



Ardahan Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi

<https://dergipark.org.tr/tr/pub/aruibfdergisi>



Küçük, eşit ve büyük örnek hacimlerinde Wald-Wolfowitz, Kolmogorov-Smirnov ve Mann-Whitney testlerinin I. tip hata oranlarının karşılaştırılması*

Comparison of type I error rates of Wald-Wolfowitz, Kolmogorov-Smirnov, and Mann-Whitney tests in small, equal, and large sample volumes

Sahib Ramazanov^{a**}, Ötüken Senger^b

^a Dr., Karabük Üniversitesi, Lisansüstü Eğitim Enstitüsü, İşletme Bölümü, Karabük, Türkiye, sahibramazanli@gmail.com, ORCID: 0000-0003-2582-3188

^b Prof. Dr., Kafkas Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, Kars, Türkiye, otukensenger@gmail.com, ORCID: 0000-0001-9815-7054

MAKALE BİLGİSİ

Makale geçmişi:

Başvuru: 6 Aralık 2023

Kabul: 27 Aralık 2023

Anahtar kelimeler:

Wald-Wolfowitz Testi,

Mann-Whitney Testi,

Kolmogorov-Smirnov İki Örnek Testi,

I. Tip Hata,

Monte Carlo Çalışması

Makale türü:

Araştırma makalesi

ARTICLE INFO

Article history:

Received: 6 December 2023

Accepted: 27 December 2023

Keywords:

Wald-Wolfowitz Test,

Mann-Whitney U Test,

Kolmogorov-Smirnov Two-Sample Test,

Type I Error,

Monte Carlo Study

Article type:

Research article

ÖZET

Sosyal bilimlerde, varyansın normalliği ve homojenliği gibi önemli varsayımların ihlal edildiği durumlarda, parametrik olmayan istatistiksel testlerin kullanılması önerilmektedir. Bu tür durumlarda, bağımsız iki örneklemin verilerini test etmek için Wald-Wolfowitz Sıra Sayıları Testi, Kolmogorov-Smirnov İki Örneklem Testi ve Mann-Whitney Testi gibi non-parametrik testlere başvurulmaktadır. Bu çalışma, bu üç farklı test yönteminin belirli koşullar altında I. tip hata oranlarını karşılaştırmak amacıyla gerçekleştirilmiştir. Bu çerçevede, her bir testin I. tip hata oranları, büyük ve küçük örneklem grupları için ayrı ayrı analiz edilmiştir. Araştırmada Uniform-like, Logistic-like ve Double exponential-like dağılımları ile 24 farklı örnek büyüklüğü kombinasyonu kullanılmıştır. Örneklerin yarısı büyük, yarısı küçüktür. Küçük örneklerin içinde hem eşit hem de farklı hacimlere sahip örnekler bulunmaktadır. Benzer şekilde, büyük örneklerin içinde de hem eşit hem de farklı hacimlere sahip örnekler bulunmaktadır. Elde edilen verilere dayalı olarak Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz testlerinin I. tip hata oranları, her kombinasyon için büyük bir Monte Carlo Simülasyonu ile hesaplanmıştır. Sonuçlar, örnek büyüklükleri ve örnek hacmi farklılıklarından bağımsız olarak incelendiğinde, Kolmogorov-Smirnov İki Örneklem Testi'nin Mann-Whitney ve Wald-Wolfowitz testlerine göre daha düşük bir I. tip hata oranına sahip olduğunu göstermektedir. Bu çalışma, araştırmacılara bağımsız iki örneklem verilerini test etmek için en uygun istatistiksel yöntemi seçerken dikkate almaları gereken önemli bir bilgi sunmaktadır.

ABSTRACT

In social sciences, when important assumptions such as the normality and homogeneity of variances are violated, the use of non-parametric statistical tests is recommended. In such cases, non-parametric tests like the Wald-Wolfowitz Rank-Sum Test, Kolmogorov-Smirnov Two-Sample Test, and Mann-Whitney U Test are employed to analyze the data of two independent samples. This study was conducted to compare the Type I error rates of these three different test methods under specific conditions. In this context, the Type I error rates of each test were analyzed separately for both large and small sample groups. The research utilized 24 different combinations of sample sizes with Uniform-like, Logistic-like, and Double exponential-like distributions. Half of the samples were large, and half were small, including samples of both equal and different sizes. Similarly, within the large samples, there were both equal and different-sized samples. Based on the obtained data, the Type I error rates of the Mann-Whitney, Kolmogorov-Smirnov, and Wald-Wolfowitz tests were calculated for each combination through an extensive Monte Carlo Simulation. The results indicate that, irrespective of variations in sample size and sample volume differences, the Kolmogorov-Smirnov Two-Sample Test has a lower Type I error rate compared to the Mann-Whitney and Wald-Wolfowitz tests. This study provides important information for researchers when selecting the most appropriate statistical method to test independent two-sample data.

* Çalışma, 19-21 Ekim 2023 tarihlerinde Bayburt'da düzenlenen, 5. Uluslararası Sosyal Bilimler Kongresi'nde özet bildiri olarak sunulmuştur.

** Sorumlu yazar / Corresponding author

E-posta / E-mail: sahibramazanli94@gmail.com

Atıf / Citation: Ramazanov, S. ve Senger, Ö. (2023). Küçük, eşit ve büyük örnek hacimlerinde Wald-Wolfowitz, Kolmogorov-Smirnov ve Mann-Whitney testlerinin I. tip hata oranlarının karşılaştırılması. *Ardahan Üniversitesi İİBF Dergisi*, 5(2), 137-144. <http://doi.org/10.58588/aru-jfeas.1388231>

1. Giriş

Eğer verilerin çekildiği ana kütle normal dağılım varsayımını karşılamıyorsa ve parametrik testler için dağılım parametrelerine dair bir önermede bulunamıyorsa, parametrik testler kullanıldığında sonuçlar hatalı olabilir. Bu tür durumlarda, non-parametrik testler daha uygun bir seçenek olabilir. Ancak, unutulmamalıdır ki non-parametrik yöntemler, normal dağılım ve veri homojenliği varsayımlarının ihlaline karşı kesin bir koruma sağlayamayabilir (Zimmerman, 1998). Non-parametrik istatistik, iki temel özellik sunar: Öncelikle, veriler düzenli bir sıralama yaparak sunulur. Daha sonra, sürekli gözlem yapılan veriler arasındaki bu sıralama ilişkisini vurgular ve tüm non-parametrik yöntemlerin temelinde olasılık teorisini gerektirir (Gibbons, 1971).

Birçok istatistiksel yöntem, analiz edilen örneğin seçildiği ana popülasyonun dağılımıyla ilgili kesin varsayımlarda bulunan parametrik istatistiklerden oluşur. Eğer ham veriler bu varsayımları karşılamıyorsa, araştırmacı alternatif yaklaşımları kullanarak verileri analiz edebilir. Bir diğer yaklaşım ise, veri dağılımı hakkında sınırlı veya hiç varsayım yapmayan, ancak parametrik alternatiflerine kıyasla daha az güçlü olan non-parametrik prosedürleri kullanmaktır (Boslough ve Watters, 2008).

