

Mann-Whitney U ile Student's t testinin I.Tip Hata ve Güç bakımından Karşılaştırılması: Monte Carlo Simülasyon Çalışması

Recep Bindak

Gaziantep Üniversitesi, Gaziantep

Geliş Tarihi:07.11.2013; Kabul Tarihi:22.03.2014

Özet

Parametrik t testinin varsayımları karşılanmadığı durumda alternatif olarak Mann-Whitney U testi kullanılmaktadır. Varsayımlar yerine getirildiğinde t testinin daha güvenli olduğu bilinmektedir. Bu çalışmanın amacı kitlenin normal dağılımlı olması durumunda ve çeşitli örnek hacimlerinde MWU ile t testinin performanslarını karşılaştırmaktır. Kitleden örnek seçme için Monte Carlo simülasyonu yapıldı. Ortalamalar arası fark $\delta=0$, $\delta=0.5\sigma$, $\delta=1\sigma$ ve varyans oranları 1:1, 1:4, 1:9 olan kitlelerin ikili kombinasyonları kullanıldı. Birinci tip hata ve test gücünü hesaplamak için her bir kombinasyon için 20000 Monte Carlo denemesi yapıldı. Her bir kombinasyon ve farklı örnek hacimleri için çekilen örneklere her iki test uygulandı, red edilen sıfır hipotezi sayısı kaydedildi. Küçük örnek ($n \leq 30$) hacminde MWU testinin daha az hata verdiği, H_0 hipotezinin ret veya kabul edilmesinde MWU ile t testlerinin %97 oranında aynı sonucu verdikleri görüldü. Eşit olmayan varyans durumunda özellikle örnek hacmi arttıkça MWU testinin I.Tip hata oranını koruyamadığı görüldü.

Anahtar kelimeler

I.Tip hata; Testin gücü; Mann-Whitney U test, Students' t test; Monte Carlo simülasyonu

Comparision Mann-Whitney U Test and Students' t Test in Terms of Type I Error Rate and Test Power: A Monte Carlo Simulation Study

Abstract

The nonparametric Mann-Whitney U test is an alternative to parametric Student's t-test, when the assumptions of t-test cannot be respected. However as one understood, it is more reliable to use the t-test if its assumptions can be met. The purpose of this study is to analyze the performances of the Mann-Whitney U test and Student's t test in cases of normal populations and various sample sizes. A Monte Carlo simulation study is carried out via data which were produced from standard normal distribution. Dual combinations of populations were used, with a difference between the means of $\delta=0$, $\delta=0.5\sigma$, $\delta=1\sigma$ and, variance ratios of 1:1, 1:4, 1:9. To compute Type I error rate and test power, 20000 trial were made for each condition. For each combination, two samples with various sample sizes were randomly selected; both tests were applied, and the number of rejected null hypotheses was recorded. The MWU test gave fewer errors in small samples ($n \leq 30$). MWU and t tests gave the same result at 98% rate in rejecting or accepting the H_0 hypothesis. In case of unequal variance, as sample size increased, the type I error ratio increased, particularly for the U test; both tests gave the same result in approximately 97% of simulations.

Keywords

Type I error, Test power, Mann-Whitney U test, Students' t test, Monte Carlo Simulation

© Afyon Kocatepe Üniversitesi

Giriş

İstatistik bir çok akademik disiplinde ve yaşam alanında giderek büyüyen uygulamalarıyla önemli ve büyüleyici bir bilim dalıdır. Özellikle son yıllarda bütün dünyada önemi gittikçe daha fazla anlaşılmaya başlanan istatistik metotlarının günümüzde uygulanmadığı alan hemen hemen yok gibidir. Uygulama alanına göre biyoistatistik, biyometri, ekonometri, sosyometri, arkeometri gibi

isimler alan uygulamalı istatistik ziraat, tıp, ormancılık, zooloji, ekoloji gibi birçok disiplini kapsayan bir uygulama alanına sahiptir (Mendeş, 2012:1). Söz gelimi tıp araştırmalarında istatistik metotlarının rolünü görmek için herhangi bir tıp dergisine kısa bir göz atmak yeterlidir (Antonisamy, 2010).

