



# MerQur'da Non-Parametrik Testler: Dağılım-Bağımsız Hipotez Sınamasının Tam Repertuarı

## Non-Parametric Tests in MerQur: The Full Repertoire of Distribution-Free Hypothesis Testing

Ömer K. Örucü<sup>1</sup>

<sup>1</sup> Süleyman Demirel Üniversitesi, Mimarlık Fakültesi, Peyzaj Mimarlığı Bölümü, Isparta/Türkiye ORCID:

0000-0002-2162-7553 E-posta: omerorucu@sdu.edu.tr · Resmi site:

<https://www.sekizgenacademy.com/journals/index.php/merqur/tr/index>

**Yazışmadan sorumlu yazar:** Ömer K. Örucü (omerorucu@sdu.edu.tr)

**Tür:** Davetli Editöryal Sunum / Invited Editorial Showcase **Geliş:** 2026-05-17 · **Kabul:** 2026-05-17 · **Yayın:** 2026-05-17 **DOI:** — (ISSN başvurusu sonrası eklenecek)

### Öz

Non-parametrik (dağılım-bağımsız) testler, veri parametrik bir dağılımdan gelmediğinde, küçük örneklem büyüklüklerinde, sıralı (ordinal) ölçek ile ölçülmüş değişkenlerle çalışırken ya da uç değerlerin etkisini sınırlandırmak gerektiğinde başvurulan yöntemler ailesidir. Bu çalışmada **MerQur** masaüstü yazılımının Non-Parametrik Testler kategorisinde sunduğu **7 analiz** ayrıntılı olarak tanıtılmıştır: Mann-Whitney U testi, Wilcoxon işaretli sıra testi, Kruskal-Wallis H testi, Friedman testi, binomial test, sign (işaret) testi ve runs (dizi) testi. Her analiz için (i) test edilen hipotez ve uygulama bağlamı, (ii) parametrik karşılığı ile karşılaştırma, (iii) gerekli varsayımlar (sıralanabilirlik, simetri, bağımsızlık), (iv) MerQur'daki form alanları ve parametre seçenekleri, (v) raporlanan istatistikler ve etki büyüklükleri (rank-biserial korelasyon,  $\epsilon^2$ , Cliff's  $\delta$ ), ve (vi) tipik bir araştırma sorusu için yorumlama önerisi sunulmuştur. Non-parametrik testlerin parametrik karşılıklarına göre asimptotik göreceli etkinliği (ARE) tartışılmıştır: normal dağılım altında Mann-Whitney U'nun bağımsız t-test'e göre ARE'si yaklaşık 0.955'tir, ancak ağır kuyruklu dağılımlarda non-parametrik testler daha güçlü olabilir. MerQur'un çıktıları her analizde tam (exact) p-değeri seçeneği sunar ve süreklilik düzeltilmesi (Yates) gibi opsiyonları içerir. Sonuç olarak MerQur'un Non-Parametrik Testler kategorisi, varsayım ihlallerinde güvenilir ve doğru çıkarımsal araç seçeneğini Türkçe arayüz ile akademik araştırmacıya sunmaktadır.

**Anahtar Kelimeler:** non-parametrik testler, Mann-Whitney, Wilcoxon, Kruskal-Wallis, Friedman, sign test, runs test, etki büyüklüğü, MerQur

### Abstract

Non-parametric (distribution-free) tests are a family of methods used when data do not come from a parametric distribution, when sample sizes are small, when variables are measured on an ordinal scale, or

