



## Biyostatistik Çalışmalarında Kullanılan Küçük Örneklerde Mann-Whitney U Testi ve Bağımsız Örneklem T (Student's-t Independent Test) Testinin Güç Yönünden Karşılaştırılması

Ömer ELTAS<sup>1a</sup>✉

1. Atatürk Üniversitesi, Veteriner Fakültesi, Biyometri Anabilim Dalı, Erzurum, TÜRKİYE.  
ORCID: 0000-0002-7415-3371<sup>a</sup>

Geliş Tarihi/Received	Kabul Tarihi/Accepted	Yayın Tarihi/Published
08.02.2021	10.03.2021	26.04.2021

**Bu makaleye atıfta bulunmak için/To cite this article:**

Eltaş Ö: Biyoistatistik Çalışmalarında Kullanılan Küçük Örneklerde Mann-Whitney U Testi ve Bağımsız Örneklem T (Student's-t Independent Test) Testinin Güç Yönünden Karşılaştırılması. Atatürk Üniversitesi Vet. Bil. Derg., 16(1): 88-94, 2021. DOI: 10.17094/ataunivbd.876777

**Öz:** Bu çalışmada örnek hacminin 3, 4 ve 5 olduğu durumlarda parametrik testlerden birisi olan Bağımsız Örneklem t testi ile bu testin nonparametrik karşılığı olan Mann-Whitney U testinin güç yönünden karşılaştırılması yapılarak, küçük örnek hacimlerinde hangi testin daha güçlü olduğu belirlenmeye çalışılmıştır. Materyal olarak Microsoft Excel programında oluşturulan rasgele sayılar kullanılmıştır. Karşılaştırmalar için 7 grup oluşturulmuş ve standart normal dağılımlı  $N(0,1)$  olan birinci grupta diğer gruplar karşılaştırılmıştır. Testin gücünü belirlemek için Monte Carlo simülasyonu kullanılmış, her örnek büyüklüğü ve her grup için 10000 tekrar yapılmıştır. Varyansların hem homojen hem de heterojen olduğu gruplarda yapılan karşılaştırmalarda, testin gücünü belirlemek için ret edilen sıfır hipotezinin oranı tespit edilmiştir. Elde edilen sonuçlara göre parametrik testlerin bir varsayımı olan varyansların homojenliği sağlanmadığı durumda hem parametrik olan t testinin hem de parametrik olmayan MWU testinin güçleri oldukça düşmüştür. Sonuç olarak, bilimsel çalışmalarda tercih edilen güç seviyesine (% 80) ulaşmak için aşırı gözlemlerin bulunmadığı, varyansların homojen olduğu, etki büyüklüğünün Cohen  $d = 2$  ve örnek büyüklüğünün  $n_1=n_2=5$  olduğu durumda t testinin uygulanabileceği ortaya çıkmıştır. Bu çalışmada incelenen diğer durumlarda hem t testinin hem de MWU testinin güçlerinin % 80'in altında kaldığı tespit edilmiştir.

**Anahtar Kelimeler:** Bağımsız örneklem t testi, Mann-Whitney U, Monte Carlo simülasyonu, Testin gücü.

### Comparison of the Mann-Whitney U test and Independent Samples t-test (Independent Student's t-test) in terms of Power in Small Samples Used in Bioistatistic Studies

**Abstract:** In this study, when the sample size is 3, 4, and 5, it was tried to determine which test was stronger in small sample volumes by comparing the Independent Sample t test, which is one of the parametric tests, and the nonparametric equivalent of this test, the Mann-Whitney U test. Random numbers generated in Microsoft Excel program were used as material. 7 groups were formed for comparisons and the first group with standard normal distribution  $N(0,1)$  was compared with the other groups. Monte Carlo simulation was used to determine the strength of the test; 10000 repetitions were performed for each sample size and each group. In the comparisons made in groups where variances are both homogeneous and heterogeneous, the ratio of the rejected null hypothesis was determined to determine the strength of the test. According to the results, if the homogeneity of the variances, which is an assumption of the parametric tests, is not provided, the power of both the parametric t-test and the non-parametric MWU test decreased considerably. As a result, it has been revealed that there are no excessive observations in scientific studies to reach the preferred power level (80%), the variances are homogeneous, the effect size is Cohen  $d = 2$  and the sample size is  $n_1 = n_2 = 5$ , the t-test can be applied. In the other cases examined in this study, it was determined that the powers of both the t-test and the MWU test remained below 80%.