Birçok istatistik metodu iki yaygın özelliğe sahiptir. Bunlardan biri temel yoğunluk fonksiyonunun

bilindiği varsayımdır, diğeri ise yoğunluk fonksiyonunun parametreleri hakkında veya parametrelerinin tahmini hakkında hipotez testleri ile ilgilenmeleridir (Hoel,1971). Hipotez testleri, kitle hakkında bazı varsayımlara dayanan istatistik tekniklere parametrik testler adı verilir. Genellikle bu varsayımlar kitle normalliği ve örneklerin kitleden rastgele çekildiği ile ilgilidir. Bu varsayımların karşılanmadığı durumlarda non-parametrik (parametrik olmayan) adı verilen tekniklerin kullanılması gerekir (Bajpai, 2010) . Parametrik olmayan metotların çoğu 1940'lı yılların sonlarında önerilmiştir. Bu tarihten sonra parametrik olmayan metotların özellikleri ile ilgili çok sayıda çalışma yapılmıştır (Gibbons, 1993;2).

Hipotez test ederken özellikle normallik varsayımının ihlal edildiği durumlarda kullanılmak üzere istatistikçiler pek az varsayım gerektiren yada hiç varsayım gerektirmeyen teknikler geliştirmişlerdir. Bu testlere parametrik olmayan testler denir. Bunlar kitlelerin sürekli oldukları dışında hemen hiçbir varsayım yapmadığımız dağılımdan bağımsız testleri, ayrıca kitlelerin parametreleri ile ilgilenmediğimiz anlamında parametrik olmayan metotları içerir (Miller ve Miller, 2006).

Parametrik metotların geçerli olması içi verilerin en az aralık düzeyinde ölçülmüş olması gerekirken parametrik olmayan istatistikler nominal veya sıralı ölçüm düzeyindeki verilere de uygulanabilmektedir. Aralık düzeyindeki ölçümler sayma veya sıralama sayılarına dönüştürülebildiğinden parametrik olmayan metotlar parametrik metotların geçerli olduğu durumlarda da kullanılabilir (bunun tersi doğru değildir). Ayrıca eğitim bilimleri, psikoloji gibi davranış bilimlerinde veriler genellikle sıralama düzeyinde ölçüldüğünden parametrik olmayan metotların kullanılması daha avantajlı olmaktadır. Özet olarak parametrik istatistiklerin uygun olmadığı oldukça geniş bir alanda parametrik olmayan istatistikler rahatlıkla kullanılabilir (Gibbons, 1993;1-2).

Örnek hacminin az olduğu ve kitle dağılımının normal olduğu durumlarda \bar{X} istatistiği ile ilgili olarak WS Gosset tarafından geliştirilen t dağılımından yararlanılmaktadır (Miller ve Miller,

2006). Uygulamada en çok kullanılan testlerden biri t-testidir.

Küçük örneklem durumunda örnek standart sapması (S), kitle standart sapması (σ) nın iyi bir tahmini olmaz. Bu nedenle test istatistiğinde kitle standart sapması yerine örnek standart sapması kullanmak tercih edilir. İşte bu Student t dağılımı olarak bilinir. Tipik student t denklemi

$$T = \frac{X\sqrt{v}}{v}$$

biçimindedir. Burada X standart normal değişken, V^2 ise v serbestlik dereceli ve X dan bağımsız χ^2 dağılımlı bir değişkendir (Hoel, 1984: 258).

t dağılımının önemli bir özelliği de bilinmeyen kitle parametrelerine bağlı olmamasıdır. t dağılımı iki ortalama arasındaki farkın önem testinde kullanılabilir. X_1 ve X_2 ortalamaları μ_1, μ_2 varyansları aynı (σ^2) olan normal dağılımlı iki kitle olsun. Örnek hacimleri sırasıyla n_1 ve n_2 olan örnekler bu kitlelerden rastgele çekilmiş olsun. Örneklem ortalama ve standart sapmaları \bar{X}_1, \bar{X}_2 ve s_1, s_2 olsun. Kitle varyansının bilinmediği durumda,

$$X = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

olmak üzere,

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \cdot \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

değişkeni $u = n_1 + n_2 - 2$ serbestlik dereceli Student t dağılımına uyar (Hoel, 1984:263). Örnek varyanslarının eşit olmaması durumunda T değişkeni,

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

biçiminde hesaplanır (Hoel, 1984:263-265).