Farklı veri örneklerinin karşılaştırılmasında bir dizi non-parametrik test kullanılmaktadır. Pratik uygulamalarda, en sık kullanılanlar arasında Mann-Whitney U testi, Wald-Wolfowitz Sıralama Testi ve Kolmogorov-Smirnov İki Örnek Testi bulunmaktadır. Bu testlerin temel amacı, iki bağımsız örnek arasındaki benzerliği değerlendirmektir. Bu benzerlik, her iki örneğin aynı ana popülasyona ait olduğunu veya popülasyonların karşılaştırılabilir olduğunu gösterebilir (Conover, 1999). Klasik bir parametrik model, X_1, \dots, X_n gözlemlerinin F_θ popülasyon dağılımına sahip olduğunu varsayar ve bu bağlamda bilinmeyen parametreler bir vektör θ ile ifade edilir. Örneğin, normal dağılım $N(\mu, \sigma^2)$ içerisinde $\theta = (\mu, \sigma^2)^t$ gibi. Bu araştırma, Kolmogorov-Smirnov İki Örnek Testi, Wald-Wolfowitz Sıralama Testi ve Mann-Whitney U Testi'nin performansı değerlendirilmiştir (Ramazanov ve Askerbeyli, 2023).

Hipotez testlerinde, sıfır hipotezinin yanlış bir şekilde kabul edilmesi veya alternatif hipotezin yanlış bir şekilde reddedilmesi durumunda hatalar ortaya çıkacaktır. Bu nedenle, iki tür hata vardır. Sıfır hipotezinin doğru bir şekilde reddedilmesi durumunda I. tip hata ortaya çıkar (Brink, 2010). Sembolik olarak ifade edildiğinde, anlamlılık seviyesi α ile tanımlanır. Bu nedenle, α için

$$\alpha = P_\theta(\text{Ret } H_0) = P_\theta(T(X) \geq c), \quad \theta \in \theta_0 \quad (1)$$

gösterilebilir (Geyer, 2001).

İstatistiksel çıkarımlarda, X 'ten $\{0,1\}$ 'e Y 'ye olan bir T testi, iki tür hata yapma olasılığını değerlendirir. Bu hatalardan biri I. tip hata olarak ifade edilir.

$$\alpha_T(P) = P(T(X) = 1) \quad P \in P_0 \quad (2)$$

ile gösterilir (Shao, 2003).

Geleneksel olarak, I. tip hata için kabul edilebilirlik düzeyi 0,05 olarak belirlenir. α 'yı 0,05'e ayarlamak, I. tip hata olasılığının %5 olarak kabul edildiği anlamına gelir (Boslough ve Watters, 2008).

Mann-Whitney testi, iki bağımsız örnek grubunun aynı ana kütle tarafından türetildiğini belirlemek için kullanılır. Eğer örnek büyüklüğü 10 veya daha az ise ($m \leq 10$ ve $n \leq 10$), bu durum da küçük örnekler için önerilen W test istatistiği kullanılır. W test istatistiği bakımından şu şekildedir:

$W_x = \sum R_1$ 1. anakütleden alınan X'lerin çoklu değişkenlerinin sıralarının toplamı

$W_y = \sum R_2$ 2. anakütleden alınan Y'lerin çoklu değişkenlerinin sıralarının toplamı

$$W_x + W_y = \frac{N(N+1)}{2} \quad (3)$$

gösterilir.

Eğer örnek büyüklüğü 10'dan fazlaysa ($m > 10$ veya $n > 10$), büyük örnekler için normal yaklaşım formülü kullanılır. Bu formül, özellikle bir örneğin büyüklüğünün 3 veya 4 olduğu ve diğerinin 12'den büyük olduğu durumlarda da önerilir. Formül aşağıda verilmiştir:

$$z = \frac{W_x \pm 0,5 - \frac{m(N+1)}{2}}{\sqrt{\frac{mn(N+1)}{12}}} \quad (4)$$

Conover (1999), Kolmogorov-Smirnov iki örneklem testi için gereken bazı ön koşulları sıralamıştır:

- Örnekler, kendi popülasyonlarından rastgele seçilir.
- Kullanılan ölçüm ölçeği en az sıralıdır veya daha yüksek bir ölçek kullanılır.
- Gözlemlenen değişken, doğası gereği sürekli bir değişkendir.
- İki örnek, birbirinden bağımsızdır.

Daniel tarafından belirtilene göre, Kolmogorov-Smirnov İki Örnek Testi için test istatistikleri, küçük ve büyük örnekler için:

$$D = \max |S_1(x) - S_2(x)| \quad (5)$$

ile gösterilmiştir.

2. Literatür Taraması

Literatür incelemesi, yerli ve yabancı kaynaklarda gerçekleştirilen yedi farklı çalışmayı kapsamaktadır. Bu çalışmalar, çeşitli istatistiksel testlerin performansını değerlendirmiştir. İnceleme, her bir çalışmanın metodolojisini ve bulgularını özetleyerek, istatistiksel güç ve I. tip hata oranlarını karşılaştırmıştır. Aşağıda, bu çalışmalar sırasıyla verilmiştir.

Blair ve Higgins (1985), çalışması, Wilcoxon işaret sıralama testinin eşleştirilmiş örnekler arasındaki farkları tespit etme gücünü değerlendirmişlerdir. Ayrıca, I. tip ve II. tip hata oranlarını araştırarak testin doğruluğunu ve güvenilirliğini değerlendirmişlerdir.

Penfield (1994), çalışmasında, normal ve normal olmayan dağılımlar için farklı örnek büyüklükleriyle yapılan testlerin I. tip hata oranlarını incelemiştir. Eşit anakütle varyansları durumunda benzer hata oranları elde edilmiş, ancak eşit olmayan örnek büyüklükleri durumunda Mann-Whitney testinin diğer testlere göre daha güçlü olduğu sonucuna varılmıştır.

Fahome (1999), çalışması, parametrik olmayan testlerin smooth symmetric, extreme asymmetric, extreme bimodal ve multimodal lumpy dağılımlardaki I. tip hata oranlarını incelemiştir. Kolmogorov-Smirnov testinin p değerinin nominal I. tip hata oranına yakınsız olduğu, ancak Mann-Whitney testinin oldukça iyi sonuçlar verdiği tespit edilmiştir.

Tekindal (2007), çalışmasında, Brown-Forsythe, Cochran, Fmax ve

Satterwaite gibi varyans homojenliği test yöntemleri farklı deneme koşullarında (dağılım şekli, varyans oranları, gözlem sayısı, grup sayısı, vb.) ampirik olarak karşılaştırmıştır. 100,000 deneme üzerinden yapılan Monte Carlo simülasyonları sonucunda genel olarak hiçbir testin tüm deneme koşulları için önerilemeyecek, her bir testin farklı koşullarda kullanılabileceği sonucuna varılmıştır. Populasyon dağılımları normal olduğunda, Cochran ve Fmax testlerinin örnek hacminden bağımsız olarak belirlenen I. tip hata olasılığını koruma konusunda en güvenilir sonuçları verdiği görülmüştür. Normal olmayan dağılımlarda ise Satterwaite testinin daha güvenilir olduğu tespit edilmiştir. Ayrıca, testlerin güç değerlerinin, dağılım şekli yerine varyans oranları ve örnek hacmi üzerinde daha fazla etkili olduğu belirlenmiştir.

