

Değişik Varyans Oranlı Normal Populasyonlardan Alınan Değişik Örnek Genişliğindeki İki Örnekte Student-t, Welch ve Ayıklanmış-t Testlerinin Uygulanması ile Elde Edilen I. Tip Hata ve Testin Gücü

Ensar BAŞPINAR¹

Geliş Tarihi: 02.01.2001

Özet : Bu çalışmada, değişik varyans oranlı populasyonlardan alınan iki örnek ortalaması arasındaki farkın Student-t, Welch ve %5 ayıklama uygulanmış Ayıklanmış-t Testleri ile karşılaştırılması sonucu elde edilen I. Tip hata olasılıkları ve testin gücü ele alınmıştır. Bu amaçla, populasyon ortalamaları arasındaki farklar $\delta_i=0,0,1,0,1,5,2,0$ standart sapma olacak şekilde 10000 simülasyon denemesi yapılmıştır. Her bir simülasyon denemesinde söz konusu populasyonlardan, örnek genişlikleri eşit ve farklı olan örnekler alınmış ve bu örnekler söz konusu testler uygulanarak gerçekleşen I. Tip hata olasılıkları ve testlerin gücü ampirik olarak elde edilmiştir. Sonuçta, örnek genişliklerine bağlı olarak populasyon varyansları homojen olduğunda her üç test, populasyon varyansları heterojen olduğunda ise Welch ve Ayıklanmış-t Testlerinin daha güvenilir olduğu görülmüştür. Ayrıca örnek genişlikleri arttıkça söz konusu testlerden özellikle Welch ve Ayıklanmış-t Testlerinin populasyon varyanslarının heterojenliğinden fazla etkilenmediği de gözlenmiştir.

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

Type I Error and Power of Tests When Applying The Student's-t, Welch and Trimmed-t Tests to Two Samples of Various Sizes from Normal Populations Having Various Variance Ratios

Abstract: In this study, Type I Error and Power of Tests were examined in two samples of various sizes from normal populations with various variance ratios when the Student's-t, Welch and Trimmed-t Tests with 5 percent trimming were applied. For this aim, the 10000 simulation experiments were run for each combinations of differences of the populations which differ from each other by $\delta_i=0,0,1,0,1,5,2,0$ standard deviation. Each of these experiments was taken with samples of equal and unequal sample sizes and applied different tests. At the end of these procedures, Type I error rates and power of the tests was observed empirically. The 10000-simulation tests show that when the population variances are homogenous all of the tests when they are heterogeneous Welch and Trimmed-t Test would be used confidentially. On the other hand, when the sample sizes were increased, Welch and Trimmed-t Tests were not influenced from the heterogeneity of populations variance.

Key Words: Student's-t test, Welch test, 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. Gözlemlerin normal dağılması ve eşit populasyon varyansına sahip olması ön şartları altında, Student-t testi ile söz konusu örnek ortalamalarının karşılaştırılması yaygın olarak başvurulan yollardan biridir. Ancak, bu test istatistiğini kullanmada iki temel problemle yaygın olarak karşılaşmaktadır. 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 kullanılmak üzere geliştirilmiş çeşitli yöntemler vardır.

Bu çalışmada, Student-t testi, Welch testi ve Ayıklanmış (trimmed)-t testi ele alınmış ve bunların, değişik varyans oranlarında eşit ve farklı ortalamalı normal

dağılım kombinasyonlarından alınan değişik örnek genişliklerindeki örneklerin karşılaştırılması her bir kombinasyon için 10000 simülasyon denemesi ile yapılarak gerçekleşen I. Tip hata olasılıkları ve testin gücü elde edilmiştir. Benzer bir durum da dağılımın normal olması ancak, populasyon varyansının bilinmemesi halinde de geçerlidir (Devenport ve Webster 1970, Yuen ve Dixon 1973, Wilcox 1995, Wilcox at al. 1998, Anonymous 1998).

Materyal ve Yöntem

Çalışmanın materyalini, simülasyon tekniği ile üretilen tesadüf sayıları oluşturmuştur. Üretilen tesadüf sayılarından standart normal tesadüf değişkenleri üretilmiş ve bunlar yardımıyla sırasıyla $N(0,1)$, $N(0,2)$, $N(0,4)$, $N(0,6)$, $N(0,8)$, $N(0,10)$, $N(0,12)$, $N(0,14)$, $N(0,16)$, $N(0,18)$

¹ Ankara Üniv. Ziraat Fak.Zootekni Bölümü - Ankara

ve $N(0,20)$ parametrelili onbir adet normal populasyon elde edilmiştir. Bu populasyonların; $N(0,1)-N(0,1)$, $N(0,1)-N(0,2)$, $N(0,1)-N(0,4)$,..., $N(0,2)-N(0,1)$, $N(0,2)-N(0,2)$, $N(0,2)-N(0,4)$,... ve $N(0,20)-N(0,1)$, $N(0,20)-N(0,2)$, $N(0,20)-N(0,4)$,..., $N(0,20)-N(0,20)$ 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 sırasıyla Student t-testi, Welch testi ve dağılımların her iki tarafından %5 ayıklama yapıp, Ayıklanmış-t testi uygulanmıştır. Bu işlem her populasyon kombinasyonu-örnek genişliği-varyans oranı kombinasyonu için 10000 defa yapıp gerçekleşen I. Tip hata olasılıkları ve populasyon ortalamaları arasındaki farklar; $\delta=0.5$, $\delta=1.0$, $\delta=1.5$, $\delta=2.0$ standart sapma için uygulanan testlerin güç değerleri, ret edilen hipotez sayıları saydırılarak ampirik olarak elde edilmiştir. Welch ve Ayıklanmış-t istatistiği, düzeltilmiş serbestlik dereceli Student t-Dağılımı göstermektedir (Devenport ve Webster 1970, Yuen 1974, Wilcox 1986).