Bağımsız iki grup ortalamasının t-testi ile karşılaştırılabilmesi için grupların alınmış oldukları kitle dağılımlarının normal olması, grup varyanslarının homojen olması ve gözlemlerin bağımsız olması varsayımlarının yerine gelmesi gerekir (Mendeş, 2012:155)

Ancak dağılımın normal olmadığı veya dağılımın şeklinin bilinmediği durumda iki örneklemlerli t testine karşılık parametrik olmayan Mann Whitney U testi geliştirilmiştir. Mann Whitney U [MWU] testi örneklemin çekildiği kitlelerin normal dağılımlara uyduğu varsayımını yapmak zorunda kalmadan “iki bağımsız örneklemin aynı kitlelerden geldiği” sıfır hipotezini test etmede kullanılmaktadır (Miller ve Miller, 2006).

MWU, davranış bilimlerinde en sık kullanılan testlerden biridir. Diğer bir çok non-parametrik testler gibi MWU testi de sıfır hipotezi altında karşılaştırılacak grupların geldikleri kitle varyanslarının eşit olduğu varsayımına dayanır ve rank hesabına dayanır (Kasuya, 2001).

MWU testi için test istatistiği hesaplanırken karşılaştırılacak örneklem değerleri bir arada büyüklük sırasına konular ve rank numarası verilir. İki örneklemin ortalamaları arasında kayda değer bir fark varsa küçük rankların çoğunun bir örnekleme; büyük rankların çoğunun da öteki örnekleme gitmesi beklenir. n_1 hacimli ilk örneklemin rankları toplamı R_1 , n_2 hacimli ikinci örneklemin rankları toplamı R_2 ve

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2} \text{ ve } U_2 = R_2 - \frac{n_2(n_2+1)}{2}$$

olmak üzere test istatistiği $U = \min\{U_1, U_2\}$ alınır (Miller ve Miller, 2006). Eğer sıfır hipotezi doğru ise U değerinin ne çok küçük ne de çok büyük olmasını bekleriz (Kreyszig, 1970:373).

Dikkat edilirse $U_1, U_2 = n_1 n_2$ ve U değerleri 0 ile $n_1 n_2$ arasında yer alır. Bu rastgele değişkenler $\frac{n_1 n_2}{2}$ nin iki yanında simetrik birbirinin aynı dağılımlara uyarlar. $n_1, n_2 \geq 8$ için U rastgele değişkeni, ortalaması ve varyansı sırasıyla,

$$\mu = \frac{n_1 n_2}{2} \text{ ve } \sigma^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

olan normal dağılıma yaklaşır. (Miller ve Miller, 2006).

Bir hipotez testi sonucunda sıfır hipotezi ret veya kabul edilir. Gerçekte doğru iken sıfır hipotezinin ret edilmesi test için I.Tip Hata olarak adlandırılır. I.Tip hata aynı zamanda testin anlamlılık düzeyini ifade eder. Gerçekte yanlış olan sıfır hipotezinin

kabul edilmesi ise II.Tip hata olarak bilinmektedir. I ve II.tip hata yapma olasılığı sırasıyla α ve β ile sembolize edilir (Hoel, 1984). Genellikle α ve β birbirine göre ters yönde artar veya azalır. Testin gücü $(1 - \beta)$

Tablo 1. Sıfır Hipotezi hakkında karar ve I.Tip ile II.Tip hatalar

		Sıfır Hipotezi (H_0)	
		Gerçekte Doğru	Gerçekte Yanlış
Karar	H_0 Kabul	doğru karar	II.Tip Hata
	H_0 Ret	I.Tip Hata	doğru karar

ile ifade edilir ve gerçekte yanlış olan sıfır hipotezinin red edilme olasılığının bir ölçüsüdür (Bajpai, 2010).