Akbulut (2008), çalışmasında, F testi, Welch-James testi, tek aşamalı yöntem ve Düzeltilmiş sıra sayısı dönüşüm testi gibi yöntemler, 2x2 faktöriyel deneme düzenindeki interaksiyon etkisini test etmek için incelenmiştir. Monte-Carlo simülasyon tekniği ile I. Tip hata ve testin gücü değerlerini elde etmiştir. F testi ve Düzeltilmiş sıra sayısı dönüşüm testi, varsayımlar sağlandığında Welch-James testi ve Tek Aşamalı yöntem göre I. Tip hata kontrolü bakımından daha geçerli ve testin gücü bakımından daha etkili sonuçlar vermiştir. Welch-James testi ve Tek Aşamalı yöntem ise varyanslar homojen olmadığında özellikle alt grup dengesizliği durumunda I. Tip hata kontrolü bakımından daha başarılı olmuştur.

Yiğit ve Gamgam (2011), klasik F testinin normallik ve homojen varyans varsayımlarına dayanmasının özellikle varyanslar homojen olmadığında sorunlara yol açtığını belirtmişlerdir. Özellikle küçük örneklem hacimlerinde bu durum daha belirgin hale gelmektedir. Literatürde bu soruna çözüm olarak geliştirilen Brown-Forsythe, genelleştirilmiş F, Scott-Smith, Welch, Xu-Wang testleri tanıtılmış ve bu testlerin deneysel birinci tip hata oranı ve güç açısından farklı yığın parametreleri ve örnek hacimleri altında nasıl performans gösterdikleri karşılaştırılmıştır.

Durand (2013), istatistiksel etkileşimlerin yaygın olduğu ancak genellikle düşük güçle tespit edildiğini belirtmiştir. Bu durumu aşmak için I. tip hata oranını artırmanın, yanlış etkileşimleri modele dahil etme olasılığını artırabileceği gösterilmiştir. Yapılan Monte Carlo simülasyonu çalışmasında, üç farklı etkileşim türü ve çeşitli senaryolarda yapılan 240 benzersiz simülasyon sonuçlarına göre, yüksek I. tip hata oranının sınırlı durumlarda savunulabilir olabileceği tespit edilmiştir. Genel olarak, etkileşim etkilerini test ederken I. tip hata oranını rutin olarak yükseltmenin tavsiye edilmediği, bunun yerine araştırmacılara etkileşim etkilerini önceden belirleme ve örneklem büyüklüğü hesaplamalarında bunları dikkate alma önerisi sunulmuştur.

Bindak (2014), parametrik t testinin varsayımlarının karşılanmadığı durumlarda alternatif olarak kullanılan Mann-Whitney U testi ile t testinin performanslarını karşılaştırmayı amaçlamıştır. Kitlenin normal dağılıma sahip olduğu durumu ele alan çalışmada, çeşitli örnek hacimleri ve farklı varyans oranları kullanılarak Monte Carlo simülasyonu gerçekleştirilmiştir. Elde edilen sonuçlar, özellikle küçük örnek hacimlerinde Mann-Whitney U testinin daha düşük hata oranları verdiğini, ancak örnek hacminin artmasıyla birlikte t testi ile benzer sonuçlar elde edildiğini göstermiştir. Ayrıca, eşit olmayan varyans durumunda özellikle örnek hacminin artmasıyla Mann-Whitney U testinin I.tip hata oranını koruyamadığı belirlenmiştir.

Meaney ve Moineddin (2014), çalışmalarında, biyomedikal araştırmalarda yaygın olarak karşılaşılan (0,1) aralığındaki yanıt değişkenlerini değerlendirmek amacıyla farklı modellerin performansını incelemişler. Lineer regresyon, beta regresyon, değişken-dispersiyonlu beta regresyon ve kesirli logit regresyon modelleri, Monte Carlo simülasyonu

karşılaştırılmıştır. Bulgular, modellerin yanıt verilerinin özelliklerine bağlı olarak farklı performans sergilediğini göstermektedir. Özellikle küçük örneklemelerde lineer regresyonun daha iyi performans gösterdiği ve beta regresyon modellerinde I. tip hata oranlarının bias düzeltme yöntemleriyle iyileştirilebileceği belirlenmiştir. Çalışmanın sonuçları, biyomedikal araştırmacılara model seçimi konusunda rehberlik sağlamaktadır.

Eltas (2021), çalışmasında, örnek hacminin 3, 4 ve 5 olduğu durumlarda parametrik Bağımsız Örneklem t testi ile Mann-Whitney U testini güç açısından karşılaştırarak, varyansların homojen olmadığı durumlarda her iki testin güçlerinin düştüğünü göstermektedir. Çalışmanın sonuçları, bilimsel araştırmalarda güçlü bir tercih olarak önerilen t testinin, homojen varyans ve belirli örnek büyüklükleri koşullarında daha etkili olduğunu ortaya koymaktadır.

Uysal ve Kılıç (2022), verilerin normal dağılıma sahip olup olmadığını belirleme amacıyla kullanılan çeşitli normalite testleri ve tanımlayıcı istatistikleri karşılaştırmışlar. Jarque-Bera, çarpıklık ve basıklık katsayılarına dayalı yöntemler I. tip hata açısından daha etkili bulunmuştur. Ancak, örneklem küçük, veri sürekli ve çarpıklık katsayısı -1 veya +1 olduğunda tüm yöntemlerin gücünde bir azalma gözlemlenmiştir.

3. Simülasyon Çalışması

Monte Carlo simülasyon çalışmaları, istatistiksel modellerin doğruluğunu, sınırlı örneklem büyüklükleri ve normal olmayan veri dağılımları gibi zorlu koşullar altında değerlendirmenin yanı sıra, yeni istatistiksel soruların keşfedilmesine ve bir istatistiğin ampirik dağılımını tahmin etme amacıyla çeşitli uygulamalar sunmaktadır (Mooney, 1997). Monte Carlo simülasyonu için SAS 9.00 bilgisayar yazılımı kullanılarak, çeşitli dağılımlar Fleishman'ın (1978) güç fonksiyonu ile oluşturulmuştur (Senger, 2013). Popülasyon dağılımlarını oluşturmak için, Fleishman'ın güç dönüşüm yöntemine uygun olarak, ortalama değeri sıfır ve standart sapma oranı bir olan standart bir normal dağılımdan rastgele sayılar üretilmiştir. Bu amaçla, SAS'ın RANNOR prosedürü kullanılmış ve Fleishman'ın güç fonksiyonu formülü ile dönüşüm işlemi gerçekleştirilmiştir.