**Keywords:** Independent sample t-test, Mann-Whitney U, Monte Carlo simulation, Power of the test.

✉ Ömer Eltaş

Atatürk Üniversitesi, Veteriner Fakültesi, Biyometri Anabilim Dalı, Erzurum, TÜRKİYE.  
e-posta: omer.eltas@atauni.edu.tr

## GİRİŞ

İstatistik birçok yaşam alanında ve akademik disiplinde giderek önemi artan bir bilim dalıdır. Bu bilim dalının önemi özellikle son yıllarda daha fazla anlaşılmaya başlamıştır. İstatistik, popülasyon verilerinin elde edilmesi zor veya imkânsız olduğunda, bu popülasyondan örnekler alınarak popülasyon hakkında tahmin yapmaktır. Bu tanımdan da anlaşılacağı üzere örnek büyüklüğü popülasyon büyüklüğüne ne kadar yaklaşırsa, yapılan tahmin o kadar isabetli ve güçlü olmaktadır. Biyoistatistikte bazı alanlarda yeterli örnek büyüklüğüne ulaşmak kolay olurken, tıp ve veterinerlik gibi bazı alanlarda yeterli sayıda örnek bulmak bazen mümkün olmayabilir. Bazen de mümkün olsa bile etik sorunlara yol açabilir. Özellikle hayvanların kullanıldığı bir araştırmada olması gerekenden daha az sayıda hayvan kullanılması, popülasyonda mevcut olsa bile herhangi bir anlamlı farkın gözden kaçmasına, daha fazla sayıda hayvan kullanılması ise etik sorunlara ve kaynakların gereksiz yere israfına yol açabilir. Özellikle biyomedikal deneylerde kullanılan hayvan sayısını azaltmak hem etik nedenler hem de hayvan kullanılarak yapılan araştırmaların zaman alıcı ve pahalı olması sebebiyle önemlidir (1). Çünkü bütçelerin ve kaynakların daraldığı zamanlarda, uygun örnek boyutlarının uygun şekilde belirlenmesi çok önemlidir (2).

Örnek büyüklüğünün küçük olması durumunda örnek standart sapması (S), popülasyon standart sapmasının ( $\sigma$ ) iyi bir tahmini olmamaktadır. Bu sebeple test istatistiğinde popülasyon standart sapması yerine örnek standart sapmasının kullanılması tercih edilmektedir. Bu da Student t dağılımı (t dağılımı) olarak bilinir (3). Bilinmeyen popülasyon parametrelerine bağlı olmaması, t dağılımının önemli bir özelliğidir. İki ortalama arasındaki farkın önem testinde t dağılımı kullanılmaktadır. Örneklem ortalama ve standart sapmaları  $\bar{X}_1$ ,  $\bar{X}_2$  ve  $S_1$ ,  $S_2$  olan iki grubu karşılaştırırken, varyansların homojen olduğu durumda ortalamalar arasındaki farkın varyansı, aşağıdaki formülle gösterilmektedir:

$$S_0^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

olmak üzere,

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{S_0 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

değişkeni ( $n_1+n_2-2$ ) serbestlik dereceli t dağılımına uymaktadır (4). Örnek varyanslarının homojen olmaması durumunda t değişkeni, aşağıdaki formül yardımı ile hesaplanmaktadır (4):

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{S_0 \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

T testinin uygulanabilmesi için grup varyanslarının homojen olması, gözlemlerin bağımsız olması ve grupların alınmış oldukları popülasyon dağılımlarının normal olması varsayımlarının yerine gelmesi gerekir (5).

Hipotez testlerinde özellikle normallik varsayımının sağlanmadığı durumlarda kullanılması için çok az varsayım gerektiren ya da hiçbir varsayım gerektirmeyen testler geliştirilmiştir. Bu testler nonparametrik testlerdir. Bu testler popülasyonların sürekli oldukları dışında neredeyse hiçbir varsayım gerektirmeyen, ayrıca popülasyon parametreleri ile ilgilenmediğimiz manasında nonparametrik metotları içerir (6).