when the influence of outliers needs to be limited. This study introduces in detail the **7 analyses** offered under the Non-Parametric Tests category of **MerQur**: Mann-Whitney U test, Wilcoxon signed-rank test, Kruskal-Wallis H test, Friedman test, binomial test, sign test, and runs test. For each analysis: (i) the hypothesis tested and application context, (ii) comparison with its parametric counterpart, (iii) required assumptions (rankability, symmetry, independence), (iv) form fields and parameter options in MerQur, (v) reported statistics and effect sizes (rank-biserial correlation,  $\epsilon^2$ , Cliff's  $\delta$ ), and (vi) interpretation guidance for a typical research question. The asymptotic relative efficiency (ARE) of non-parametric tests relative to their parametric counterparts is discussed: under the normal distribution, the ARE of Mann-Whitney U relative to the independent t-test is approximately 0.955; however, under heavy-tailed distributions, non-parametric tests can be more powerful. MerQur's outputs provide exact p-value options where feasible and include options such as continuity correction (Yates). In conclusion, MerQur's Non-Parametric Tests category offers Turkish-localised, reliable, and correct inferential tools for situations where parametric assumptions are violated.

**Keywords:** non-parametric tests, Mann-Whitney, Wilcoxon, Kruskal-Wallis, Friedman, sign test, runs test, effect size, MerQur

## 1. Giriş

Parametrik testlerin ardındaki güçlü varsayım — verinin belirli bir parametrik dağılımdan (genellikle normal) geldiği — pek çok araştırma bağlamında sağlanmaz. Küçük örneklem büyüklükleri ( $n < 30$ ), aşırı çarpık dağılımlar, sıralı (Likert ölçek vb.) ölçümler, aşırı değerlerin (outlier) etkisi olabilecek veriler — bu durumların hepsinde parametrik testlerin güvenilirliği zayıflar. Non-parametrik testler, dağılım varsayımlarını minimum tutan veya tamamen ortadan kaldıran bir alternatif yöntem ailesidir (Conover, 1999; Hollander ve ark., 2014).

Non-parametrik testlerin temel mekanizması verinin **sıra değerlerine (rank)** dönüştürülmesidir; sıralar bir kez hesaplandığında orijinal ölçek bilgisinin çoğu kaybolur ama sıralı ilişkiler korunur. Bu yaklaşım sayesinde testler dağılım şeklinden büyük ölçüde bağımsızlaşır — bu nedenle “dağılım-bağımsız” (distribution-free) denir. Önemli bir kavramsal yanlış: non-parametrik testler “varsayimsız” değildir; yalnızca farklı (ve genellikle daha hafif) varsayımlar yapar (bağımsızlık, sıralanabilirlik, simetri vb.).

Bir başka önemli kavram **asimptotik göreceli etkinlik (ARE)** — non-parametrik testin parametrik rakibine karşı, aynı güce ulaşmak için ihtiyaç duyduğu örneklem büyüklüğünün oranı. Normal dağılım altında Mann-Whitney U'nun bağımsız t-testine karşı ARE'si yaklaşık 0.955; yani Mann-Whitney aynı gücü %4-5 daha büyük örneklemle elde eder. Buna karşılık ağır-kuyruklu dağılımlarda (laplace, t-distribution with low df) bu oran 1'i aşar — non-parametrik testler daha güçlü olur. Pratik sonuç: dağılım belirsizliğinde non-parametrik testler düşük “ücret” karşılığında güçlü güvence sağlar.

Bu çalışmanın amacı, **MerQur** masaüstü yazılımının Non-Parametrik Testler kategorisinde sunduğu 7 analizi ayrıntılı olarak tanıtmak; her birinin hangi araştırma sorusunu adresler, hangi parametrik karşılığı varsa onunla nasıl ilişkilendiğini ve MerQur arayüzünde nasıl uygulandığını göstermektir.

## 2. Mann-Whitney U Testi

### 2.1 Yöntem

Mann-Whitney U testi (eşdeğer: Wilcoxon sıra-toplam testi), iki bağımsız grubun **dağılımsal konumlarını** karşılaştırır. Bağımsız iki örneklem t-testinin non-parametrik karşılığıdır. Hipotezler:

- $H_0$ : İki grup aynı dağılımdan gelir (medyanlar eşit, dağılımlar çakışıyor)
- $H_1$ : Bir grup sistematik olarak diğerinden büyük/küçük değerler içerir

Test, iki grubun birleştirilmiş sıralarını hesaplayıp her gruba düşen sıra-toplamlar üzerinden U istatistiğini hesaplar:

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

Test istatistiği  $U = \min(U_1, U_2)$  veya bazen  $U_1$  kullanılır.