Formül:

$$Y = a + [(d \times X + c) \times X + b] \times X \quad (6)$$

Fleishman'ın güç fonksiyonu formülü, sabitlere bağlıdır ve bu formül, SAS programı içerisindeki RANNOR prosedürü kullanılarak üretilen bir dağılım değişkeni olan Y ve ortalama değeri sıfır, standart sapma oranı bir olan normal olarak dağılmış bir rastgele değişken olan X'i içerir. a, b, c ve d değerleri, a'nın sabit olduğu ve a = -c, b, c ve d'nin değişken değerlere sahip olduğu Fleishman'ın yöntemi kullanılarak hesaplanır. Örnek popülasyonların oluşturulmasının ardından, güç simülasyonları PROC NPAR1WAY prosedürü ile gerçekleştirilir (Senger, 2013).

Çalışmada, bağımsız iki örneğe ait verilerin Mann-Whitney, Kolmogorov-Smirnov iki örnek ve Wald-Wolfowitz dizi sayıları testleri üzerinden karşılaştırılması gerçekleştirilmiştir. Söz konusu veriler, sıralı ikililer şeklinde (n₁, n₂) ifade edilmiştir. Burada, n₁ ve n₂, birinci ve ikinci örneğin örnek büyüklüklerini temsil etmektedir. Örneğin, (5, 5) ifadesi, her iki örneğin de 5 elemandan oluştuğunu göstermektedir.

Çalışma, üç farklı ana kütle dağılımı olan Uniform-Like, Logistic-Like ve Double Exponential-Like dağılımlarına odaklanmıştır. Her bir dağılım için, farklı örneklem büyüklükleri (5, 8, 10, 12, 16, 20 küçük ve eşit, (4, 16), (8, 16), (10, 20), (16, 4), (16, 8), (20, 10) küçük ve farklı, 25, 50, 75, 100 büyük ve eşit, (10, 30), (30, 10), (50, 75), (50, 100), (75, 50), (75, 100),

(100, 50), (100, 75) büyük ve farklı gibi çeşitli koşullar altında istatistiksel gücün nasıl değiştiği araştırılmıştır. Araştırmada (standart sapma oranı 1) kullanılmıştır. Çalışma, büyük örneklem boyutlarına yönelik 12 soru ve küçük örneklem boyutlarına ilişkin 12 soru olmak üzere toplam 24 sorunun analizini içermektedir. Daha spesifik olarak, küçük örneklem boyutları için 8 farklı çalışma sorusu incelenirken, büyük örneklem boyutları için de 8 farklı çalışma sorusu değerlendirilmiştir. Sonuç olarak, bu araştırma kapsamında toplamda 1152 syntax (24 x 6 x 8) analiz edilmiştir. Monte Carlo simülasyonları, her bir durum için ayrı ayrı 20000 simülasyon çalıştırılarak, SAS yazılım paketi kullanılarak gerçekleştirilmiştir.

Tablo 1. $\mu=0$ ve $\sigma=1$ için Fleishman'ın güç fonksiyonu

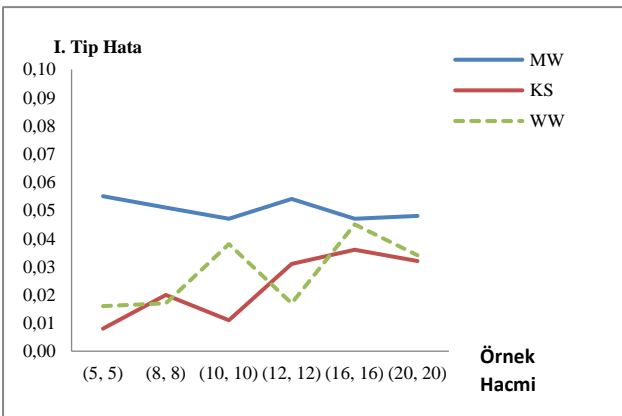
Dağılım	Çarpıklık (γ_1)	Basıklık (γ_2)	a	b	c	d
Uniform-Like ²	0,00	-1,20	0,00	1,2237300	0,00	-0,0636881
Logistic-Like ²	0,00	1,30	0,00	0,8807330	0,00	0,0382866
Double Exponential-Like ²	0,00	3,00	0,00	0,7823562	0,00	0,06790456

Kaynak: Lee (2007).

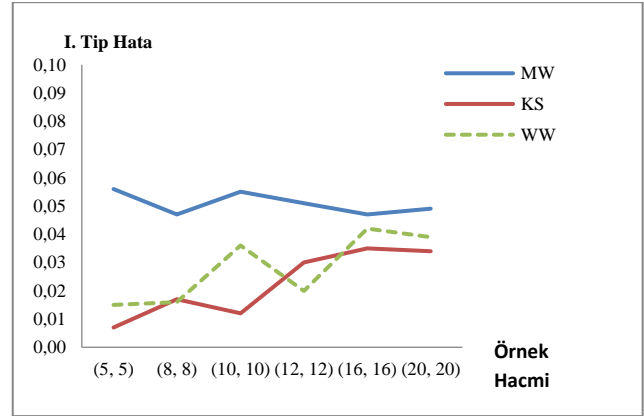
4. Simülasyon Sonuçları

Küçük ve eşit örneklem büyüklüklerine sahip Uniform-Like, Logistic-Like ve Double Exponential-Like dağılımları için, standart sapma oranının 1 olduğu varsayımıyla Mann-Whitney U testi, Kolmogorov-Smirnov testi ve Wald-Wolfowitz testi için hesaplanan I. tip hataları incelenmiştir. Örneklem büyüklükleri 5, 8, 10, 12, 16 ve 20 olarak seçilmiş ve bu büyüklükler arasında karşılaştırma yapılmıştır.

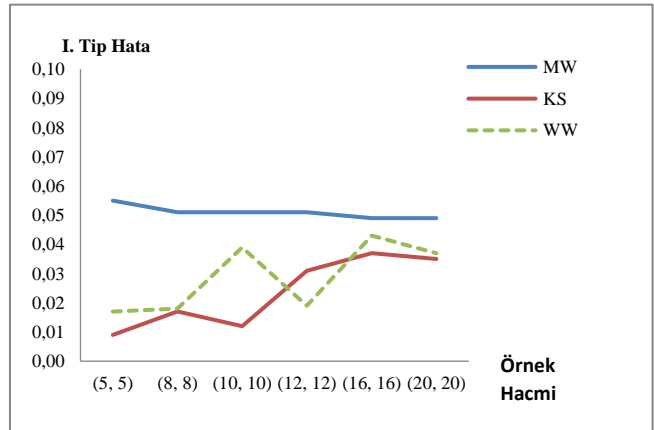
Mann-Whitney U testi için, standart sapma 1 durumu için I. tip hata oranları şu şekildedir: Uniform-Like: 0.047, Logistic-Like: 0.047, Double Exponential-Like: 0.049. Kolmogorov-Smirnov testi için, standart sapma 1 durumu için I. tip hata oranları şu şekildedir: Uniform-Like: 0.008, Logistic-Like: 0.007, Double Exponential-Like: 0.009. Wald-Wolfowitz testi için, standart sapma 1 durumu için I. tip hata oranları şu şekildedir: Uniform-Like: 0.016, Logistic-Like: 0.015, Double Exponential-Like: 0.017.