Dağılımın şeklinin bilinmediği veya dağılımın normal olmadığı durumda iki örneklem t testi yerine kullanılacak olan nonparametrik bir test olan Mann Whitney U (MWU) testi geliştirilmiştir. "İki bağımsız örneklemin aynı popülasyondan geldiği" sıfır hipotezini test etmek için kullanılan MWU testi, örneklemin çekildiği popülasyonların normal dağılıma uyduğu varsayımını gerektirmemektedir (6).

MWU testine ait test istatistiği hesaplanırken karşılaştırılacak her iki grubun örneklem değerleri birlikte büyüklük sırasına göre dizilir ve rank numarası verilir.  $n_1$  hacimli ilk örneklemin rankları toplamı  $R_1$ ,  $n_2$  hacimli ikinci örneklemin rankları toplamı  $R_2$  olmak üzere,

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

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

test istatistiği  $U = \min\{U_1, U_2\}$  alınır. Eğer hesaplanan en küçük U değeri Tablo U değerinden büyük ise sıfır hipotezi kabul edilir (7).

Bir hipotez testi yapılırken anlamlılık düzeyi, etki büyüklüğü, örneklem hacmi ve istatistiksel güç dikkat edilmesi gereken önemli faktörlerdir. Bir istatistiksel anlamlılık testinin sıfır hipotezini reddedebilme olasılığı olan istatistiksel güç, anlamlılık seviyesi, etki büyüklüğü ve örneklem hacminin doğrusal bir fonksiyonudur (8).

İstatistiki olarak etki büyüklüğü; iki değişken arasındaki istatistiksel ilişkinin gücünün ölçüsüdür ve sıfır hipotezinin yanlış olmasının derecesidir (8). Etki büyüklüğü için kullanılan farklı ölçüler bulunmakla birlikte deney grupları arasındaki ortalamalardaki farklılıkları karakterize etmek için en yaygın olarak kullanılan Cohen-d istatistiğidir (9). İki grup arasındaki etki büyüklüğünün hesaplanması için gerekli formül:

$$d = \frac{m_1 - m_2}{S_p}$$

şekindedir (10). Burada,

d : Cohen's d sini yani bu durumda etki büyüklüğü değerini,

$M_1$  : 1. grubun ortalaması,

$M_2$  : 2. grubun ortalaması,

$S_p$  : Harmanlanmış (pooled) standart sapmayı ifade eder. Varyansların eşit olduğu durumda  $S_p = \sigma$ , varyansların eşit olmadığı durumda ise;

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2}}$$

şeklinde olup burada,

$n_1$  : 1. grubun örnek sayısı,

$n_2$  : 2. grubun örnek sayısı,

$S_1$  : 1. grubun standart sapması,

$S_2$  : 2. grubun standart sapmasını ifade etmektedir.

Ayrıca Cohen'e göre etki büyüklüğü (effect size) sıfır "0" ise, deney ile kontrol grubu arasında herhangi bir farkın olmadığı, etki büyüklüğü negatif (-) ise, durumun kontrol grubu lehine olduğu, etki

büyüklüğü pozitif (+) ise, durumun deney grubu lehine olduğu söylenebilir (11).

Hipotez testlerinde sıfır hipotezi ya kabul edilir ya da ret edilir. Gerçekte doğru olan sıfır hipotezinin ret edilmesi Tip I Hata olarak adlandırılır ve  $\alpha$  ile sembolize edilir. Tip I hata aynı zamanda testin anlamlılık seviyesini ifade eder. Gerçekte yanlış olan sıfır hipotezinin kabul edilmesi ise Tip II hata olarak adlandırılır ve  $\beta$  ile sembolize edilir. Genellikle  $\alpha$  ve  $\beta$  birbirine göre ters yönde artar veya azalır. Testin gücü ( $1 - \beta$ ) ile gösterilir ve gerçekte yanlış olan sıfır hipotezinin ret edilme olasılığının bir ölçüsüdür (12). Bilimsel çalışmalarda genellikle testin gücü 0.80 olarak kabul edilir (10, 13, 14).