## 2.2 MerQur'da uygulama

*İstatistik → Non-Parametrik Testler → Mann-Whitney U*

. Form alanları:

- **Grup Sütunu:** iki kategorili grup değişkeni
- **Değer Sütunu:** sıralanabilen sayısal/sıralı değişken
- **Alternatif hipotez:** iki yönlü / büyük / küçük
- **Süreklilik düzeltmesi:** dahil/hariç (varsayılan dahil)
- **Tam p-değeri:** n küçükse ( $\leq 20$ ) exact, büyükse normal yaklaşıklık

Çıktıda U istatistiği, z skoru (büyük örneklem için), p-değeri, her grup için sıra ortalaması ve **rank-biserial korelasyon** etki büyüklüğü raporlanır.

## 2.3 Uygulama örneği

Bir kentsel ve kırsal bölgeden alınan 18 ağaç boyu örneğinin karşılaştırılması. Veri çarpık dağılımlı, t-test uygun değil.  $U = 92$ ,  $z = -2.31$ ,  $p = 0.021$ , rank-biserial  $r = 0.43$  (orta etki).

## 3. Wilcoxon İşaretli Sıra Testi

### 3.1 Yöntem

Wilcoxon işaretli sıra testi (signed-rank test), eşleştirilmiş t-testinin non-parametrik karşılığıdır. Aynı denekten alınan iki ölçümün farklarının medyanını sıfıra karşı test eder. Test, farkların mutlak değerlerini sıralar, ardından sıra-toplamlarına işaretlerini geri uygular.

Önemli varsayım: **farkların dağılımı simetrik** olmalıdır. Bu, fark dağılımının normal olmasını gerektirmez; medyana göre simetrik olması yeterlidir.

### 3.2 MerQur'da uygulama

*İstatistik → Non-Parametrik Testler → Wilcoxon İşaretli Sıra*

. Form alanları:

- **1. Ölçüm:** ilk ölçüm sütunu
- **2. Ölçüm:** ikinci ölçüm sütunu (aynı satır = aynı denek)
- **Alternatif hipotez:** iki yönlü / artış / azalış

- **Sıfır farklar:** dahil/hariç (Pratt veya Wilcoxon yöntemi)
- **Tam p-değeri:** n küçükse ( $\leq 25$ ) exact

Çıktıda W istatistiği, p-değeri, **matched-pairs rank-biserial r** etki büyüklüğü ve **Hodges-Lehmann tahmini** (medyan fark) sunulur.

### 3.3 Uygulama örneği

22 deneğin müdahale öncesi ve sonrası likert ölçek puanları. Veri ordinal — t-test uygun değil.  $W = 35$ ,  $p = 0.006$ , rank-biserial  $r = -0.58$  (büyük etki, müdahale lehine).

## 4. Kruskal-Wallis H Testi

### 4.1 Yöntem

Kruskal-Wallis H testi, üç veya daha fazla bağımsız grubun dağılım konumlarını karşılaştırır — tek yönlü ANOVA'nın non-parametrik karşılığıdır. Tüm gözlemler birleştirilip sıralanır; her grup için sıra-toplamı hesaplanır. Test istatistiği:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

ki-kare dağılımıyla  $k - 1$  serbestlik derecesinde değerlendirilir.

Anlamlı H bulgusu hangi grupların farklı olduğunu söylemez; ayrıntı için **Dunn post-hoc** veya **Conover-Iman post-hoc** testleri uygulanır.