Student t testi ile Mann Whitney U testinin çeşitli özellikleri bakımından performanslarının ele alındığı simülasyon çalışmalarına rastlamak mümkündür (Munro, 2005; Zimmerman & Zumbo, 1990; Zimmerman & Zumbo, 1993; Zimmerman, 2004; Nachar, 2004; Zimmerman, 2006; Akkartal, Mendes, and Mendes, 2010; Nthangeni and Algina, 2001). Bu çalışmada parametrik olmayan MWU testi ile parametrik t testinin performansları karşılaştırılmıştır. Ayrıca bu çalışmada literatürdeki simülasyon çalışmalarında yer almayan testlerin sıfır hipotezini aynı anda ret-kabul oranlarının belirlenmesi amaçlanmış, bu oranlar ile parametrik olmayan MWU testinin parametrik t - testine göre performansı incelenmiştir. Bu amaçla Monte Carlo simülasyonu kullanılarak MWU testi ile iki örneklem t testi çeşitli örnek hacimlerinde I.tip hata oranı ve testin gücü bakımından karşılaştırılmıştır.

2. Materyal ve Metot

Çalışmanın materyalini Microsoft Excel ile oluşturulan standart normal ($N(0,1)$) dağılımlı 500 birimlik rastgele sayılar oluşturmuştur. Çalışmada, I. Tip hata olasılığı simülasyon denemeleri başında $\alpha=0.05$ olarak kararlaştırılmıştır.

Araştırmada, “ H_0 : Örneklem aynı kitleden çekilmiştir” sıfır hipotezi test edilmiştir. Örneklem aynı ortalamalı kitlelerden çekilmesiyle sıfır hipotezinin red edildiği durumda I.tip hata olduğu kabul edilmiştir. Örneklem

ortalamaları farklı kitlelerden çekilmesi durumunda ret edilen sıfır hipotezi oranı testin gücü olarak değerlendirilmiştir. Heterojen varyans oluşturmak için ilk kitle birimleri istenen oranın karekökü ile çarpılmak suretiyle ikinci kitle oluşturulmuş ve örnekler bu iki kitleden çekilmiştir.

MWU testi ile t testinin I. Tip hata oranları ve testin gücünün belirlenmesi için ortalamaları arasındaki fark $\delta=0$, $\delta=0.5\sigma$, $\delta=1\sigma$ ve varyans oranları 1:1, 1:4, 1:9, 4:1, 9:1 olan kitlelerin ikili kombinasyonu kullanılmıştır. Her bir örnek hacminde simülasyon ile 20000'er defa denenmiştir. Her bir kombinasyon için farklı örnek hacimleri için rastgele iki örnek seçilerek her iki test uygulanmış ve ret edilen sıfır hipotezi sayısı kaydedilmiştir. Ayrıca aynı anda her iki test tarafından da ret edilen ve kabul edilen sıfır hipotezi sayısı belirlenmiş, böylece MWU ve t testlerinin benzer sonuç üretme oranı tespit edilmiştir. Bütün hesaplamalar Microsoft Excel ile yapılmıştır. Monte Carlo simülasyon algoritması aşağıdaki gibi uygulanmıştır:

1-N=500 birimli ve $N(0,1)$ dağılımlı kitle oluşturuldu. I. tip hata için aynı kitle kullanıldı, farklı varyans için aynı ortalamaya sahip ancak varyansı farklı kitle oluşturuldu. Testin gücü için ikinci kitle olarak ilk kitleye standart sapma cinsinden sabit sayı eklenerek oluşturulan ikinci kitle kullanıldı. Birimin büyüklük sıra numarası birim numarası olarak verildi.