**Tablo 1.** Sıfır hipotezi hakkında karar ve I.tip ile II.tip hatalar.

**Table 1.** Decision on zero hypothesis and type I and type II errors.

		Gerçek Durum	
		$H_0$ Doğru	$H_0$ Yanlış
Test Sonucu	$H_0$ Kabul	Testin Güvenilirliği ( $1 - \alpha$ )	Tip I Hata ( $\alpha$ )
	$H_0$ Ret	Tip II Hata ( $\beta$ )	Testin Gücü ( $1 - \beta$ )

Güçü göz ardı etmek ve uygun bir güç karşılaştırması olmadan örnek büyüklüklerini veya grupları karşılaştırmak, verilerimizin sağlamlığını ve dolayısıyla sonuçlarımızın güvenilirliğini riske atacaktır (2). İstatistiksel güç, araştırmada bir hipotezin denetlenmesi için kullanılan istatistiksel testler sonucunda varılan kararın ne kadar güvenilir ve geçerli olduğunu olasılık olarak tahmin eden bir yaklaşımdır (15).

MWU ile t testlerinin bazı özellikleri açısından performanslarının karşılaştırıldığı simülasyon çalışmaları mevcuttur (3, 16-18). Bu çalışmada parametrik olan t testi ile parametrik olmayan MWU testinin güç bakımından performansları karşılaştırılmıştır. Ayrıca, Monte Carlo simülasyonu ile MWU testi ve t testi için küçük örnek hacimleri

üretilecek, sıfır hipotezinin ret oranının belirlenmesi amaçlanmış, bu oran bakımından testlerin performansları incelenmiş ve elde edilen sonuçlar yapılan benzer çalışmalarla karşılaştırılmıştır.

#### MATERYAL ve METOT

Microsoft Excel programı ile oluşturulan rastgele sayılar çalışma materyalini oluşturmuştur. Araştırmada, testin güvenilirliği  $(1-\alpha)=0.95$  olarak belirlenmiştir. Çalışmada, “ $H_1$ : Örnekler farklı kitlelerden çekilmiştir” alternatif hipotezi test edilmiştir. Sıfır hipotezinin ret edilme oranı testin gücü olarak değerlendirilmiştir. Çalışmada varyansların hem homojen hem de heterojen olduğu ve gruplarda örnek büyüklüğünün eşit olduğu durumlarda elde edilen sonuçlar karşılaştırılmıştır.

MWU testi ve T testinin güçlerinin belirlenmesi için, her grubun örnek büyüklüğünün 3, 4 ve 5 olduğu durumların her biri için 7 grup oluşturulmuş ve her grup için 10000 tekrar yapılmıştır. Birinci grup ortalaması 0 ve varyansı 1 olan standart normal  $(N(0,1))$  dağılımlıdır. İlk kitlenin standart sapmasına  $1\sigma$  ve  $2\sigma$  eklemek suretiyle heterojen varyans grupları oluşturulmuştur. Grupların ortalamaları aralarındaki fark yine ilk kitlenin ortalamasına  $1\sigma$  ve  $2\sigma$  birimleri eklemek suretiyle oluşturulmuştur. Oluşturulan 7 grubun ortalama ve standart sapmaları sırasıyla (0,1), (1,1), (2,1), (1,2), (1,3), (2,2) ve (2,3) olarak belirlenmiştir. Her bir örnek hacmi için simülasyon 10000'er defa tekrar edilmiştir. Her bir kombinasyon ve farklı örnek hacimleri için birinci grup ile diğer gruplar karşılaştırılmış, sıfır hipotezinin ret edilme oranı belirlenmiştir. Bütün grup karşılaştırmalarında hem t testi hem de MWU testi uygulanmıştır. T testi ile elde edilen sonuçlardan, varyansların eşitliğine bağlı olarak ilgili olan sonuçlar kullanılmıştır. Rasgele sayıların oluşturulması işlemi Microsoft Excel ile, MWU ve t testleri SPSS 26 (19) paket programıyla yapılmıştır. Monte Carlo simülasyon çalışması aşağıdaki gibi uygulanmıştır;

1-Örnek büyüklüğü 3 ve  $N(0,1)$  dağılımlı kitle oluşturuldu.