Student-t testi; bağımsız iki örnek ortalaması arasındaki farkın tesadüfi olup olmadığına ilişkin hipotez testini,

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2}}} \text{ ifadesine göre}$$

hesaplanan t istatistiğinin $(n_1-1)+(n_2-1)$ serbestlik dereceli t-dağılımı gösterdiğini varsayarak kontrol etme esasına dayanmaktadır. Welch testi ise aynı hipotezi, Student-t testindeki gibi hesaplanan t' istatistiğinin,

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}} \text{ serbestlik dereceli t-dağılımı}$$

gösterdiğini varsaymaktadır. Bu şekilde hesaplanan serbestlik derecesi kendisine en yakın küçük tam sayıya yuvarlanmaktadır (Zar 1999).

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

$$\bar{X}_{k,a} = \frac{1}{n-2k} (x_{k+1} + x_{k+2} + \dots + x_{n-k}) \text{ ve}$$

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

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

$$SKT_{kd} = (k+1)(x_{k+1} - \bar{X}_{kd})^2 + (x_{k+2} - \bar{X}_{kd})^2 + \dots + (x_{n-k-1} - \bar{X}_{kd})^2 + (k+1)(x_{n-k} - \bar{X}_{kd})^2$$

$h=n-2k$ olmak üzere, $S_{kd}^2 = \frac{SKT_{kd}}{h-1}$ ifadelerine göre

hesaplandıktan sonra, bağımsız iki ayıklanmış örnek ortalaması arasındaki fark için Ayıklanmış-t istatistiği;

$$t = \frac{(\bar{X}_{ka1} - \bar{X}_{ka2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_{kd1}^2}{h_1} + \frac{S_{kd2}^2}{h_2}}} \text{ şeklinde hesaplanır. Bu}$$

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

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

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, Zar 1999).

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 ($\delta=0.0$, $\delta=0.5$, $\delta=1.0$, $\delta=1.5$, $\delta=2.0$) göre 10 000 simülasyon denemesi sonunda elde edilen I. Tip hata olasılıkları Çizelge 1'de, uygulanan testlerin güç değerleri ise Çizelge 2, Çizelge 3, Çizelge 4 ve Çizelge 5'de topluca verilmiştir.

Çizelge 1'de, değişik genişlikteki örneklerin alındığı populasyon ortalamaları arasındaki fark, $\delta=0.0$ ($\mu_1=\mu_2$) olacak şekilde düşünülmüş ve 10 000 simülasyon sonunda gerçekleşen I. Tip Hata olasılıkları Student t, Welch ve dağılımların her iki tarafından da %5'lik ayıklama oranı uygulanmış Ayıklanmış-t testleri için ampirik olarak elde edilmiştir. Örneklerin alınmış oldukları populasyon ortalamaları eşit olmakla birlikte, varyansları arasında sırasıyla, 1,2,4,6,8,10,12,14,16,18 ve 20 kat fark olmasına rağmen örnek genişlikleri eşit ($n_1=n_2=20,40,60$) olduğunda bütün testlerde I. Tip hata olasılıklarının deneme başında karşılaştırılan α =%5'ten aşırı bir sapma göstermedikleri Çizelge 1'den anlaşılmaktadır. Bu durumda, populasyon varyansları arasında 20 kat fark (standart sapma cinsinden) olduğu veya populasyon varyansları bilinmediği durumlarda örnek genişlikleri eşit olmak kaydıyla, ele alınan testlerin $1 - \alpha = 0.95$ güven

katsayısı ile kullanılabileceği söylenebilir. Ancak örnek genişliklerinin eşit olmadığı durumlarda, Student-t testinin sadece varyans oranları 1:1 olduğu durumlarda I. Tip hata olasılığını beklenen seviyede gerçekleştirebildiği, $n_1 > n_2$ olması halinde (varyansı küçük olan populasyondan alınan bireylerin çok olması halinde) I. Tip hata olasılığında aşırı düşme, $n_1 < n_2$ olması halinde de (varyansı büyük olan populasyondan alınan bireylerin çok olması halinde) I. Tip hata olasılığında aşırı yükselme olduğu görülmektedir. Buna karşılık gerek Welch ve gerekse Ayıklanmış-t testleri I. Tip hata olasılığını başlangıçta belirlenen %5 sınırında muhafaza etmektedirler.

Çizelge 2'de, değişik genişlikteki örneklerin alındığı populasyon ortalamaları arasındaki fark, $\delta=0.5$ ($\mu_1=0$ ve $\mu_2=0.5$) standart sapma olacak şekilde düşünülmüş ve 10000 simülasyon sonunda, uygulanan Student t, Welch ve Ayıklanmış-t testlerinin güç değerleri ampirik olarak elde edilmiştir. Bu Çizelgede, populasyon varyansları ve örnek genişlikleri eşit olmak kaydıyla ancak $n_1=n_2=60$ olması halinde testlerin gücünün güvenilir güç sınırları içinde (%75'ten büyük, Zar 1999) kaldığı görülmektedir. Diğer bütün kombinasyonlarda güç değerleri oldukça düşüktür. Ancak populasyon ortalamaları arasındaki farkın $\delta = 0.5$ olmasının, böyle bir sonuç elde

etmede rolü olduğu da göz ardı edilmemelidir. Çünkü ortalamalar arasındaki fark düştükçe güç değerlerinin de düşmesi beklenen bir durumdur (Akdeniz 1984, Zar 1999)

