.8
		1 : 15	57.2	59.0	59.1	61.0
40	40	1 : 1	100.0	100.0	100.0	100.0
		1 : 5	99.9	99.9	100.0	99.9
		1 : 10	96.3	96.5	96.5	96.6
		1 : 15	86.6	87.7	88.3	88.8
60	60	1 : 1	100.0	100.0	100.0	100.0
		1 : 5	100.0	100.0	100.0	100.0
		1 : 10	99.6	99.6	99.7	99.7
		1 : 15	96.8	97.0	97.0	97.0

Çizelge 3. (Devamı) Normal dağılım gösteren iki populasyonda δ=2.0 için elde edilen ampirik testin gücü (%)

Gözlem sayıları		Varyanslar	Her iki taraftan ayıklama oranları			
n_1	n_2	$\sigma^2_1 : \sigma^2_2$	%0	%5	%10	%20
20	40	1 : 1	100.0	100.0	100.0	100.0
		1 : 5	99.7	99.8	99.7	99.7
		1 : 10	94.7	94.9	94.9	95.3
		1 : 15	85.1	86.2	85.7	86.8
40	20	1 : 1	100.0	100.0	100.0	100.0
		1 : 5	95.6	95.8	95.7	96.1
		1 : 10	75.1	76.9	76.3	78.4
		1 : 15	57.7	59.2	59.5	62.4
20	60	1 : 1	100.0	100.0	100.0	100.0
		1 : 5	99.9	100.0	100.0	100.0
		1 : 10	98.9	98.9	99.0	98.8
		1 : 15	94.5	95.2	95.4	95.3
60	20	1 : 1	100.0	100.0	100.0	100.0
		1 : 5	95.9	96.3	96.1	96.2
		1 : 10	75.0	77.2	77.1	78.4
		1 : 15	58.8	60.7	60.4	62.5
40	60	1 : 1	100.0	100.0	100.0	100.0
		1 : 5	100.0	100.0	100.0	100.0
		1 : 10	99.5	99.6	99.4	99.4
		1 : 15	96.3	96.5	96.8	96.8
60	40	1 : 1	100.0	100.0	100.0	100.0
		1 : 5	99.9	99.9	99.9	99.9
		1 : 10	97.1	96.8	96.9	97.3
		1 : 15	87.7	87.8	88.4	89.0

Sonuç

Bu simülasyon çalışması sonunda, populasyon varyansları bilinmediğinde veya homojen olmadıklarında, mümkün olduğunca eşit genişlikte örneklerle çalışılarak Ayıklanmış-t testinin güvenilir bir şekilde kullanılabileceği sonucuna varılmıştır.

Kaynaklar

Anonymous, 1998. Robust Determination of Mean Trimmed Mean. Eriţim Adresi: http://osprey6.npac.syr.edu:8080/foilsets/cps713stat/node199.html. Erişim Tarihi:29.04.1999.

Wilcox, R.R. 1986. New Monte Carlo Results of the Robustnes of the ANOVA F, W and F* Statistics. Commun. Statist.-Simula., 15 (4), 933-943.

Wilcox, R. R. 1995. The Practical Importance of Heteroscedastic Methods, Using Trimmed Means Versus Means, and Designing Simulation Studies. Brit.Jour. of Mat. and Stat. Psychology. 48 (1), 99-114.

Wilcox, R. R. 1997. A Bootstrap Modification of the Alexander-Govern ANOVA Method, Plus Comments on Comparing Trimmed Means. Educational and Psychological Measurement. 57 (4), 655-665.

Wilcox, R. R., H. J. Keselman ve R. K. Kowalchuk, 1998. Can Test for Treatment Group Equality Be Improved - The Bootstrap and Trimmed Means Conjecture. Brith.Jour. of Math. & Stat. Psychology. 51, Iss MAY, 123-134.

Winer, B. J. 1971. Statistical Principles in Experimental Design. Second Ed. McGraw-Hill Book Co., New York. 907 s.

Yuen, K. K. ve W. J. Dixon, 1973. The Approximate Behavior and Performance of the Two-Sample Trimmed t. Biometrika 60, (2), 369-374.

Yuen, K. K. 1974. The Two-Sample Trimmed t for Unequal Population Variances. Biometrika, 61 (1), 165-170.