2-Varyansları ilk gruba eşit fakat ortalamaları farklı olan 2. ve 3. gruplar oluşturuldu. Heterojen varyans için ilk gruptan hem varyansları hem de ortalamaları farklı olan 4., 5., 6. ve 7. gruplar oluşturuldu.

3-Simülasyon kullanılarak örnek büyüklüğü kadar veri üretildi, karşılık gelen veri örneğe alındı.

4-Aynı verilere t testi ve MWU testi uygulanarak sonuçlar kaydedildi.

5-Üçüncü adım ve dördüncü adım 10000 kez tekrarlandı.

6-Birinci adımdaki örnek büyüklüğü değiştirilerek işlemler tekrarlandı.

#### BULGULAR

Örnek büyüklüğünün 3, 4 ve 5 olduğu, grup ortalamalarının farklı ve varyanslarının eşit olduğu durumlarda Mann-Whitney U ve Bağımsız Örneklem t testi ile elde edilen testin gücü değerleri tablo 2'de sunulmuştur.

**Tablo 2.** Grup ortalamasının farklı ve varyanslarının eşit olduğu durumlardaki testin gücü.

**Table 2.** The power of the test when the group mean is different and the variances are equal.

Örnek Hacmi ( $n_1, n_2$ )	$\bar{x}_1, \bar{x}_2 : 0,1$ $\sigma_1^2, \sigma_2^2 : 1,1$		$\bar{x}_1, \bar{x}_2 : 0,2$ $\sigma_1^2, \sigma_2^2 : 1,1$	
	MWU (%)	t-test (%)	MWU (%)	t-test (%)
3.3	26.32	16.25	62.47	46.53
4.4	23.38	22.29	65.94	65.37
5.5	28.98	28.28	79.03	80.29

MWU: Mann-Whitney U Test

Birinci grup  $(N(0,1))$  ile ikinci grup  $(N(1,1))$  arasında yapılan karşılaştırmalarda örnek büyüklüğü arttıkça t testinin gücü artmasına rağmen MWU testinin gücü örnek büyüklüğünün 3'ten 4'e çıkarıldığında azaldığı belirlenmiştir. Birinci grup  $(N(0,1))$  ile üçüncü grubun  $(N(2,1))$  karşılaştırılmasında elde edilen sonuçlara göre hem MWU testinin hem de t testinin gücü örnek büyüklüğüyle paralel bir şekilde artış gösterdiği tespit edilmiştir (Tablo 2).

Örnek büyüklüğünün 3, 4 ve 5 olduğu, grup ortalamalarının ve varyanslarının farklı olduğu

durumlarda Mann-Whitney U ve Bağımsız Örneklem t testi ile elde edilen testin gücü değerleri tablo 3'de sunulmuştur.

**Tablo 3.** Grup ortalamalarının ve varyanslarının farklı olduğu durumlarda testin gücü.

**Table 3.** The power of the test in cases where group means and variances are different.

Örnek Hacmi ( $n_1, n_2$ )	$\bar{x}_1, \bar{x}_2 : 0,1$ $\sigma_1^2, \sigma_2^2 : 1,4$		$\bar{x}_1, \bar{x}_2 : 0,1$ $\sigma_1^2, \sigma_2^2 : 1,9$		$\bar{x}_1, \bar{x}_2 : 0,2$ $\sigma_1^2, \sigma_2^2 : 1,4$		$\bar{x}_1, \bar{x}_2 : 0,2$ $\sigma_1^2, \sigma_2^2 : 1,9$	
	MWU (%)	t-test (%)	MWU (%)	t-test (%)	MWU (%)	t-test (%)	MWU (%)	t-test (%)
3.3	14.36	10.05	18.27	10.52	38.22	24.95	28.78	17.28
4.4	15.18	13.56	12.46	11.11	36.59	34.43	24.08	21.96
5.5	14.92	14.74	11.56	11.76	41.64	43.16	25.21	26.93