2-Örnek hacmi n_1, n_2 seçildi

Tablo 2. Örnek hacimlerinin eşit ve grup varyanslarının farklı olması durumunda gerçekleşen I. tip hata oranları

Varyans oranı	$\sigma_1^2 : \sigma_2^2 = 1:1$			$\sigma_1^2 : \sigma_2^2 = 1:4$			$\sigma_1^2 : \sigma_2^2 = 1:9$		
	Örnek Hacmi n_1, n_2	MWU	t-test	Aynı sonuç %	MWU	t-test	Aynı sonuç %	MWU	t-test
8,8	.0490	.0498	98.22	.0579	.0510	97,54	.06705	.05195	96.83
10,10	.0506	.0507	98.05	.0504	.0496	98,08	.07080	.05005	95.98
15,15	.0486	.0510	98.28	.0597	.0498	96,99	.07130	.05025	95.78
20,20	.0474	.0499	98.16	.0582	.0503	96,99	.07070	.05055	95.66
30,30	.0496	.0505	98.14	.0591	.0474	96,64	.07190	.04940	95.47
40,40	.0509	.0506	98.11	.0587	.0490	96,63	.07180	.04980	95.64
50,50	.0499	.0498	98.36	.0584	.0469	96,43	.07300	.05220	95.1
60,60	.0492	.0499	98.19	.0608	.0501	96,4	.07525	.05165	95.09

Popülasyon varyansları arasında 1:4 oranı olduğunda t testinin I. tip hata oranı %5 civarında

3-Simülasyon ile örnek hacmi kadar birim numarası üretildi, karşılık gelen birim örneğe alındı.

4-Aynı örneklere MWU ve t testi uygulandı sonuçlar kaydedildi.

5-Üçüncü ve dördüncü adımlar 20000 defa tekrarlandı.

6-İkinci adımdaki örnek hacimleri değiştirilerek işlemler tekrar edildi.

BULGULAR

2.1.1. Tip Hata ile ilgili bulgular:

Örnek hacimlerinin eşit olduğu $n_1=n_2$; 8,10,15,2,30,40,50,60 ve kitle varyansları arasında 1:1, 1:4, 1:9 oranları olduğunda gerçekleşen I. tip hata oranları ile her bir durumda testlerin aynı sonucu verme oranları Tablo 2'de verilmektedir. Eşit varyans durumunda; eşit örnek hacmi ve eşit varyans durumunda hem U hem t testleri araştırmanın başlangıcında kararlaştırılan I. tip hata oranlarını korumuşlardır. Küçük örneklerde ($n \leq 30$) U testi daha az hata vermiştir. U testinin küçük örnek hacimlerinde biraz konservatif (tutucu) olduğu söylenebilir. Bütün örnek hacimlerinde H_0 hipotezini red veya kabul etmede U ve t testleri ortalama %98'in üstünde aynı sonucu vermiştir.

kalmıştır. Ancak U testi için hata oranının Örnek hacmi arttıkça 0.05 oranını biraz aştığı görülmüştür.

Eşit örnek hacmi ve 1:4 varyans oranlarında her iki testin sıfır hipotezi hakkında aynı sonucu verme oranı %96-97 civarında gerçekleşmiştir. Varyans oranı 1:9 iken örnek hacmi arttıkça her iki testin I.tip hata oranlarının arttığı, U testi için bu oranın %7'yi aştığı görülmüştür. Bu durumda t testinin başlangıçta karşılaştırılan hata oranını küçük bir miktar aştığı söylenebilir. Varyans oranları büyüdüğünde H_0 hipotezini ret veya kabul etmede U ve t testleri daha az simülasyon denemelerinde (%95 civarında) aynı sonucu vermişlerdir (bakınız Tablo 2).

İki örnek hacminin farklı olduğu eşit olmayan varyans durumunda; genel olarak t testinin I.tip hata oranını koruduğu görülmektedir. Bu durumda U testinin I. tip hata oranı farklı davranış göstermektedir. Şöyle ki, örnek hacmi ile varyans dengesi aynı yönlü iken (büyük örnek büyük varyans) testin oldukça konservatif (tutucu) olduğu yani I. tip hata oranının %5'in oldukça altında gerçekleştiği anlaşılmaktadır. Örnek hacmi ile varyans dengesi ters yönlü iken (küçük örnek büyük varyans) ise I. tip hata oranının %5'in oldukça üzerine çıktığı anlaşılmaktadır. Örneğin örnek hacmi 30,50 ve varyans oranı 9:1 iken gerçekleşen I. tip hata oranı %10'a ulaşmıştır.