### 4.2 MerQur'da uygulama

*İstatistik → Non-Parametrik Testler → Kruskal-Wallis*

. Form alanları:

- **Grup Sütunu:** kategorik (3+ seviye)
- **Değer Sütunu:** sıralanabilen sayısal/sıralı değişken
- **Post-hoc:** Dunn (Bonferroni düzeltmeli) / Conover-Iman / yok
- **Bağlı sıra düzeltmesi:** otomatik

Çıktıda H istatistiği, df, p-değeri, her grup için ortanca ve sıra ortalaması, **epsilon<sup>2</sup> ( $\epsilon^2$ )** etki büyüklüğü, ve post-hoc karşılaştırma matrisi sunulur.

### 4.3 Uygulama örneği

Üç farklı bölgeden toplanan toprak verimlilik puanları (sıralı ölçek).  $H(2) = 11.4$ ,  $p = 0.003$ ,  $\epsilon^2 = 0.18$  (orta etki). Dunn post-hoc: Bölge A > Bölge C ( $p_{adj} = 0.002$ ), B vs C anlamsız.

## 5. Friedman Testi

### 5.1 Yöntem

Friedman testi, aynı deneklerden 3+ koşul veya zaman noktasında alınan ölçümleri karşılaştırır — tekrarlı ölçümler ANOVA'nın non-parametrik karşılığıdır. Her denek içinde koşullar sıralanır (yani sıra hesabı denek içi yapılır). Test istatistiği:

$$\chi_F^2 = \frac{12}{nk(k+1)} \sum_{j=1}^k R_j^2 - 3n(k+1)$$

ki-kare dağılımıyla  $k - 1$  df'de değerlendirilir.

### 5.2 MerQur'da uygulama

*İstatistik → Non-Parametrik Testler → Friedman*

. Form alanları:

- **Ölçüm Sütunları:** her satırın aynı deneğe ait olduğu, koşul başına bir sütun
- **Post-hoc:** Nemenyi (eşit n ve k ile) / Conover

Çıktıda Friedman  $\chi^2$  istatistiği, df, p-değeri, **Kendall's W** uyum katsayısı (rater-eşgüdüm yorumu), her koşul için sıra ortalaması ve post-hoc karşılaştırmaları sunulur.

### 5.3 Uygulama örneği

10 değerlendirici, 4 farklı tasarımı sıraladı.  $\chi^2_{F(3)} = 14.3$ ,  $p = 0.003$ , Kendall's  $W = 0.48$  (orta-yüksek uyum). Nemenyi post-hoc: tasarım 4 > tasarım 1 (anlamlı).

## 6. Binomial Test

### 6.1 Yöntem

Binomial test, bir iki-kategorili (başarı/başarısızlık) değişkenin gözlenen oranının önceden belirlenmiş bir referans oranla (genellikle 0.5) uyumlu olup olmadığını test eder. Tam binomial olasılık formülü kullanır:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Tam p-değeri, gözlenen veya daha ekstrem sonuçların binomial dağılım altındaki kümülatif olasılığıdır.

### 6.2 MerQur'da uygulama

*İstatistik → Non-Parametrik Testler → Binomial Test*

. Form alanları:

- **Değişken:** iki kategorili sütun (veya başarı sayısı + toplam giriş)
- **Başarı kategorisi:** hangi değer "başarı" sayılacağı
- **Test edilen p:** referans oran (varsayılan 0.5)
- **Alternatif hipotez:** iki yönlü / büyük / küçük

Çıktıda gözlenen oran, %95 güven aralığı (Clopper-Pearson exact), p-değeri ve gözlenen vs beklenen frekans karşılaştırması raporlanır.

### 6.3 Uygulama örneği

20 hastadan 14'ü iyileşti. Beklenen oran %50; gözlenen  $14/20 = 0.70$ . Binomial p (iki yönlü) = 0.115. Anlamlı değil ama trend pozitif; daha büyük örneklem gerekli.