MWU: Mann-Whitney U test

Varyansların homojen olmadığı durumlarda yapılan karşılaştırmalarda birinci grup ( $N \sim (0,1)$ ) ile dördüncü grup ( $N \sim (1,2)$ ) arasında yapılan karşılaştırmalarda örnek büyüklüğünün artmasıyla birlikte t testinin gücünün arttığı, örnek büyüklüğünün 5'e yükseltildiği durumda MWU testinin gücünün azaldığı tespit edilmiştir. Birinci grup ( $N \sim (0,1)$ ) ile beşinci grubun ( $N \sim (1,3)$ ) karşılaştırıldığı durumda örnek büyüklüğünün artmasına bağlı olarak t testinin gücünün arttığı ve MWU testinin gücünün azaldığı tespit edilmiştir. Birinci grup ( $N \sim (0,1)$ ) ile altıncı grubun ( $N \sim (2,2)$ ) ve birinci grup ( $N \sim (0,1)$ ) ile yedinci grubun ( $N \sim (2,3)$ ) karşılaştırıldığı durumlarda benzer sonuçlar elde edilmiş olup, örnek büyüklüğünün artmasına paralel olarak t testinin gücünün arttığı fakat örnek büyüklüğünün 4 olması durumunda MWU testinin gücünün azaldığı belirlenmiştir (Tablo 3).

### TARTIŞMA ve SONUÇ

MWU testi ile t testi karşılaştırıldığında örnek büyüklüğünün  $n_1=n_2=3$  ve  $n_1=n_2=4$  olduğu durumlarda, varyansların homojen veya heterojen olduğu bütün gruplardaki karşılaştırmalarda MWU testinin daha güçlü olduğu tespit edilmiştir. Örnek büyüklüğünün  $n_1=n_2=5$  olduğu durumda yapılan karşılaştırmalarda birinci grup ( $N \sim (0,1)$ ) ile ikinci ( $N \sim (1,1)$ ) ve dördüncü gruplar ( $N \sim (1,2)$ ) arasındaki karşılaştırmalarda MWU testinin t testine göre daha

güçlü sonuçlar verdiği, birinci gruba diğer grupların karşılaştırıldığı durumlarda t testinin MWU testine göre daha güçlü sonuçlar verdiği tespit edilmiştir.

Bütün karşılaştırmalarda örnek büyüklüğünün artmasına paralel olarak t testinin gücünün arttığı ortaya çıkmıştır. Fakat MWU testinin gücünün özellikle örnek büyüklüğünün  $n_1=n_2=4$  olduğu durumlarda,  $n_1=n_2=3$  olduğu duruma göre azaldığı belirlenmiştir.

Varyansların homojen ve heterojen olduğu durumlarda elde edilen sonuçlara göre, varyansların heterojen olması durumunda her iki testin güçlerinin oldukça azaldığı belirlenmiştir. İki grubun varyansları arasındaki farklılık arttıkça testlerin gücünün azaldığı tespit edilmiştir. Parametrik testler için gerekli bir varsayım olan varyansların homojen olmasının, nonparametrik bir test olan MWU testinin gücünü de etkilediği ortaya çıkmıştır. Bu sonuçlara benzer olarak, Bidak (3) yaptığı çalışmada varyanslardaki farklılık arttıkça MWU ve t testlerinde tip I hata oranının arttığını bildirmiştir. Bu çalışmada elde edilen sonuçların aksine, Zimmerman (17) örnek büyüklüğünün  $n_1=n_2=25$  olduğu durumda varyanslar arasındaki farklılık arttıkça testin gücünün arttığını bildirmiştir.

Genel olarak değerlendirildiğinde varyansların heterojen olduğu durumlarda ve varyansların homojen ve etki büyüklüğünün  $d = 1$  olduğu durumda her iki testin gücünün % 50'nin altında kaldığı belirlenmiştir. Varyansların homojen olduğu