Tablo 3. Örnek hacimlerinin farklı olduğu ve varyans oranının dengesiz olduğu durumda testlerin gerçekleşen I.tip hata oranları

Varyans Oranı	$\sigma_1^2: \sigma_2^2 = 1:4$			$\sigma_1^2: \sigma_2^2 = 1:9$			$\sigma_1^2: \sigma_2^2 = 4:1$			$\sigma_1^2: \sigma_2^2 = 9:1$		
	Örnek hacmi n_1, n_2	MWU	t-test	aynı %	MWU	t-test	Aynı sonuç %	MWU	t-test	Aynı sonuç %	MWU	t-test
5,10	.0251	.0489	96.97	.0227	.0498	96.6	.0712	.0583	97.53	.0829	.0583	97.01
8,12	.0418	.0517	97.26	.0420	.0497	96.83	.0719	.0517	97.14	.0874	.0521	95.69
10,15	.0429	.0526	97.10	.0442	.0508	96.66	.0738	.0526	96.83	.0897	.0509	95.10
20,30	.0447	.0518	97.08	.0475	.0514	96.29	.0760	.0516	96.14	.0965	.0511	94.25
30,50	.0398	.0488	97.24	.0395	.0479	96.65	.0782	.0497	95.60	.1015	.0499	93.40
40,60	.0437	.0506	96.87	.0502	.0495	96.36	.0760	.0489	95.65	.1032	.0507	93.22

2.2. Testin Gücü ile ilgili bulgular:

Testlerin yanlış olan sıfır hipotezini red etme oranları ile aynı sonucu üretme oranları Tablo 4'te verilmektedir. Eşit örnek hacmi ve eşit varyans durumunda t testinin biraz daha güçlü olduğu anlaşılmaktadır. Etki büyüklüğü $\delta=0.5\sigma$ iken $n_1=n_2=30$ örnek hacmine kadar her iki test için güç

%50'nin altındadır. Bu durumda ancak büyük örnek hacminde ($n>60$) test güçlerinin %80'i aştığı görülmüştür. Öte yandan etki büyüklüğü $\delta=1\sigma$ iken her iki testin de güçlerinin $n>15$ için %80'e ulaştığı görülmüştür. Bir diğer dikkat çekici bulgu test güçlerinin düşük olduğu örnek hacimlerde bile iki testin aynı sonucu üretme oranı %95 civarında gerçekleşmiş olmasıdır.

Tablo 4. Testin Gücü ile ilgili bulgular

$\Delta=\mu_1-\mu_2$	$\Delta=0.5\sigma$			$\Delta=1\sigma$		
Örnek Hacmi n_1, n_2	MWU %	t-test %	Aynı sonuç %	MWU %	t-test %	Aynı sonuç %
8,8	14.245	14.965	95.85	42.455	44.92	92.58
10,10	17.485	17.785	95.62	53.345	55.71	92.75
15,15	24.525	25.725	94.84	71.72	74.89	94.12
20,20	30.970	33.370	94.16	83.675	86.83	95.45
30,30	43.915	47.075	93.49	94.90	96.43	98.12
40,40	56.077	59.607	93.16	98.765	99.21	99.42
50,50	65.355	69.070	93.72	99.74	99.88	99.83

60,60	73.875	77.500	94.58	99.93	99.98	99.92
80,80	85.320	88.105	96.09	99.99	100	99.99
100,100	91.840	93.755	97.32	99.99	100	100