Çizelge 3'te, değişik genişlikteki örneklerin alındığı populasyon ortalamaları arasındaki fark, $\delta=1.0$ ($\mu_1=0$ ve $\mu_2=1.0$) olacak şekilde düşünülmüş ve 10000 simülasyon sonunda uygulanan Student t, Welch ve Ayıklanmış-t testinin güç değerleri ampirik olarak elde edilmiştir. Bu Çizelgede, populasyon varyansları eşit olduğunda, örnek genişlikleri ve bunların kombinasyonu ne olursa olsun testin gücünün her yöntemde güvenilir güç sınırları (%75<güç<%90, Zar 1999) içinde kaldığı görülmektedir. Ancak, populasyon varyansları arasında 6 kat fark olması halinde testin gücünün $n_1=n_2=60$ örnek genişlikleri kombinasyonunda bütün yöntemlerde, $n_1=40$, $n_2=60$ kombinasyonunda (büyük varyanslı populasyondan alınan örnek genişliği büyük olduğunda) Welch ve Ayıklanmış-t testlerinin, $n_1=60$, $n_2=40$ kombinasyonunda (büyük varyanslı populasyondan alınan örnek genişliği küçük olduğunda) da sadece Student-t testinin güvenilir olduğu görülmektedir. Bu durum, populasyon varyansları arasındaki fark arttıkça ve örnek genişliklerinin de büyük varyanslı populasyon lehinde dengesiz olması halinde Welch ve Ayıklanmış-t testlerinin, Student-t testine göre daha güvenilir oldukları şeklinde yorumlanabilir.

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

n_1	n_2	Uygulanan tst	$\sigma_1 : \sigma_2$										
			1:1	1:2	1:4	1:6	1:8	1:10	1:12	1:14	1:16	1:18	1:20
20	20	Student-t	4.9	4.9	5.3	5.3	5.5	5.5	5.5	5.3	5.8	5.4	5.3
		Welch	4.4	4.3	4.5	4.4	4.4	4.6	4.4	4.3	4.8	4.4	4.2
		Ayıklanmış-t	5.0	5.0	4.8	5.1	5.3	5.2	5.2	5.2	4.9	4.8	4.9
40	40	Student-t	5.0	4.8	5.0	5.7	5.6	5.3	5.3	5.6	5.4	4.8	5.7
		Welch	4.8	4.5	4.6	5.2	5.0	4.7	4.8	5.1	4.7	4.2	5.0
		Ayıklanmış-t	4.9	4.9	4.7	4.8	4.9	5.3	5.1	5.0	4.5	5.0	5.1
60	60	Student-t	5.2	5.1	5.0	4.8	5.5	5.3	5.5	5.2	5.6	5.1	4.9
		Welch	5.0	4.9	4.7	4.5	5.2	5.1	5.2	4.9	5.2	4.7	4.5
		Ayıklanmış-t	4.9	5.0	5.2	5.2	4.7	4.9	4.8	4.9	4.9	5.2	5.2
20	40	Student-t	4.9	2.9	1.6	1.4	1.1	1.1	1.1	0.9	1.0	0.9	1.0
		Welch	4.6	4.5	4.5	4.5	4.7	4.8	4.8	4.5	5.1	5.0	4.5
		Ayıklanmış-t	5.0	5.3	5.1	4.9	5.0	4.7	5.3	5.4	4.9	5.1	5.1
40	20	Student-t	5.0	8.3	11.3	12.9	13.3	14.8	15.1	15.4	16.3	14.9	15.5
		Welch	4.5	4.8	4.7	4.6	4.2	4.6	4.7	4.8	4.9	4.5	4.6
		Ayıklanmış-t	5.3	4.8	4.9	5.1	4.9	5.2	4.6	5.1	5.4	5.0	5.1
20	60	Student-t	4.7	2.1	0.9	0.6	0.3	0.3	0.3	0.2	0.1	0.1	0.1
		Welch	4.5	4.7	4.4	4.7	4.6	4.8	4.9	4.8	4.6	5.0	4.7
		Ayıklanmış-t	5.1	5.0	4.7	4.6	4.8	5.2	4.9	4.7	4.7	5.2	5.2
60	20	Student-t	4.9	9.2	15.7	18.3	18.9	21.0	21.8	22.7	23.1	23.4	23.6
		Welch	4.6	4.1	4.8	4.6	4.0	4.5	4.3	4.4	4.5	4.5	4.2
		Ayıklanmış-t	5.4	5.3	4.9	5.1	5.3	4.9	5.4	5.3	4.9	5.2	5.5
40	60	Student-t	4.8	3.4	2.8	2.4	2.0	2.5	2.2	2.0	2.0	2.3	2.1
		Welch	4.6	4.4	4.9	4.7	4.2	4.9	4.8	4.6	4.8	5.0	4.9
		Ayıklanmış-t	4.8	4.9	5.1	5.3	5.1	4.9	5.3	5.0	5.0	5.3	5.0
60	40	Student-t	5.3	6.9	8.2	9.1	9.3	10.0	10.0	10.3	10.6	10.7	11.4
		Welch	5.0	5.0	4.8	4.7	4.8	4.8	4.7	4.9	4.6	5.0	5.2
		Ayıklanmış-t	5.2	4.7	5.4	5.1	4.7	4.9	5.3	5.0	4.9	5.2	4.8