Şekil 1. Küçük ve eşit örnek büyüklüklerinde Uniform-Like dağılımı için standart sapma oranı 1 olduğunda Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz dizi sayıları testlerinin I. tip hata oranları



Şekil 2. Küçük ve eşit örnek büyüklüklerinde Logistic-Like dağılımı için standart sapma oranı 1 olduğunda Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz dizi sayıları testlerinin I. tip hata oranları



Şekil 3. Küçük ve eşit örnek büyüklüklerinde Double Exponential-Like dağılımı için standart sapma oranı 1 olduğunda Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz dizi sayıları testlerinin I. tip hata oranları

Tablo 2. Küçük örneklem durumlarında, örnek büyüklüklerinin eşit olduğu senaryolarda Mann-Whitney U testi, Kolmogorov-Smirnov testi ve Wald-Wolfowitz testinin I. tip hata oranları (Standart sapma oranı=1)

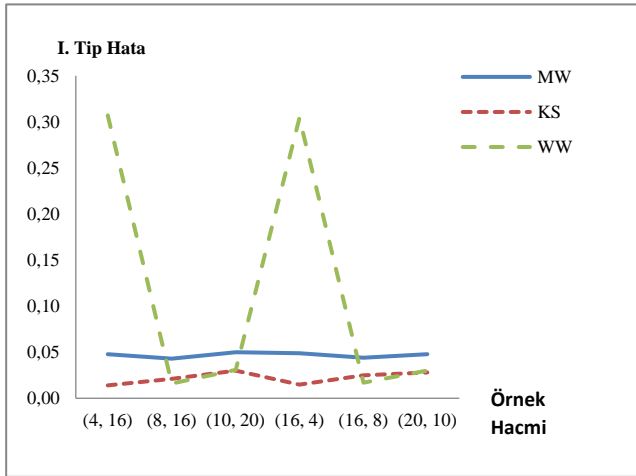
Popülasyon Dağılımı	n_1	n_2	I. Tip Hata Oranı		
			MW	KS	WW
Uniform-Like	5	5	0.055	0.008*	0.016 *
	8	8	0.051	0.020*	0.017*
	10	10	0.047*	0.011*	0.038*
	12	12	0.054	0.031*	0.017*
	16	16	0.047*	0.036*	0.045*
Logistic-Like	20	20	0.048*	0.032*	0.034*
	5	5	0.056	0.007*	0.015*
	8	8	0.047*	0.017*	0.016*
	10	10	0.055	0.012*	0.036*
	12	12	0.051	0.030*	0.020*
	16	16	0.047*	0.035*	0.042*
	20	20	0.049*	0.034*	0.039*

Double Exponential-Like	5	5	0.055	0.009*	0.017*
	8	8	0.051	0.017*	0.018*
	10	10	0.051	0.012*	0.039*
	12	12	0.051	0.031*	0.019*
	16	16	0.049*	0.037*	0.043*
	20	20	0.049*	0.035*	0.037*

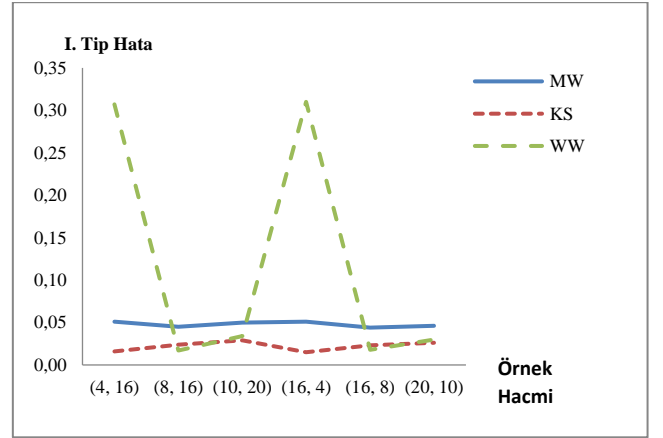
* $\alpha=0,05$ 'den küçük veya eşit olan değerler

Elde edilen verilere dayanarak, standart sapma oranı 1 olarak kabul edilen Uniform-Like, Logistic-Like ve Double Exponential-Like dağılımları için Mann-Whitney U testi, Kolmogorov-Smirnov testi ve Wald-Wolfowitz testi ile hesaplanan I. tip hataları, (4, 16), (8, 16), (10, 20), (16, 4), (16, 8) ve (20, 10) küçük ve farklı örneklem büyüklükleri arasında karşılaştırılmıştır.

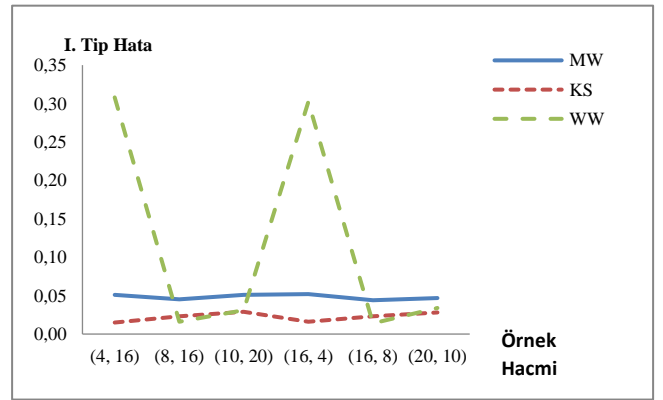
Mann-Whitney U testi için, standart sapma oranı 1 olduğunda I. tip hata oranları şu şekildedir: Uniform-Like: 0.043, Logistic-Like: 0.044, Double Exponential-Like: 0.044. Kolmogorov-Smirnov testi için, standart sapma oranı 1 olduğunda I. tip hata oranları şu şekildedir: Uniform-Like: 0.014, Logistic-Like: 0.016, Double Exponential-Like: 0.015. Wald-Wolfowitz testi için, standart sapma oranı 1 olduğunda I. tip hata oranları şu şekildedir: Uniform-Like: 0.016, Logistic-Like: 0.017, Double Exponential-Like: 0.014.



Şekil 4. Küçük ve farklı örnek büyüklüklerinde Uniform-Like dağılımı için standart sapma oranı 1 olduğunda Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz dizi sayıları testlerinin I. tip hata oranları



Şekil 5. Küçük ve farklı örnek büyüklüklerinde Logistic-Like dağılımı için standart sapma oranı 1 olduğunda Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz dizi sayıları testlerinin I. tip hata oranları



Şekil 6. Küçük ve farklı örnek büyüklüklerinde Double Exponential-Like dağılımı için standart sapma oranı 1 olduğunda Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz dizi sayıları testlerinin I. tip hata oranları

Tablo 3. Küçük örneklem durumlarında, örnek büyüklüklerinin farklı olduğu senaryolarda Mann-Whitney U testi, Kolmogorov-Smirnov testi ve Wald-Wolfowitz testinin I. tip hata oranları (Standart sapma oranı=1)

Popülasyon Dağılımı	n_1	n_2	I. Tip Hata Oranı		
			MW	KS	WW
Uniform-Like	4	16	0.048*	0.014*	0.307
	8	16	0.043*	0.021*	0.016*
	10	20	0.050*	0.030*	0.031*
	16	4	0.049*	0.015*	0.304
	16	8	0.044*	0.025*	0.017*
	20	10	0.048*	0.028*	0.030*
Logistic-Like	4	16	0.051	0.016*	0.307
	8	16	0.045*	0.024*	0.017*
	10	20	0.050*	0.029*	0.034*
	16	4	0.051	0.015*	0.310
	16	8	0.044*	0.023*	0.018*
	20	10	0.046*	0.026*	0.030*