## 7. Sign (İşaret) Testi

### 7.1 Yöntem

Sign testi, eşleştirilmiş gözlemlerin ( $X_i, Y_i$ ) çiftlerinde  $X_i > Y_i$  sayısının binomial dağılım altında 0.5'e karşı test edilmesidir. Wilcoxon işaretli sıra testinin daha basit ama daha az güçlü versiyonudur — yalnızca farkların işaretini kullanır; büyüklüğü göz ardı eder. Fark dağılımı simetrik olmadığında veya sıralanması mümkün olmayan ordinal verilerde uygundur.

### 7.2 MerQur'da uygulama

*İstatistik → Non-Parametrik Testler → Sign Testi*

. Form alanları:

- **1. Ölçüm:** ilk ölçüm sütunu
- **2. Ölçüm:** ikinci ölçüm sütunu
- **Alternatif hipotez:** iki yönlü / artış / azalış

Çıktıda pozitif fark sayısı, negatif fark sayısı, sıfır fark sayısı (genellikle çıkarılır), exact binomial p-değeri ve oranın %95 GA'sı sunulur.

### 7.3 Uygulama örneği

15 kişiden 12'sinde uygulamadan sonra puan yükseldi. Sign testi: pozitif = 12, negatif = 3,  $p = 0.018$  (iki yönlü). Anlamlı yükselme.

## 8. Runs (Dizi) Testi

### 8.1 Yöntem

Runs testi, ikili veya ikili-dönüştürülmüş bir dizide gözlenen run (ardışık aynı değer dizisi) sayısının rastgelelik altında beklenen değerle uyumlu olup olmadığını test eder — yani **gözlem sıralamasının bağımsız** mı yoksa **bir örüntü içeren mi** olduğunu sorgular. Zaman serileri ön analizinde, kalıntıların rastgeleliği kontrolünde ve seri korelasyon tespitinde kullanılır.

### 8.2 MerQur'da uygulama

*İstatistik → Non-Parametrik Testler → Runs Testi*

. Form alanları:

- **Değişken:** sıralı (zaman/sırada) gözlenen sütun
- **İkilemek için kesme noktası:** medyan / ortalama / kullanıcı tanımlı
- **Alternatif hipotez:** iki yönlü / az run (kümeleşme) / çok run (alternasyon)

Çıktıda gözlenen run sayısı, beklenen run sayısı, z istatistiği, p-değeri ve dizinin görsel temsili (renkli + - işaretleri) sunulur.

### 8.3 Uygulama örneği

50 günlük üretim hata sayılarının medyana göre ikilemesi. Gözlenen run = 18, beklenen  $\approx 25.5$ ;  $z = -2.10$ ,  $p = 0.036$ . Üretim hatalarında **kümelenme** var — rastgelelik reddedilir.

## 9. Hangi Test Ne Zaman? Bir Karar Matrisi

Tablo 1, parametrik–non-parametrik testler arasındaki eşleşmeleri ve seçim kriterlerini özetler.

**Tablo 1.** Parametrik testler ve non-parametrik karşılıkları — seçim rehberi.

Araştırma sorusu	Parametrik	Non-Parametrik	Ne zaman non-parametrik?
Tek örneklemin referansla karşılaştırılması	Tek örneklem t	Sign testi (ortancaya karşı)	Çarpık dağılım, küçük n
İki bağımsız grup ortalama karşılaştırma	Bağımsız t	Mann-Whitney U	Normal değil, ordinal, küçük n
Eşleştirilmiş iki ölçüm	Eşleştirilmiş t	Wilcoxon işaretli sıra / Sign	Fark normal değil, simetrik değilse Sign
3+ bağımsız grup karşılaştırma	Tek yönlü ANOVA	Kruskal-Wallis	Normal değil, ordinal
3+ tekrarlı ölçüm karşılaştırma	RM-ANOVA	Friedman	Normal değil, küresellik sağlanmıyor
İki kategorili oran test	z testi (oranlar)	Binomial test	Küçük n, exact istenince
Dizi rastgeleliği	Durbin-Watson (regresyon)	Runs testi	Modelsiz sıralama testi

## 10. Karşılaştırmalı Değerlendirme

Tablo 2, MerQur'un Non-Parametrik Testler kategorisindeki 7 analiz için diğer GUI tabanlı açık erişimli alternatiflerle karşılaştırmasını özetler.