4. Tartışma ve Sonuç

Bu çalışmada kitle dağılımının normal olması durumunda *MWU* ve *t* testlerinin I.tip hata ve testin gücü bakımından performansları karşılaştırılmıştır. Ayrıca *MWU* ve *t* testlerinin sıfır hipotezini kabul veya ret etme bakımından benzer sonuç üretme oranı belirlenmiştir. Sonuçta *t* testinin 0.05 düzeyinde I.tip hata oranını koruduğu, *MWU* testinin ise örnek hacminin artmasına bağlı olarak ve özellikle heterojen örnek varyansı olması durumunda I.tip hata oranını bir miktar aştığı görülmüştür. Testin gücü bakımından yine *t* testinin daha iyi performans gösterdiği elde edilmiştir. Ancak *MWU* için kitle varyanslarına bakılmaksızın tek tür istatistik hesaplanmış, buna karşılık t-testi için biri homojen varyans diğeri homojen olmayan varyans için olmak üzere iki farklı istatistik ve serbestlik derecesi kullanıldığı unutulmamalıdır. Ayrıca kitlelerin normal dağılımlı olduğu göz önüne alındığında bunun beklenen bir sonuç olduğu söylenebilir.

Bu çalışmada Monte Carlo simülasyonu sonucu elde edilen sonuçların literatür bulguları ile paralellik gösterdiği anlaşılmıştır. Çünkü, hem parametrik hem de parametrik olmayan testlerin eşit olmayan varyansa (heteroscedasticity) karşı duyarlı olduğu bilinmektedir (Leys & Schumann, 2010). Varyansların homojenliği Student's t testi için gerekli bir varsayım olup bu varsayım yerine gelmediğinde planlanan I.tip hata oranından sapmalar meydana gelmektedir (Bradstreet, 1996). Ayrıca Mann-Whitney U testinin özellikle eşit olmayan örnek varyansı durumunda ve örnek hacimlerinin eşit olmaması durumunda I.tip hata oranının aşırı yükseldiği rapor edilmiştir (Zimmerman, 2006; Kasuya, 2001).

Literatürde *t* ve *F* testlerinin yerine ancak aşağıdaki şartlarda parametrik olmayan alternatiflerinin kullanılması önerilmektedir (Zimmerman, Zumbo, 1993; Antonisamy et al, 2010) : 1-normallik varsayımı ihlal edildiğinde 2-varyansların

homojenliği ihlal edildiğinde 3-ölçme aracının aralık düzeyinde değil de sıralı (ordinal) olması durumunda.

Ne kadar ciddi ihlalin parametrik test kullanımını gerektireceği araştırmacılar arasında görüş birliğine varılmayan bir konudur. Birçok yazar (araştırmacı) parametrik testlerin robust olduğunu ve çok şiddetli ilk iki ihlal olmadıkça alternatiflerin kullanılmasını önermemektedirler. Genellikle varsayımların sağlandığı durumlarda parametrik testlerin tercih edilmesi gerektiği çünkü parametrik testlerin parametrik olmayan muadillerine göre daha güçlü oldukları kabul edilmektedir (Zimmerman & Zumbo, 1993: 481).

Bu çalışmada elde edilen, testlerin sıfır hipotezini aynı anda ret-kabul oranlarına ilişkin bulguları incelendiğinde *MWU* testinin *t* testine göre performansı hakkında bir fikir edinilebilir. Şöyle ki, simülasyon sonuçlarından *t* testinin planlanan (nominal) I.tip hata oranını genelde koruduğu ve U testine göre bir miktar daha güçlü olduğu görülmektedir. Bu sonuçla birlikte testlerin aynı sonucu üretme oranları *MWU* testinin *t* testine göre performansı olarak okunabilir.

Bu çalışmada örneklerin çekildiği kitlelerin normal dağılımlı olduğu durumda U ve *t* testlerinin I.Tip hata ve testin gücü karşılaştırılmıştır. Ayrıca *MWU* ve *t* testlerinin sıfır hipotezini kabul veya ret etme bakımından benzer sonuç verme oranları belirlenmiştir. Kitlenin normal dağılımlı olmadığı durumlarda ve farklı örnek hacimlerinde söz konusu testlerin performansları incelenmeye değer bir konu olarak karşımıza çıkmaktadır. Ayrıca *MWU* testi için heterojen varyans durumunda I.tip hata oranını nominal düzeye düşürmek için düzeltme faktörü çalışmaları yapılması önerilebilir.

