OKÜ Fen Bilimleri Enstitüsü Dergisi
6(3): 2106-2122, 2023

OKU Journal of The Institute of Science and Technology, 6(3): 2106-2122, 2023

Osmaniye Korkut Ata Üniversitesi
Fen Bilimleri Enstitüsü
Dergisi

Osmaniye Korkut Ata University
Journal of The Institute of Science
and Technology

İki Örneklem Behrens-Fisher Problemi İçin Farklı İstatistiksel Test Yöntemlerinin Karşılaştırılması

Esin BUDAK[1*], Zeki YILDIZ[2], Mehmet SANDAL[3]

[1]Eskişehir Osmangazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, 26040, Eskişehir
[2]Eskişehir Osmangazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, 26040, Eskişehir
[3]Manisa Celal Bayar Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, 45140, Manisa

[1] https://orcid.org/0000-0002-9016-3385
[2] https://orcid.org/0000-0003-1907-2840
[3] https://orcid.org/0000-0001-7396-0801
*Sorumlu yazar: esin01karahan@hotmail.com

Araştırma Makalesi

Makale Tarihçesi:
Geliş tarihi: 24.03.2022
Kabul tarihi:20.12.2022
Online Yayınlanma: 04.12.2023

Anahtar Kelimeler:
Hotelling'in T^2 istatistiği
Çok değişkenli Behrens-Fisher problemi
Simülasyon
Tip I hata

ÖZ

Çok değişkenli istatistiksel araştırmalarda, iki ortalama vektörün eşitliğini test etme problemi ile sıklıkla ilgilenilmektedir. Ancak varsayımlar ihlal edildiğinde klasik yöntemlerin kullanılması yanıltıcı sonuçlar elde edilmesine neden olabilmektedir. Bu çalışmanın amacı da iki örneklem Behrens-Fisher problemleri için önerilen test istatistiklerini I. tip hata olasılıkları bakımından karşılaştırmaktır. Bu amaçla varsayımların ihlal edilmesi durumunda iki grup ortalama vektörünün eşitliğini test etmek için literatürde önerilen test istatistikleri bir simülasyon çalışması ile karşılaştırılmıştır. Ayrıca önerilen test istatistiklerinin gerçek bir veri örneği üzerinde karşılaştırılması yapılmıştır. Çalışmanın sonuçları, test istatistiklerinin performanslarının bağımlı değişken sayısına ve gözlem büyüklüklerine göre değiştiğini göstermiştir. Ancak Yanagihara ve Yuan (2005) tarafından önerilen test istatistiğinin oldukça iyi bir performans ortaya koyduğu görülmüştür. Ayrıca Hotelling T^2 test istatistiğinin varsayım ihlallerinden oldukça fazla etkilendiği gözlemlenmiştir.

Comparison of Different Statistical Test Methods for Two Samples Behrens-Fisher Problem

Research Article

Article History:
Received: 24.03.2022
Accepted: 20.12.2022
Published online: 04.12.2023

Keywords:
Hotelling's T-squared statistic
Multivariate Behrens-Fisher problem
Simulation
Type I error

ABSTRACT

In multivariate statistical research, the problem of testing the equality of two mean vectors is often dealt with. However, when the assumptions are violated, the use of classical methods can lead to misleading results. The aim of this study is to compare the proposed test statistics for two sample Behrens-Fisher problems in terms of their probability of type I error. For this purpose, test statistics proposed in the literature are compared with a simulation study to test the equality of the two group mean vectors in case the assumptions are violated. In addition, the proposed test statistics were compared on a real data sample. The results of the study showed that the performances of the test statistics vary according to the number of dependent variables and the size of the observations. However, it was seen that the test statistics proposed by Yanagihara and Yuan (2005) performed quite well. In addition, Hotelling T^2 test statistics is highly affected by assumption violations.

To Cite: Budak E., Yıldız Z., Sandal M. İki Örneklem Behrens-Fisher Problemi İçin Farklı İstatistiksel Test Yöntemlerinin Karşılaştırılması. Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi 2023; 6(3): 2106-2122.