**Tablo 2.** Non-parametrik test ailesinin MerQur, JASP, jamovi ve PSPP'deki desteği.

Analiz	MerQur	JASP	jamovi	PSPP
Mann-Whitney U	✓	✓	✓	✓
Wilcoxon İşaretli Sıra	✓	✓	✓	✓
Kruskal-Wallis	✓	✓	✓	✓
Friedman	✓	✓	✓	+
Binomial Test	✓	✓	✓	✓
Sign Testi	✓	+	-	-
Runs Testi	✓	-	-	✓

MerQur'un ayrıştığı noktalar: (i) Sign ve Runs testlerinin doğrudan panel olarak yer alması, (ii) her testte exact p-değeri seçeneği, (iii) post-hoc karşılaştırma matrisinin görsel olarak sunulması, (iv) rank-biserial / Cliff's  $\delta / \epsilon^2$  etki büyüklüklerinin otomatik raporlanması.

**Sınırlılıklar:** (i) **Mood medyan testi** ve **Brown-Mood medyan testi** doğrudan panel olarak bulunmaz (Ki-Kare Bağımsızlık ile yaklaşık olarak yapılabilir); (ii) **Kolmogorov-Smirnov iki örneklem testi** Normallik kategorisinde tek örneklem KS olarak mevcuttur, iki örneklem için ayrı panel yoktur; (iii) **Anderson-Darling iki örneklem testi** yoktur — uygun olduğunda Mann-Whitney tercih edilir.

## 11. Sonuç

Bu çalışma, MerQur masaüstü yazılımının Non-Parametrik Testler kategorisinde sunulan 7 analizi sistemli olarak tanıtmıştır. Mann-Whitney'den Friedman'a, sign ve binomial gibi temel testlerden runs testi gibi sıralama-tabanlı diagnostik araçlara uzanan kapsam; parametrik varsayımların sağlanmadığı durumlarda araştırmacının güvenle başvurabileceği bir alternatif portföy sunar. Her testte exact p-değeri seçeneği, etki büyüklüğü hesaplaması (rank-biserial  $r$ ,  $\epsilon^2$ , Cliff's  $\delta$ , Kendall's  $W$ ) ve post-hoc karşılaştırmalar otomatik olarak sağlanır. MerQur, Türkçe akademik camianın özellikle küçük örneklemler ve ordinal ölçekli çalışmalarda doğru çıkarımsal araca komutsuz erişimini güçlendirir. Sonraki davetli editöryal sunumlarda **İlişki** ve **Regresyon** kategorileri ayrıntılı olarak incelenecektir.

## Beyanlar

**Etik Kurul Onayı:** Bu çalışma insan ya da hayvan denek içermediğinden etik kurul onayı gerektirmemiştir.

**Çıkar Çatışması:** Yazar, MerQur yazılımının geliştiricisidir. Bu durum dışında herhangi bir çıkar çatışması bildirmemiştir.

**Finansman:** Spesifik bir dış fon alınmamıştır.

**Veri ve Kod Erişim Beyanı:** Bu derleme orijinal araştırma verisi içermez. MerQur yazılımı <https://merqur.sdu.edu.tr> adresinden ücretsiz indirilebilir.

**Yapay Zekâ Kullanımı:** Bu makalenin yazımı sırasında üretken yapay zekâ (Claude, Anthropic) dil ve yapı düzeltmesi amacıyla destekleyici olarak kullanılmıştır. Bilimsel içerik yazarın özgün katkısıdır.

**Yazar Katkı Beyanı (CRediT):** Ömer K. Örcü — Kavramsallaştırma, Yöntem, Yazılım, Doğrulama, Yazma (orijinal taslak), Yazma (gözden geçirme & düzenleme).