ve etki büyüklüğünün  $d = 2$  olduğu durumda bütün örnek büyüklükleri için MWU testinin gücü % 60'ın üzerinde tespit edilirken, örnek büyüklüğünün  $n_1=n_2=4$  ve  $n_1=n_2=5$  olması durumunda t testinin gücü de % 60'ın üzerinde olduğu belirlenmiştir. Örnek büyüklüğünün  $n_1=n_2=5$ , varyansların homojen ve etki büyüklüğünün  $d = 2$  olduğu durumda t testinin (% 80.29) MWU testine (% 79.03) göre daha güçlü olduğu tespit edilmiştir. Yapılan benzer çalışmalarda, Zimmerman (18) varyansların eşit,  $n_1=n_2=30$  ve  $\alpha=0.01$  olduğu durumda MWU ve t testlerinin güçlerini sırasıyla %9 ve %10 olarak, Bindak (3) etki büyüklüğünün  $d=1$  ve örnek büyüklüğünün  $n_1=n_2=8$  olduğu durumda MWU ve t testlerinin güçlerini sırasıyla %42 ve %44, Dixon (20) etki büyüklüğünün  $d=3$ ,  $\alpha=0.025$  ve  $n_1=n_2=5$  olduğu durumda medyan kullanılarak testin gücünün %78 olarak elde edildiğini, Doğan ve Doğan (2020)  $n_1=n_2=6$  ve etki büyüklüğünün  $d \geq 1.8$  olduğu durumda testin gücünün %80 olduğunu bildirmişlerdir.

Yapılan bazı benzer simülasyon çalışmalarında bu çalışmanın sonuçlarına benzer olarak, Leys & Schumann (21) ve Bindak (3) hem parametrik hem de nonparametrik testlerin heterojen varyansa (heteroscedasticity) karşı duyarlı olduğunu, Kartal (22) aynı koşullar altındaki parametrik testlerin nonparametrik testlerden daha güçlü olduğunu, Sohn ve ark. (23) aynı istatistiksel güç düzeyine ulaşmak için t testlerinin nonparametrik eşdeğerlerinden daha küçük örnek boyutları gerektirdiğini bildirmişlerdir. Yapılan diğer çalışmalarda, Dixon (20) dört nonparametrik testin (rank-sum, maximum deviation, median, and total number of runs) normal etki büyüklüğü için t-testi ile karşılaştırıldığında, çok küçük örnek hacmi (3, 4, 5) ve küçük  $\alpha$  için daha güçlü olduğunu, Zimmerman (18) ve Kasuya (24) Mann-Whitney U testinin örnek hacimlerinin eşit olmaması durumunda ve özellikle varyansların homojen olmaması durumunda I. tip hata oranının çok yükseldiğini, Botta-Dukat (25) normal dağılım varsayımının karşılandığı çalışmalarda standardize edilmiş etki büyüklüğünün kullanılabilirliğini bildirmişlerdir. Ayrıca

varsayımların ne kadar ihlal edildiğinde nonparametrik testlerin kullanılması gerektiği, araştırmacıların fikir birliğine varamadıkları bir konudur. Araştırmacıların birçoğu parametrik testlerin daha güçlü olduğunu ve çok şiddetli birkaç ihlal olmadıkça nonparametrik testlerin kullanılmasını önermemektedirler (3,16).

Sonuç olarak varyansların homojen olmadığı veya varyansların homojen olduğu ve etki büyüklüğünün  $d = 1$  olduğu küçük örneklerde MWU ve t testinin güç bakımından yetersiz kaldığı tespit edilmiştir. Varyansların homojen olduğu ve etki büyüklüğünün  $d = 2$  olduğu durumda, örnek büyüklüğünün  $n_1=n_2=3$  ve  $n_1=n_2=4$  olduğu zaman MWU testinin t testine göre daha güçlü olduğu fakat bilimsel çalışmalarda istenilen güce ulaşamadığı belirlenmiştir. Bilimsel çalışmalarda tercih edilen durum testin gücünün % 80'in üzerinde olmasıdır. Belirtilen bu güce ulaşmak için aşırı gözlemlerin bulunmadığı, varyansların homojen, etki büyüklüğünün  $d = 2$  ve örnek büyüklüğünün  $n_1=n_2=5$  olması durumunda t testinin uygulanabileceği ortaya çıkmıştır. Bunun dışındaki durumlarda hem t testinin hem de MWU testinin güçlerinin % 80'in altında kaldığı tespit edilmiştir.

#### Çıkar Çatışması

Yazar, çıkar çatışması olmadığını beyan eder.