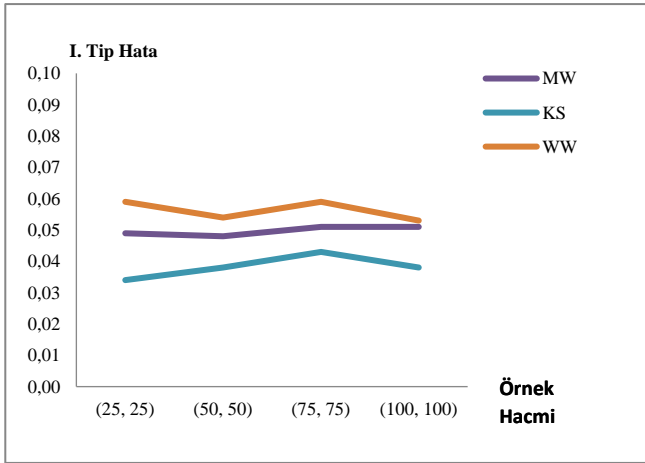
Double Exponential-Like	4	16	0.051	0.015*	0.308
	8	16	0.045*	0.023*	0.016*
	10	20	0.051	0.029*	0.031*
	16	4	0.052	0.016*	0.302
	16	8	0.044*	0.023*	0.014*
	20	10	0.047*	0.028*	0.034*

* $\alpha=0,05$ 'den küçük veya eşit olan değerler

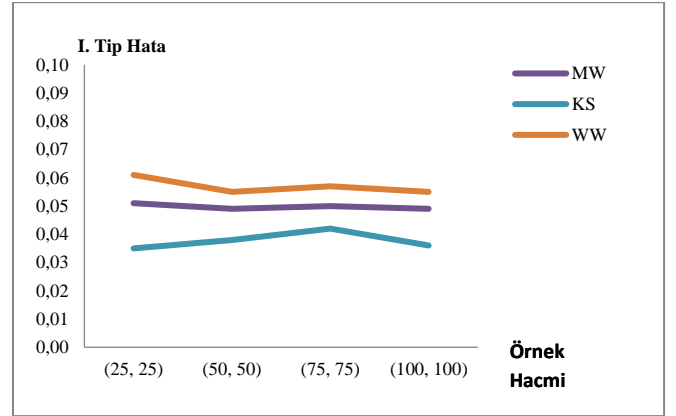
Bu tabloda, α anlamlılık seviyesi 0,05'in altında veya 0,05'e eşit olan değerler '*' sembolü ile işaretlenmiştir.

Elde edilen verilere dayalı olarak, standart sapma oranı 1 olarak ele alınan Uniform-Like, Logistic-Like ve Double Exponential-Like dağılımı için Mann-Whitney U testi, Kolmogorov-Smirnov testi ve Wald-Wolfowitz testi için hesaplanan I. tip hataları, 25, 50, 75 ve 100 büyük ve eşit örneklem büyüklükleri arasında karşılaştırılmıştır.

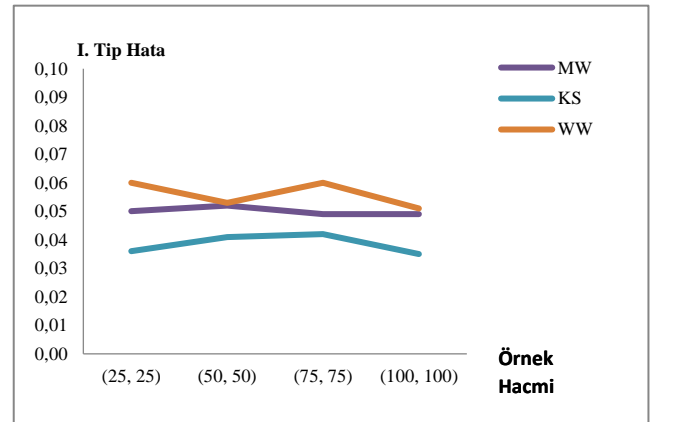
Mann-Whitney U testi için, standart sapma 1 durumu için I. tip hata oranları şu şekildedir: Uniform-Like: 0.048, Logistic-Like: 0.049, Double Exponential-Like: 0.049. Kolmogorov-Smirnov testi için, standart sapma 1 durumu için I. tip hata oranları şu şekildedir: Uniform-Like: 0.034, Logistic-Like: 0.035, Double Exponential-Like: 0.035. Wald-Wolfowitz testi için, standart sapma 1 durumu için I. tip hata oranları şu şekildedir: Uniform-Like: 0.053, Logistic-Like: 0.055, Double Exponential-Like: 0.053.



Şekil 7. Büyük ve eşit örnek büyüklüklerinde Uniform-Like dağılımı için standart sapma oranı 1 olduğunda Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz dizi sayıları testlerinin I. tip hata oranları



Şekil 8. Büyük ve eşit örnek büyüklüklerinde Logistic-Like dağılımı için standart sapma oranı 1 olduğunda Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz dizi sayıları testlerinin I. tip hata oranları



Şekil 9. Büyük ve eşit örnek büyüklüklerinde Double Exponential-Like dağılımı için standart sapma oranı 1 olduğunda Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz dizi sayıları testlerinin I. tip hata oranları

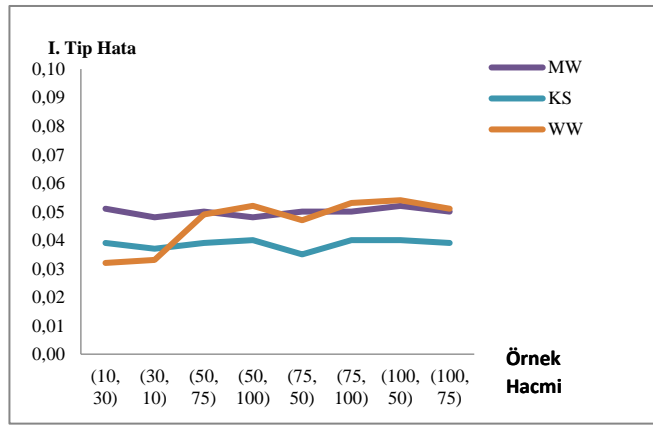
Tablo 4. Büyük örneklem durumlarında, örnek büyüklüklerinin eşit olduğu senaryolarda Mann-Whitney U testi, Kolmogorov-Smirnov testi ve Wald-Wolfowitz testinin I. tip hata oranları (Standart sapma oranı=1)