Kaynaklar

Akkartal, E., Mendes, M. and Mendes, E., 2010. Determination suitable permutation numbers in comparing independent group means: a monte carlo

- simulation study. *Journal of Science & Industrial Research*, **69**, 422–425.
- Antonisamy, B., Christopher, S. and Samuel, P.P., 2010. *Biostatistics: Principles and Practice*. New Delhi, India: Tata McGraw Hill, pp 1–3.
- Bajpai, N., 2010. *Business Statistics*. New Delhi: Dorling Kindersley.
- Bradstreet, T.E., 1996. A Monte Carlo study of Type I error rates for the two-sample Behrens-Fisher problem with and without rank transformation. *Computational Statistics & Data Analysis*, **25**, 167–179.
- S. Chakraborti, S., Michaelson, G. And Mccracken, A.K., 2012. Shortest Expected Length Confidence Interval for the Power of the *t*-Test. *Communications in Statistics—Simulation and Computation*, **41**: 1336–1345, 2012. DOI: 10.1080/03610918.2011.600502
- Gibbons, J.D., 1993. *Nonparametric Statistics An Introduction*. USA: Sage University Paper, pp 1–2.
- Gibbons, J.D. and Chakraborti, S., 2003. *Nonparametric Statistical Inference (4th ed.)*. New York: Marcel Dekker Inc, pp 3.
- Hoel, P.G., 1971. *Introduction to Mathematical Statistics*. 4th ed. USA: John Wiley & Sons Inc, pp 258-263.
- Kasuya E., 2001. Mann–Whitney *U* test when variances are unequal. *Animal Behaviour*, **61**, 1247–1249. DOI:10.1006/anbe.2001.1691
- Kreyszig, E., 1970. *Introductory Mathematical Statistics*. USA: John Willey&Sons, Inc, pp 373.
- Leys, C. and Schumann, S., 2010. A nonparametric method to analyze ineration. *Journal of Experimental Social Psychology*, **46**, 684–688. DOI:10.1016/j.jesp.2010.02.007
- Mendeş, M., 2012. *Uygulamalı Bilimler İçin İstatistik ve Araştırma Yöntemleri*. İstanbul: Kriter, p. 155.
- Miller, I. ve Miller, M., 2006. *John Freund Matematiksel İstatistik*. Altıncı baskıdan çev. Ed.: Ü.Şenesen. İstanbul: Literatür, pp 551-552.
- Montgomery, D.C. and Runger, G.C., 2011. *Applied Statistics and Probability for Engineer (Fifth Ed.)*. USA: John Wiley Sons Inc, pp 1.
- Munro, B.H., 2005. *Statistical Methods for Health Care Research*. Fifth Edition. USA: Lippincott Williams & Wilkins, pp 87-88.
- Nachar, N., 2004. The Mann Whitney U: A test for assesing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology*, **4**: 13–20.
- Nthangeni, M and Algini, J., 2001. Type I error rate and power of some alternative methods to the independent samples *t* test. *Educational and Psychological Measurement*, **61(6)**, 937-957.
- Pett, M.A., 1997. *Nonparametric Statistics for Health Care Research*. California: Sage Publications, Inc, pp 169-171.
- Zimmerman, D.W., 2004. Inflation of Type I error rates by unequal variances associated with parametric, nonparametric and rank transformation tests. *Psicologica*, **25**, 103-133.
- Zimmerman, D.W., 2006. Two separate effects of variance heterogeneity on the validity and power of significance tests of location. *Statistical Methodology*, **3**, 351–374. DOI:10.1016/j.stamet.2005.10.002
- Zimmerman, D.W. and Zumbo, B.D., 1990. The relative power of the wilcoxon Mann Whitney test and Student *t* test under simple bounded transformations. *The Journal of General Psychology*, **117**, 425–436.
- Zimmerman, D.W. and Zumbo, B.D., 1993. The relative power of parametric and nonparametric statistical methods (In: Keren G, Lewis C, editors). *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues*. New Jersey: Lawrence Erlbaum, Hillsdale, pp 481.