Çizelge 2. Normal dağılım gösteren iki populasyonda $\delta=0.5$ için elde edilen ampirik testin gücü (%)

n ₁	n ₂	Uygulanan Test	$\sigma_1^2 : \sigma_2^2$										
			1:1	1:2	1:4	1:6	1:8	1:10	1:12	1:14	1:16	1:18	1:20
20	20	Student-t	33.8	24.2	16.1	13.3	12.1	11.2	9.5	9.1	8.6	8.5	8.0
		Welch	31.7	22.4	14.5	11.7	10.4	9.6	7.9	8.0	7.1	7.0	6.6
		Ayıklanmış-t	34.1	24.3	16.7	13.1	11.2	10.0	9.7	9.0	8.5	8.1	7.1
40	40	Student-t	59.0	43.3	29.0	22.7	17.8	16.3	13.3	12.8	11.9	11.3	11.3
		Welch	58.1	42.3	27.8	21.5	16.8	15.3	12.5	11.8	11.1	10.5	10.4
		Ayıklanmış-t	59.9	43.8	29.7	22.0	18.5	16.1	14.3	12.6	12.1	11.0	10.6
60	60	Student-t	76.8	60.3	40.9	31.7	24.6	21.6	19.4	17.0	15.1	15.1	13.5
		Welch	76.3	59.7	40.1	30.7	23.6	20.6	18.6	16.0	14.3	14.4	12.9
		Ayıklanmış-t	76.9	61.1	40.8	30.8	24.7	21.1	18.5	16.8	14.9	14.5	13.5
20	40	Student-t	43.9	26.2	13.2	8.3	6.0	5.0	3.9	3.3	3.0	2.8	2.5
		Welch	41.7	32.9	23.5	18.7	15.0	13.7	12.5	11.4	10.8	10.6	9.4
		Ayıklanmış-t	42.3	34.2	25.1	20.0	16.5	14.6	13.7	13.3	11.1	10.9	9.9
40	20	Student-t	44.4	35.6	29.1	26.3	24.7	22.9	22.3	21.0	21.2	21.3	20.4
		Welch	41.9	26.0	16.3	12.2	10.6	9.2	8.3	7.3	7.5	7.5	6.7
		Ayıklanmış-t	43.6	28.8	17.7	14.0	11.8	10.5	10.0	8.9	8.4	8.7	7.9
20	60	Student-t	47.8	28.9	11.1	5.9	4.0	3.0	2.2	1.7	1.5	1.1	1.0
		Welch	44.8	38.0	29.3	23.7	20.1	18.8	16.2	14.7	14.1	12.8	12.2
		Ayıklanmış-t	45.7	40.0	30.0	25.1	21.6	18.7	16.6	15.7	14.4	13.1	12.4
60	20	Student-t	47.8	41.7	36.2	33.9	32.3	31.8	30.5	29.8	29.3	29.1	29.4
		Welch	44.4	27.7	16.8	13.3	10.7	9.9	8.4	7.5	7.8	6.5	6.7
		Ayıklanmış-t	48.0	30.1	18.4	14.4	12.1	10.6	9.9	9.0	8.6	7.9	8.3
40	60	Student-t	67.7	47.6	27.8	20.3	14.8	11.7	10.1	9.0	7.9	7.6	6.3
		Welch	66.8	52.2	36.1	28.7	23.1	19.3	17.4	15.6	14.7	13.8	12.1
		Ayıklanmış-t	68.0	53.4	37.7	28.8	23.8	20.4	18.1	16.8	15.2	13.8	13.4
60	40	Student-t	68.3	53.2	38.8	32.2	27.0	23.4	22.9	21.7	20.6	19.0	18.8
		Welch	67.3	46.8	28.8	22.1	17.3	14.4	13.7	12.3	11.5	10.4	10.3
		Ayıklanmış-t	68.1	48.6	30.4	22.6	18.6	17.0	13.9	12.9	12.4	11.5	10.5

Çizelge 3. Normal dağılım gösteren iki populasyonda $\delta=1.0$ için elde edilen ampirik testin gücü (%)