Popülasyon Dağılımı	n ₁	n ₂	I. Tip Hata Oranı		
			MW	KS	WW
Uniform-Like	25	25	0.049*	0.034*	0.059
	50	50	0.048*	0.038*	0.054
	75	75	0.051	0.043*	0.059
	100	100	0.051	0.038*	0.053
Logistic-Like	25	25	0.051	0.035*	0.061
	50	50	0.049*	0.038*	0.055
	75	75	0.050*	0.042*	0.057
	100	100	0.049*	0.036*	0.055
Double Exponential-Like	25	25	0.050*	0.036*	0.060
	50	50	0.052	0.041*	0.053
	75	75	0.049*	0.042*	0.060
	100	100	0.049*	0.035*	0.051

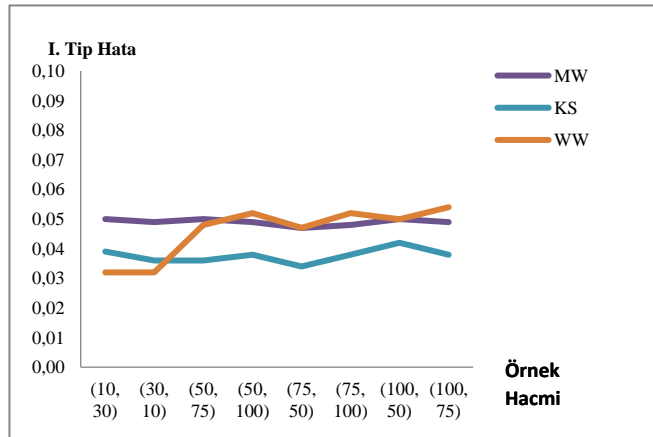
* $\alpha=0,05$ 'den küçük veya eşit olan değerler

Elde edilen verilere dayanarak, standart sapma oranı 1 olarak kabul edilen Uniform-Like, Logistic-Like ve Double Exponential-Like dağılımları için Mann-Whitney U testi, Kolmogorov-Smirnov testi ve Wald-Wolfowitz testi ile hesaplanan I. tip hataları, (10, 30), (30, 10), (50, 75), (50, 100), (75, 50), (75, 100), (100, 50) ve (100, 75) büyük ve farklı örneklem büyüklükleri arasında karşılaştırılmıştır.

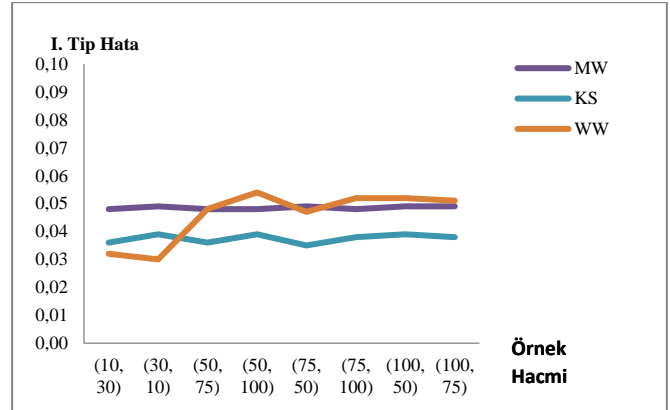
Mann-Whitney U testi için, standart sapma oranı 1 olduğunda I. tip hata oranları şu şekildedir: Uniform-Like: 0.048, Logistic-Like: 0.047, Double Exponential-Like: 0.048. Kolmogorov-Smirnov testi için, standart sapma oranı 1 olduğunda I. tip hata oranları şu şekildedir: Uniform-Like: 0.035, Logistic-Like: 0.034, Double Exponential-Like: 0.035. Wald-Wolfowitz testi için, standart sapma oranı 1 olduğunda I. tip hata oranları şu şekildedir: Uniform-Like: 0.032, Logistic-Like: 0.032, Double Exponential-Like: 0.030.



Şekil 10. Büyük ve farklı örnek büyüklüklerinde Uniform-Like dağılımı için standart sapma oranı 1 olduğunda Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz dizi sayıları testlerinin I. tip hata oranları



Şekil 11. Büyük ve farklı örnek büyüklüklerinde Logistic-Like dağılımı için standart sapma oranı 1 olduğunda Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz dizi sayıları testlerinin I. tip hata oranları



Şekil 12. Büyük ve farklı örnek büyüklüklerinde Double Exponential-Like dağılımı için standart sapma oranı 1 olduğunda Mann-Whitney, Kolmogorov-Smirnov ve Wald-Wolfowitz dizi sayıları testlerinin I. tip hata oranları

Tablo 5. Büyük örneklem durumlarında, örnek büyüklüklerinin farklı olduğu senaryolarda Mann-Whitney U testi, Kolmogorov-Smirnov testi ve Wald-Wolfowitz testinin I. tip hata oranları (Standart sapma oranı=1)

Popülasyon Dağılımı	n ₁	n ₂	I. Tip Hata Oranı		
			MW	KS	WW
Uniform-Like	10	30	0.051	0.039*	0.032*
	30	10	0.048*	0.037*	0.033*
	50	75	0.050*	0.039*	0.049*
	50	100	0.048*	0.040*	0.052
	75	50	0.050*	0.035*	0.047*
	75	100	0.050*	0.040*	0.053
	100	50	0.052	0.040*	0.054
	100	75	0.050*	0.039*	0.051
Logistic-Like	10	30	0.050*	0.039*	0.032*
	30	10	0.049*	0.036*	0.032*
	50	75	0.050*	0.036*	0.048*
	50	100	0.049*	0.038*	0.052
	75	50	0.047*	0.034*	0.047*
	75	100	0.048*	0.038*	0.052
	100	50	0.050*	0.042*	0.050*
	100	75	0.049*	0.038*	0.054
Double Exponential-Like	10	30	0.048*	0.036*	0.032*
	30	10	0.049*	0.039*	0.030*
	50	75	0.048*	0.036*	0.048*
	50	100	0.048*	0.039*	0.054
	75	50	0.049*	0.035*	0.047*
	75	100	0.048*	0.038*	0.052
	100	50	0.049*	0.039*	0.052
	100	75	0.049*	0.038*	0.051

* $\alpha=0,05$ 'den küçük veya eşit olan değerler

5. Sonuç

Çalışmamız, Uniform-Like, Logistic-Like ve Double Exponential-Like dağılımlarının standart sapma oranını 1 olarak kabul ederek, Mann-Whitney U testi, Kolmogorov-Smirnov testi ve Wald-Wolfowitz testi ile hesaplanan I. tip hataları karşılaştırmaktadır. Farklı örneklem büyüklükleri üzerinde yapılan bu karşılaştırma sonuçlarına dayanarak, Mann-Whitney testinin özellikle küçük örnek hacimlerinde diğer non-parametrik testlere göre daha yüksek I. tip hata oranlarına sahip olduğu gözlemlenmiştir. Ayrıca, büyük örneklem gruplarında da Mann-Whitney testinin I. tip hata oranlarının diğer testlere göre yüksek olduğu belirlenmiştir. Bu bulgular, özellikle Kolmogorov-Smirnov iki örnek testinin Mann-Whitney ve Wald-Wolfowitz dizi sayıları testlerine göre daha düşük I. tip hata oranlarına sahip olduğunu ortaya koymaktadır.