## Kaynakça

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions.

---

*Psychological Bulletin*

, 114(3), 494–509. <https://doi.org/10.1037/0033-2909.114.3.494>

Conover, W. J. (1999).

---

*Practical nonparametric statistics*

(3rd ed.). Wiley.

Conover, W. J., & Iman, R. L. (1979). On multiple-comparisons procedures (Tech. Rep. LA-7677-MS). Los Alamos Scientific Laboratory.

Dunn, O. J. (1964). Multiple comparisons using rank sums.

---

*Technometrics*

, 6(3), 241–252. <https://doi.org/10.2307/1266041>

Fagerland, M. W., Lydersen, S., & Laake, P. (2017).

---

*Statistical analysis of contingency tables*

. Chapman & Hall/CRC.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance.

---

*Journal of the American Statistical Association*

, 32(200), 675–701. <https://doi.org/10.2307/2279372>

Hodges, J. L., & Lehmann, E. L. (1963). Estimates of location based on rank tests.

---

*Annals of Mathematical Statistics*

, 34(2), 598–611. <https://doi.org/10.1214/aoms/1177704172>

Hollander, M., Wolfe, D. A., & Chicken, E. (2014).

---

*Nonparametric statistical methods*

(3rd ed.). Wiley.

Kendall, M. G. (1948).

---

*Rank correlation methods*

. Griffin.

Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation.

---

*Comprehensive Psychology*

, 3, 1–9. <https://doi.org/10.2466/11.IT.3.1>

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis.

---

*Journal of the American Statistical Association*  
, 47(260), 583–621. <https://doi.org/10.2307/2280779>

Lehmann, E. L. (1975).

---

*Nonparametrics: Statistical methods based on ranks*  
. Holden-Day.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other.

---

*Annals of Mathematical Statistics*  
, 18(1), 50–60. <https://doi.org/10.1214/aoms/1177730491>

Nemenyi, P. (1963).

---

*Distribution-free multiple comparisons*  
[Doctoral dissertation, Princeton University].

Pratt, J. W. (1959). Remarks on zeros and ties in the Wilcoxon signed rank procedures.

---

*Journal of the American Statistical Association*  
, 54(287), 655–667. <https://doi.org/10.2307/2282543>

Sheskin, D. J. (2011).

---

*Handbook of parametric and nonparametric statistical procedures*  
(5th ed.). Chapman & Hall/CRC.

Siegel, S., & Castellan, N. J. (1988).

---

*Nonparametric statistics for the behavioral sciences*  
(2nd ed.). McGraw-Hill.

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited: An overview of some recommended measures of effect size.

---

*Trends in Sport Sciences*  
, 1(21), 19–25.

Wald, A., & Wolfowitz, J. (1940). On a test whether two samples are from the same population.

---

*Annals of Mathematical Statistics*  
, 11(2), 147–162. <https://doi.org/10.1214/aoms/1177731909>

Wilcoxon, F. (1945). Individual comparisons by ranking methods.

---

---

*Biometrics Bulletin*

, 1(6), 80–83. <https://doi.org/10.2307/3001968>

Zimmerman, D. W. (2003). A warning about the large-sample Wilcoxon-Mann-Whitney test.

---

*Understanding Statistics*

, 2(4), 267–280. [https://doi.org/10.1207/S15328031US0204\\_03](https://doi.org/10.1207/S15328031US0204_03)

---

*Bu makale "Davetli Editöryal Sunum" bölümü kapsamında yayımlanmıştır. Bölüm politikası gereği harici hakem değerlendirmesinden geçmemiş, MerQur Veri Bilimi ve Yöntemleri Dergisi Yayın Kurulu tarafından editöryal incelemeye tabi tutulmuştur. Bu makale Creative Commons Atıf 4.0 Uluslararası (CC-BY 4.0) lisansı altında yayımlanmıştır.*