n ₁	n ₂	Uygulanan Test	$\sigma_1^2 : \sigma_2^2$										
			1:1	1:2	1:4	1:6	1:8	1:10	1:12	1:14	1:16	1:18	1:20
20	20	Student-t	87.1	70.9	49.3	38.5	30.7	26.3	23.3	21.6	18.8	18.0	16.6
		Welch	85.9	68.8	46.3	35.5	27.6	23.6	20.8	18.8	16.4	15.6	14.3
		Ayıklanmış-t	87.5	71.2	49.6	38.3	30.7	26.2	21.9	20.9	19.0	17.4	16.0
40	40	Student-t	99.2	95.2	79.9	64.5	55.0	47.1	40.8	36.9	32.7	30.7	27.8
		Welch	99.1	94.6	78.8	63.0	53.5	45.2	39.1	35.3	30.9	29.0	26.0
		Ayıklanmış-t	99.3	95.5	80.2	66.2	55.5	47.2	41.0	35.9	33.1	29.5	27.8
60	60	Student-t	99.9	99.3	93.5	83.2	73.0	63.1	56.8	50.8	46.4	41.6	39.0
		Welch	99.9	99.3	93.1	82.4	72.1	62.0	55.8	49.6	44.9	40.3	37.9
		Ayıklanmış-t	100.0	99.4	92.9	82.5	71.9	64.4	57.0	51.1	46.1	42.5	37.7
20	40	Student-t	94.8	81.8	55.9	38.4	27.9	21.5	17.3	14.1	12.9	10.5	10.0
		Welch	94.0	86.3	71.5	57.5	48.2	42.3	37.0	33.0	30.5	27.2	25.5
		Ayıklanmış-t	93.8	87.3	72.3	59.8	50.3	43.8	38.5	33.8	31.3	28.4	26.7
40	20	Student-t	95.0	85.1	68.0	58.5	51.2	46.3	42.5	40.0	37.6	36.2	33.8
		Welch	94.1	76.3	50.6	38.2	29.6	25.0	21.7	18.9	16.6	15.9	14.6
		Ayıklanmış-t	95.0	78.9	54.1	40.5	31.9	26.7	23.7	21.1	18.5	17.6	16.2
20	60	Student-t	97.2	86.0	57.4	37.8	26.3	18.3	14.3	10.5	8.9	6.9	6.0
		Welch	96.3	91.7	80.7	69.8	62.5	55.8	49.8	44.8	41.0	37.9	34.2
		Ayıklanmış-t	96.3	92.3	81.8	72.9	63.0	56.7	50.7	46.7	41.8	38.6	36.2
60	20	Student-t	97.1	88.7	75.7	67.3	60.5	56.0	52.4	49.8	45.7	44.7	43.1
		Welch	96.1	78.1	51.8	38.1	29.9	25.0	22.0	19.3	16.4	15.4	14.6
		Ayıklanmış-t	96.7	81.9	55.3	40.9	32.8	26.4	23.6	21.0	19.3	17.0	16.6
40	60	Student-t	99.8	97.8	85.1	70.7	57.1	47.1	40.3	35.1	30.6	27.8	25.3
		Welch	99.8	98.3	89.7	79.7	68.4	60.6	53.7	48.5	44.1	40.4	37.8
		Ayıklanmış-t	99.8	98.3	90.4	79.7	70.1	62.7	55.5	50.2	44.2	41.3	37.6
60	40	Student-t	99.8	98.0	87.4	77.4	67.1	59.8	54.3	51.3	46.7	43.9	39.8
		Welch	99.8	97.1	81.3	66.8	54.3	46.2	40.2	35.9	31.8	29.5	26.7
		Ayıklanmış-t	99.8	97.0	82.9	67.5	56.5	47.7	42.2	36.8	32.7	30.0	27.5

Çizelge 4'de, değişik genişlikteki örneklerin alındığı populasyon ortalamaları arasındaki fark, $\delta=1.5$ ($\mu_1=0$ ve $\mu_2=1.5$) olacak şekilde düşünülmüş ve 10000 simülasyon sonunda uygulanan Student t, Welch ve Ayıklanmış-t testinin güç değerleri ampirik olarak elde edilmiştir. Bu Çizelgede, populasyon varyansları eşit olduğunda, örnek genişlikleri ve bunların kombinasyonu ne olursa olsun testin gücünün her yöntemde %95'in üzerinde olduğu görülmektedir.

$n_1=n_2=60$ ($n=120$) olduğunda, populasyon varyansları arasında 18 kat fark olması halinde bile her üç test yönteminin de güç bakımından birbirine yakın ve yaklaşık %75 olarak gerçekleşirken, $n_1=20$, $n_2=40$ örnek genişliği kombinasyonunda (büyük varyanslı populasyondan alınan birey sayısı fazla olduğunda) Welch ve Ayıklanmış-t Testlerinin güç değerleri, populasyonların varyansları arasındaki fark 10 kat oluncaya kadar Student-t Testinden daha yüksek olarak gerçekleşmiştir. $n_1=40$, $n_2=20$ örnek genişliği kombinasyonunda (küçük varyanslı populasyondan alınan birey sayısı fazla olduğunda) ise populasyonların varyansları arasındaki fark 8 kat artıncaya kadar Student-t Testi daha güçlü olarak bulunmuştur. $n_1=20$, $n_2=60$ örnek genişliği kombinasyonunda (büyük varyanslı populasyondan alınan birey sayısı fazla olduğunda) Welch ve Ayıklanmış-t Testlerinin güç değerleri, populasyonların varyansları arasındaki fark 16 kat oluncaya kadar Student-t Testinden daha yüksek olarak gerçekleşmiştir. $n_1=60$, $n_2=20$ örnek genişliği kombinasyonunda (küçük varyanslı populasyondan alınan birey sayısı fazla olduğunda) ise populasyonların varyansları arasındaki fark 12 kat artıncaya kadar Student-t Testi daha güçlü olarak bulunmuştur. $n_1=40$, $n_2=60$ ve

$n_1=60$, $n_2=40$ örnek genişlikleri kombinasyonlarında da benzer sonuçlar elde edilmiştir. Bunların bir sonucu olarak, populasyon ortalamaları arasındaki fark 1.5 standart sapma olduğunda, örnek genişlikleri eşit ve populasyonların varyans oranları 4 kat oluncaya kadar Student-t, Welch ve Ayıklanmış-t Testlerinden her hangi biri aynı güvenilirlikte kullanılabilirken, populasyonların varyans oranları arttıkça ve büyük varyanslı populasyondan alınan birey sayısı küçük varyanslı populasyondan alınan birey sayısından fazla olduğunda Welch ve Ayıklanmış-t Testleri, Student-t Testinden daha güvenilir, küçük varyanslı populasyondan alınan birey sayısı büyük varyanslı populasyondan alınan birey sayısından fazla olduğunda da Student-t Testi, Welch ve Ayıklanmış-t Testlerinden daha güvenilir olarak kullanılabilir.