#### KAYNAKLAR

1. Ya Doğan İ., Doğan, N., 2020. Deney Hayvanı Kullanılan Çalışmalarda Örneklem Büyüklüğünün Kaynak Eşitlik Yöntemi ile Tahmini. Türkiye Klinikleri J Biostat, 12, 211-218.
2. Meyners M., Carr BT., Hasted A., 2020. To replicate or not to replicate, or when did we start to ignore the concept of statistical power?. Food Qual Prefer, 79.
3. Bindak R., 2014. Mann-Whitney U ile Student't testinin I. Tip Hata ve Güç bakımından Karşılaştırılması: Monte Carlo Simülasyon Çalışması. AKÜ FEMÜBİD, 14, 5-11.
4. Puzyr R., Haikova T., Majerník J., Karkova M.,

- Kmec J., 2018. Experimental study of the process of radial rotation profiling of wheel rims resulting in formation and technological flattening of the corrugations. *journalmt*, 18, 106-111.
5. Mendes M., 2012. Uygulamalı Bilimler İçin İstatistik ve Araştırma Yöntemleri. p: 155, Kriter Yayınevi, Çanakkale.
  6. Miller I., Miller M., 2006. John Freund Matematiksel İstatistik. 6th çev. ed.: Ü. Şenesen., 551-552, Literatür Yayınevi, İstanbul.
  7. Baştürk R., 2011. Nonparametrik İstatistik Yöntemler. 2. Basım, 99-110, Arı Yayıncılık, Ankara.
  8. Özçomak MS., Çebi K., 2017. İstatistiksel Güç Analizi: Atatürk Üniversitesi İktisadi Ve İdari Bilimler Dergisi Üzerine Bir Uygulama. *atauniiibd*, 31, 413-431.
  9. Funder DC., Ozer DJ., 2019. Evaluating effect size in psychological research: Sense and nonsense, *AMPPS*, 2, 156-168.
  10. Cohen J., 1992. Statistical Power Analysis. *Curr Dir Psychol Sci*, 1, 98-101.
  11. Orhan AT., Men DD., 2018. Web Tabanlı Öğretimin Fen Dersi Başarısına Ve Fen Dersine Yönelik Tutuma Etkisi: Bir Meta Analiz Çalışması. *cbayarsos*, 16, 245-284.
  12. Yalçiner AY., Günday ÜR., 2020. Yalın altı sigma metodu ve bankacılık sektöründe uygulanması. *DÜBİTED*, 8, 188-209.
  13. Olejnik SF., 1984. Planning educational research: Determining the necessary sample size. *JXE*, 53, 40-48.
  14. Kul S., 2011. Klinik araştırmalarda örnek genişliği belirleme. *Plevra Bülteni*, 2, 129-132.
  15. Keskin B., 2020. İstatistiksel güç bir araştırmanın sonuçlarına etki eder mi? örneklem büyüklüğüne nasıl karar verilmeli?. *cbayarsos*, 18 (Armağan Sayısı), 157-174.
  16. Zimmerman DW., Zumbo BD., 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*. pp 481, Lawrence Erlbaum, Hillsdale, New Jersey.
  17. Zimmerman DW., 2004. Inflation of type I error rates by unequal variances associated with parametric, nonparametric and rank transformation tests. *Psicologica: International Journal of Methodology and Experimental Psychology*, 25(1), 103-133.
  18. Zimmerman DW., 2006. Two separate effects of variance heterogeneity on the validity and power of significance tests of location. *J. statmet*, 3, 351-374.
  19. IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp.
  20. Dixon WJ., 1954. Power under normality of several nonparametric tests. *AMS*, 25, 610-614.
  21. Leys C., Schumann S., 2010. A nonparametric method to analyze interactions: The adjusted rank transform test. *JESP*, 46, 684-688.
  22. Kartal M., 2006. Bilimsel Araştırmalarda Hipotez Testleri Parametrik ve Nonparametrik Teknikler, 3. Basım. s:148, Nobel Basımevi, Ankara.
  23. Sohn W., Jeong M., Jeong K., 2020. Theoretical comparative study of t tests and nonparametric tests for final status surveys of MARSSIM at decommissioning sites. *Ann Nucl Energy*, 135, 106945.
  24. Kasuya E., 2001. Mann-Whitney U test when variances are unequal. *Anim Behav*, 61, 1247-1249.
  25. Botta-Dukat Z., 2018. Cautionary note on calculating standardized effect size (SES) in randomization test. *Community Ecol*, 19, 77-83.