Özellikle Kolmogorov-Smirnov iki örnek testinin diğer testlere göre daha düşük I. tip hata oranlarına sahip olduğu sonucu, istatistiksel analizlerde tercih edilirken örnek büyüklüğü veya örnek hacmi fark etmeksizin dikkate alınması gerektiğini vurgular. Bu sonuçlar, bilimsel araştırmalarda doğru sonuçlara ulaşmak için uygun istatistiksel yöntemlerin seçilmesinde önemli bir rehber olabilir.

Literatürdeki benzer çalışmalara baktığımızda, Blair ve Higgins'ın (1985) çalışması, Wilcoxon işaretli sıralama testinin eşleştirilmiş örneklerdeki farkları tespit etme gücünü ve testin doğruluğunu değerlendirmiştir. Penfield (1994), örnek büyüklükleri ve varyanslar arasındaki ilişkiyi inceleyerek Mann-Whitney testinin eşit olmayan örnek büyüklükleri durumunda diğer testlere göre daha güçlü olduğunu belirtmiştir. Fahoome (1999), parametrik olmayan testlerin dağılımlardaki I. tip hata oranlarını değerlendirerek Kolmogorov-Smirnov testinin p değerinin nominal I. tip hata oranına yakınsız olduğunu, Mann-Whitney testinin ise daha iyi sonuçlar verdiğini göstermiştir. Bu çalışmada ise Mann-Whitney testinin özellikle küçük örnek hacimlerinde diğer testlere göre daha yüksek I. tip hata oranlarına sahip olduğunu ortaya koymuştur. Bu sonuçlar, istatistiksel analizlerde test seçiminde dikkatli olunması gerektiğini vurgulamış, Kolmogorov-Smirnov iki örnek testinin, diğer testlere kıyasla daha düşük bir I. tip hata oranına sahip olduğunu belirlemiştir. Çalışmanın sonuçları, istatistiksel test seçimi konusundaki bilinç ve doğruluk açısından araştırmacılara faydalı bir rehber sunmaktadır.

Yazar Katkı Oranı Beyanı

Veri, Sahib Ramazanov ve Ötügen Senger tarafından toplanmıştır. Analiz, Sahib Ramazanov ve Ötügen Senger tarafından gerçekleştirilmiştir. Literatür taraması, Sahib Ramazanov ve Ötügen Senger yapılmıştır. Sonuç ve tartışma bölümü yazarlar tarafından ortak olarak yazılmıştır.

Çatışma Beyanı

Çalışmada yazarlar arasında çıkar çatışması yoktur.

Destek Beyanı

Bu çalışma için herhangi bir kurumdan destek alınmamıştır.

Kaynaklar

Akbulut, M. C. (2008). *Faktöriyel düzende varyans analizi tekniğinde varyansların homojenliği ve normal dağılım ön şartları yerine gelmediğinde interaksiyona ilişkin I. tip hata ve testin gücü* [Yayımlanmamış Doktora Tezi]. Ankara Üniversitesi, Fen Bilimleri

- Enstitüsü.
- Bindak, R. (2014). Mann-Whitney U ile Student's t testinin I. Tip Hata ve Güç bakımından karşılaştırılması: Monte Carlo simülasyon çalışması (011302). *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 14(1), 5-11.
- Blair, R. C. ve Higgins, J. J. (1985). Comparison of the power of the paired samples t-test to that of Wilcoxon's sign-ranks test under various population shapes. *Psychological Bulletin*, 97 (1), 119-128.
- Boslaugh, S. ve Watters, P. A. (2008). *Statistics in a Nutshell*. O'Reilly Media, Inc.
- Brink, D. (2010). *Statistics*. Ventus Publishing ApS.
- Conover, W. J. (1999). *Practical Nonparametric Statistics* (Vol. 350). John Wiley & Sons.
- Daniel, W. W. (1990). Procedures That Utilize Data from Two Independent Samples. *Applied Nonparametric Statistics* (s. 82-143) içinde. PWS-Kent Publishing Co.
- Durand, C. P. (2013). Does raising type 1 error rate improve power to detect interactions in linear regression models? A simulation studies. *PloS one*, 8(8), e71079.
- Eltas, Ö. (2021). Biyoistatistik çalışmalarında kullanılan küçük örneklerde Mann-Whitney u testi ve bağımsız örneklem t (student's t independent test) testinin güç yönünden karşılaştırılması. *Atatürk Üniversitesi Veteriner Bilimleri Dergisi*, 16(1), 88-94.
- Fahoome, G. (1999). *A Monte Carlo study of twenty-one nonparametric statistics with normal and nonnormal data* [Yayımlanmamış Doktora Tezi]. Wayne State University.
- Geyer, C. J. (2001). Probability and statistics. <http://www.stat.umn.edu/geyer/old/5102/#notes> adresinden 20.10.2023 tarihinde alınmıştır.
- Gibbons, J. D. (1971). *Nonparametric Statistical Inference*. McGraw Hill.
- Lee, C. H. (2007). *A Monte Carlo Study of Two Nonparametric Statistics with Comparisons of Type I Error Rates and Power*. Oklahoma State University.
- Meaney, C. ve Moineddin, R. (2014). A Monte Carlo simulation study comparing linear regression, beta regression, variable-dispersion beta regression and fractional logit regression at recovering average difference measures in a two-sample design. *BMC Medical Research Methodology*, 14, 1-22.
- Mooney, C. Z. (1997). *Monte Carlo Simulation* (No. 116). Sage.
- Penfield, D. A. (1994). Choosing a two-sample location test. *Journal of Experimental Education*, 4(62), 343-350.
- Ramazanov, S. ve Askerbeyli, R. (2023). *Power Analysis of Mann Whitney, Two Sample Kolmogorov Smirnov, and Wald-Wolfowitz Runs Tests: Comparison with Monte Carlo Simulation in the Case of Variance Heterogeneity* (No. 10501). EasyChair.
- Senger, Ö. (2013). A statistical power comparison of the Kolmogorov-Smirnov two-sample test and the Wald-Wolfowitz test in terms of fixed skewness and fixed kurtosis in large sample sizes. *Chinese Business Review*, 12(7), 469-476.
- Senger, Ö. (2013). Statistical power comparisons for equal skewness different kurtosis and equal kurtosis different skewness coefficients in nonparametric tests. *Istanbul University Econometrics and Statistics e-Journal*, (18), 81-115.
- Shao, J. (2003). *Mathematical Statistics*. Springer Science & Business Media.
- Tekindal, B. (2007). Varyansların homojenliğinde kullanılan bazı yöntemlerin I. tip hata ve testin gücü bakımından karşılaştırılması: Bir Monte Carlo simülasyon çalışması. *Gazi Üniversitesi Endüstriyel Sanatlar Eğitimi Fakültesi Dergisi Sayı*, (20), 90-101.
- Uysal, İ. ve Kılıç, A. (2022). Normal dağılım ikilemi. *Anadolu Journal of Educational Sciences International*, 12(1), 220-248.
- Yiğit, E. ve Gamgam, H. (2011). Homojen olmayan varyans varsayımı altında ortalamaların eşitliği için bazı test istatistikleri ve karşılaştırmaları. *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, 1 (1), 57-71.
- Zimmerman, D. W. (1998). Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. *The Journal of Experimental Education*, 67(1), 55-68.