Çizelge 5'de, değişik genişlikteki örneklerin alındığı populasyon ortalamaları arasındaki fark, $\delta=2.0$ ($\mu_1=0$ ve $\mu_2=2.0$) olacak şekilde düşünülmüş ve 10000 simülasyon sonunda uygulanan Student t, Welch ve Ayıklanmış-t testinin güç değerleri ampirik olarak elde edilmiştir. Bu Çizelgede, populasyon varyansları eşit olduğunda, örnek genişlikleri ve bunların kombinasyonu ne olursa olsun testin gücünün her yöntemde %100 olduğu görülmektedir. Örnek genişlikleri eşit olmak şartıyla, populasyonların varyans oranları arttıkça her üç testinde güç değerlerinde de bir azalma olmaktadır. $n_1=n_2=20$ ($n=40$) olduğunda, populasyon varyansları arasında 10 kat, $n_1=n_2=40$ ($n=80$) olduğunda, populasyon varyansları arasında 20 kat fark olması halinde her üç test yönteminin de istenilen güç değerlerine ulaştıkları görülmektedir.

Çizelge 4. Normal dağılım gösteren iki populasyonda $\delta=1.5$ için elde edilen ampirik testin gücü (%)

n_1	n_2	Uygulanan Test	$\sigma_1 : \sigma_2$										
			1:1	1:2	1:4	1:6	1:8	1:10	1:12	1:14	1:16	1:18	1:20
20	20	Student-t	99.6	96.5	83.2	69.3	58.4	49.5	44.3	40.0	35.4	32.7	30.6
		Welch	99.6	95.9	81.2	66.5	55.1	45.9	40.5	36.2	31.6	29.4	27.1
		Ayıklanmış-t	99.8	96.5	83.2	69.6	58.7	51.5	43.8	40.0	35.6	31.4	30.1
40	40	Student-t	100.0	100.0	96.7	94.4	87.7	79.8	73.3	67.6	63.0	57.5	53.8
		Welch	100.0	100.0	98.5	94.0	86.8	78.3	71.6	66.1	61.1	55.3	51.7
		Ayıklanmış-t	100.0	99.9	98.7	94.3	87.9	80.6	74.2	68.3	61.6	57.6	53.4
60	60	Student-t	100.0	100.0	99.9	99.2	97.0	93.2	89.1	84.3	80.0	75.4	71.3
		Welch	100.0	100.0	99.9	99.1	96.8	92.8	88.4	83.4	78.9	74.4	70.3
		Ayıklanmış-t	100.0	100.0	99.9	99.1	96.8	93.7	88.9	84.9	80.5	75.1	71.2
20	40	Student-t	100.0	99.4	91.6	78.3	64.8	53.6	45.3	38.3	31.9	27.7	25.9
		Welch	99.9	99.6	96.4	90.5	82.7	75.5	69.0	62.9	57.4	52.7	50.7
		Ayıklanmış-t	99.9	99.7	97.2	90.8	84.2	77.3	71.0	64.4	59.5	55.4	51.6
40	20	Student-t	100.0	99.2	94.0	85.3	78.1	72.6	66.2	61.6	57.5	54.5	51.7
		Welch	99.9	97.9	84.8	69.3	57.2	48.8	42.1	37.5	33.0	30.4	27.1
		Ayıklanmış-t	99.9	98.6	86.9	72.3	60.8	52.7	44.8	40.1	35.7	33.5	29.8
20	60	Student-t	100.0	99.7	95.3	82.8	69.4	55.6	44.7	37.2	30.8	26.0	21.8
		Welch	100.0	99.9	99.1	96.8	93.1	88.5	83.1	79.1	74.4	70.9	65.6
		Ayıklanmış-t	100.0	99.9	99.2	97.2	93.6	88.6	84.3	80.4	75.6	71.7	67.9
60	20	Student-t	100.0	99.7	96.3	90.4	84.8	79.4	74.5	69.5	67.5	64.3	62.1
		Welch	100.0	98.8	85.6	70.4	56.2	49.5	41.9	36.9	33.5	30.4	27.7
		Ayıklanmış-t	100.0	98.8	88.2	73.9	61.6	52.7	45.9	40.2	36.1	32.3	30.5
40	60	Student-t	100.0	100.0	98.5	97.2	92.4	85.5	78.8	71.6	66.1	59.7	54.6
		Welch	100.0	100.0	99.6	98.7	96.0	91.9	87.6	82.6	78.6	73.2	68.9
		Ayıklanmış-t	100.0	100.0	99.9	98.7	96.3	92.6	87.8	83.3	78.2	74.8	70.2
60	40	Student-t	100.0	100.0	99.6	97.3	93.8	88.4	83.3	78.5	74.2	70.2	67.8
		Welch	100.0	100.0	99.1	94.5	87.9	80.8	73.0	65.6	60.9	55.7	52.7
		Ayıklanmış-t	100.0	100.0	99.0	95.3	89.4	81.9	74.7	69.4	62.7	58.7	54.7

Çizelge 5. Normal dağılım gösteren iki popülasyonda $\delta=2.0$ için elde edilen ampirik testin gücü (%)

n ₁	n ₂	Uygulanan Test	$\sigma_1^2 : \sigma_2^2$										
			1:1	1:2	1:4	1:6	1:8	1:10	1:12	1:14	1:16	1:18	1:20
20	20	Student-t	100.0	99.9	97.4	90.4	82.6	74.8	67.7	61.0	55.7	51.5	47.8
		Welch	100.0	99.9	96.7	89.1	80.2	71.4	64.2	57.0	51.5	47.5	43.9
		Ayıklanmış-t	100.0	99.9	97.2	90.6	82.2	74.0	66.5	61.5	55.9	50.8	47.0
40	40	Student-t	100.0	100.0	100.0	99.6	98.7	96.4	93.4	90.0	85.7	81.7	78.5
		Welch	100.0	100.0	100.0	99.6	98.5	96.1	92.7	89.1	84.6	80.4	77.0
		Ayıklanmış-t	100.0	100.0	100.0	99.7	98.8	96.5	93.3	90.0	85.4	81.7	77.8
60	60	Student-t	100.0	100.0	100.0	100.0	100.0	99.6	98.8	97.6	96.1	94.1	91.7
		Welch	100.0	100.0	100.0	100.0	99.9	99.6	98.7	97.5	95.8	93.6	91.3
		Ayıklanmış-t	100.0	100.0	100.0	100.0	99.9	99.6	99.0	97.9	96.3	94.2	91.5
20	40	Student-t	100.0	100.0	99.6	96.9	91.5	83.9	76.2	67.5	60.7	54.6	48.7
		Welch	100.0	100.0	99.9	99.2	97.3	94.5	91.2	86.7	82.8	78.4	74.6
		Ayıklanmış-t	100.0	100.0	99.9	99.3	97.5	94.9	91.6	88.4	83.3	79.8	76.0
40	20	Student-t	100.0	100.0	99.5	97.2	93.5	89.5	85.3	81.2	76.9	73.5	69.7
		Welch	100.0	100.0	97.9	90.6	82.2	73.8	65.2	59.5	53.5	48.5	44.3
		Ayıklanmış-t	100.0	100.0	98.4	92.7	84.6	76.9	68.9	62.2	57.4	52.8	48.3
20	60	Student-t	100.0	100.0	99.9	98.5	95.1	87.8	80.7	71.3	63.6	55.8	49.5
		Welch	100.0	100.0	100.0	99.9	99.5	98.6	97.3	95.9	93.5	91.1	88.9
		Ayıklanmış-t	100.0	100.0	100.0	99.9	99.6	98.9	97.6	96.4	94.1	92.0	89.3
60	20	Student-t	100.0	100.0	99.8	98.7	96.3	93.1	90.3	87.0	83.4	81.5	78.4
		Welch	100.0	100.0	98.3	91.7	82.5	73.2	66.1	59.2	53.8	49.5	44.9
		Ayıklanmış-t	100.0	100.0	98.6	92.8	84.8	77.2	69.3	63.4	57.1	52.5	49.1
40	60	Student-t	100.0	100.0	100.0	99.9	99.6	98.6	96.8	93.7	90.4	86.5	82.7
		Welch	100.0	100.0	100.0	100.0	99.8	99.5	98.6	96.9	95.2	93.1	90.5
		Ayıklanmış-t	100.0	100.0	100.0	100.0	99.9	99.6	98.6	97.3	95.5	93.8	91.6
60	40	Student-t	100.0	100.0	100.0	99.9	99.6	98.7	97.2	94.8	92.2	89.6	86.9
		Welch	100.0	100.0	100.0	99.8	98.8	96.7	93.5	88.9	85.0	81.1	76.8
		Ayıklanmış-t	100.0	100.0	100.0	99.8	98.8	96.8	93.8	90.1	86.1	82.1	78.1

n₁=20, n₂=40 (n=60) örnek genişliği kombinasyonunda (büyük varyanslı popülasyondan alınan birey sayısı fazla olduğunda) Welch ve Ayıklanmış-t Testlerinin güç değerleri, popülasyonların varyansları arasındaki fark 20 kat oluncaya kadar Student-t Testinden daha yüksek olarak gerçekleşmiştir. n₁=40, n₂=20 örnek genişliği kombinasyonunda (küçük varyanslı popülasyondan alınan birey sayısı fazla olduğunda) ise popülasyonların varyansları arasındaki fark 18 kat artıncaya kadar Student-t Testi daha güçlü olarak bulunmuştur. Benzer sonuçlar, n₁=20, n₂=60 ve n₁=60, n₂=20 (n=80) örnek genişliği kombinasyonları için de geçerlidir. n₁=40, n₂=60 ve n₁=60, n₂=40 (n=100) örnek genişlikleri kombinasyonlarında da ise bütün test yöntemleri, popülasyonların varyansları arasında 20 kat olsa bile, %80'in üzerinde güç değerine ulaşmaktadırlar. Bunların bir sonucu olarak, popülasyon ortalamaları arasındaki fark 2.0 standart sapma olduğunda, örnek genişlikleri eşit (en az 40'ar birey) ve popülasyonların varyans oranları 20 kat oluncaya kadar Student-t, Welch ve Ayıklanmış-t Testlerinden her hangi biri ile aynı güvenilirlikte kullanılabilirken, popülasyonların varyans oranları arttıkça ve büyük varyanslı popülasyondan alınan birey sayısı küçük varyanslı popülasyondan alınan birey sayısından fazla oldukça Welch ve Ayıklanmış-t Testleri, Student-t Testinden daha güvenilirdir, küçük varyanslı popülasyondan alınan birey sayısı büyük varyanslı popülasyondan alınan birey sayısından fazla olduğunda da Student-t Testi, Welch ve Ayıklanmış-t Testlerinden daha güvenilirdir olarak kullanılabileceğini ileri sürmek mümkündür.

Sonuç

Bu çalışmada sonuç olarak, popülasyon ortalamaları arasındaki fark sıfır ve popülasyon varyansları arasında 20 kat fark (standart sapma cinsinden) olduğu veya popülasyon varyansları bilinmediği durumlarda örnek genişlikleri eşit olmak kaydıyla, ele alınan üç test yönteminin de $1-\alpha=0.95$ güven katsayısı ile kullanılabilirliği söylenebilir. Ancak örnek genişliklerinin eşit olmadığı durumlarda, Student-t testi sadece varyans oranları 1:1 olduğu durumlarda I. Tip hata olasılığını beklenen seviyede gerçekleştirebildiği halde, örnek genişlikleri farklı olduğunda, Student-t testi etkilenirken Welch ve Ayıklanmış-t testlerinin I. Tip hata olasılığını başlangıçta belirlenen %5 sınırında koruduğu görülmüştür.

Popülasyon ortalamaları arasındaki fark, $\delta=0.5$ standart sapma olduğunda, popülasyon varyansları ve örnek genişlikleri eşit olmak kaydıyla ancak n₁=n₂=60 olması halinde testlerin gücünün güvenilir güç sınırları içinde kalırken, diğer bütün kombinasyonlarda güç değerleri oldukça düşük olmuştur. Ancak popülasyon ortalamaları arasındaki farkın $\delta=0.5$ olmasının, böyle bir sonuç elde etmede rolü olduğu da göz ardı edilmemelidir.

Popülasyon ortalamaları arasındaki fark, $\delta=1.0$ standart sapma olduğunda, popülasyon varyansları eşit olduğunda, örnek genişlikleri ve bunların kombinasyonu ne olursa olsun testin gücünün her yöntemde güvenilir güç sınırları içinde kalmıştır. Ancak, popülasyon varyansları heterojenleştikçe, testlerin istenilen güç

gösterebilmesinin örnek genişlikleri ile ilişkili olduğu görülmüştür. Mesela, populasyonların varyans oranları 1/6 oluncaya kadar testin gücünün $n_1=n_2=60$ örnek genişlikleri kombinasyonunda bütün yöntemlerde kabul edilebilir seviyelerde bulunmuştur. Bu durum, populasyon varyansları arasındaki fark arttıkça ve örnek genişliklerinin de büyük varyanslı populasyon lehinde dengesiz olması halinde Welch ve Ayıklanmış-t testlerinin, Student-t testine göre daha güvenilir oldukları şeklinde yorumlanabilir.

Populasyon ortalamaları arasındaki fark, $\delta=1.5$ standart sapma ve populasyon varyansları da eşit olduğunda, örnek genişlikleri ve bunların kombinasyonu ne olursa olsun testlerin gücünün her yöntemde %95'in üzerinde olarak gerçekleşmiştir. Populasyonların varyansları heterojenleştikçe testlerin gücünün, örnek genişlikleri ile ilişkili olduğu ve örnek genişlikleri eşit ve fazla (en az $n_1=n_2=60$) olduğunda, populasyon varyansları arasında 18 kat fark olması halinde bile her üç test yönteminin de güç değerleri birbirine yakın ve yaklaşık %75 olarak gerçekleşmektedir. Örnek genişliklerinin farklı olması durumlarında da, populasyonların varyans oranları arttıkça ve büyük varyanslı populasyondan alınan birey sayısı küçük varyanslı populasyondan alınan birey sayısından fazla oldukça Welch ve Ayıklanmış-t Testleri, Student-t Testinden daha güvenilir, küçük varyanslı populasyondan alınan birey sayısı büyük varyanslı populasyondan alınan birey sayısından fazla oldukça da Student-t Testi, Welch ve Ayıklanmış-t Testlerinden daha güvenilir olarak bulunmuştur.

Populasyon ortalamaları arasındaki fark, $\delta=2.0$ standart sapma olduğunda ise, populasyon varyansları eşit, örnek genişlikleri ve bunların kombinasyonu ne olursa olsun testlerin güçleri her yöntemde %100 olarak gerçekleşmiştir. Örnek genişlikleri eşit olmak şartıyla, populasyonların varyans oranları arttıkça her üç testinde güç değerlerinde de bir azalma olmaktadır. Populasyon ortalamaları arasındaki fark 2.0 standart sapma olduğunda, örnek genişlikleri eşit (en az 40'ar birey) ve populasyonların varyans oranları 20 kat oluncaya kadar Student-t, Welch ve Ayıklanmış-t Testlerinden her hangi biri, aynı güvenilirlikte kullanılabilirken, populasyonların varyans oranları arttıkça ve büyük varyanslı populasyondan alınan birey sayısı küçük varyanslı populasyondan alınan birey sayısından fazla oldukça Welch ve Ayıklanmış-t Testleri, Student-t Testinden daha

güvenilir, küçük varyanslı populasyondan alınan birey sayısı büyük varyanslı populasyondan alınan birey sayısından fazla oldukça da Student-t Testi, Welch ve Ayıklanmış-t Testlerinden daha güvenilir olarak kullanılabilirliği sonucuna varılmıştır.

Kaynaklar

- Akdeniz, F. 1984. Olasılık ve İstatistik. Ankara Üniversitesi Fen Fakültesi Yayınları, No: 138. Ankara. 519 s.
- Anonymous, 1998. Robust Determination of Mean Trimmed Mean. Erişim Adresi: <http://osprey6.npac.syr.edu:8080/foilssets/cps713stat/node199.html>. Erişim Tarihi:29.04.1999.
- Devenport, J. M. and J. T. Webster, 1970. The Behrens-Fisher Problem, an Old Solution Revisited. *Metrika*, 22, 47-54.
- Wilcox, R. R. 1986. New Monte Carlo Results of the Robustness 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 and 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. and 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.
- Zar, J. H. 1999. *Biostatistical Analysis*. Fourth Ed. Prentice Hall Inc., Upper Saddle River, New Jersey, USA. 663 s.