1. Giriş

İstatistiksel çalışmalarda, araştırma problemi için uygulanacak analiz yaklaşımının belirlenmesi oldukça önemlidir. Analiz yöntemi doğru seçilmediği takdirde yanıltıcı sonuçlar elde edilebilmektedir. Bu nedenle çeşitli varsayımlara dayalı olarak geliştirilen parametrik testler kullanıldığında da anlamlı sonuçlar elde etmek için ilgili varsayımların sağlanması gerekmektedir. İstatistiksel yönteme ilişkin varsayım sağlanmadığı takdirde yeni çözümleme yöntemleri geliştirilmeli ya da alternatif teknikler kullanılmalıdır.

Bağımlı değişken sayısının tek olduğu çalışmalarda iki ve daha fazla ortalamanın karşılaştırılmasında tek değişkenli istatistiksel yöntemler kullanılmaktadır. Ancak bağımlı değişken sayısının birden fazla olduğu durumlarda ise çok değişkenli istatistiksel yöntemlere başvurulması gerekmektedir. Çok değişkenli istatistiksel yöntemlerde, değişkenlerin ve birimlerin karşılıklı ilişkileri de söz konusu olmaktadır. Bu nedenle karşılıklı ilişkileri de dikkate alan çok değişkenli istatistiksel tekniklerin kullanılmasına ihtiyaç duyulmaktadır (Budak, 2021).

Parametrik analiz yöntemleri arasında en sık başvurulan yöntemlerin başında bağımsız grup ortalamalarının karşılaştırmalarına ilişkin hipotezleri sınama problemleri karşımıza çıkmaktadır (Sandal, 2020). İki anakütle ortalamasının karşılaştırılması ise araştırmacılar tarafından ilgi duyulan konuların başında gelmektedir. Tek değişkenli grup ortalamalarının ya da çok değişkenli gruplara ait ortalama vektörlerinin karşılaştırılma probleminde genellikle değişkenler için en temel varsayımların başında normallik varsayımı ile varyansların ya da varyans-kovaryans matrislerinin eşit olma koşulları gelmektedir. Tek değişkenli problemler için varyansların ve çok değişkenli problemler için varyans-kovaryans matrislerinin eşit olmaması durumunda, test istatistiklerinin I. tip hata olasılığı anlamlılık düzeyinden uzaklaşmaktadır. Varyansların ve varyans-kovaryans matrislerinin eşit olmaması durumunda grup ortalamaları ya da ortalama vektörleri arasındaki farkın anlamlılığını test etme problemi ise "Behrens (1929)-Fisher (1935) Problemi" olarak adlandırılmaktadır.

Fisher (1935), Welch (1947) ve Wald (1955) başta olmak üzere Behrens-Fisher problemlerinin çözümü için çeşitli istatistiksel yöntemler geliştirilmiştir. Ancak ele alınan grup sayısı ve bağımlı değişken sayısına bağlı olarak literatürde farklı analiz yöntemleri önerilmiştir. Welch 'in (1947) geliştirdiği Welch-t testi, asimptotik olarak oldukça güçlü bir test istatistiğidir. Wald (1955) istatistiği ise yalnızca gözlem sayılarının eşit olduğu durumlarda kullanılabilmektedir (Pfanzagl, 1974). Tek değişkenli Behrens-Fisher problemleri için Scheffe (1943) tarafından önerilen test istatistiği, Bennett (1951) tarafından çok değişkenli durumlara genelleştirilmiştir. Ayrıca James (1954), Yao (1965), Subrahmaniam ve Subrahmaniam (1973), Johansen (1980), Nel ve Van der Merwe (1986), Kim (1992), Christensen ve Rencher (1997), Krishnamoorthy ve Yu (2004), Yanagihara ve Yuan (2005), Zezula (2009), Krishnamoorthy ve Yu (2012), Kawasaki ve Seo (2015) ile Erdoğan (2018) tarafından yaklaşık çözümler önerilmiştir.

Bu çalışmanın amacı; varyans-kovaryans matrislerinin eşitliği varsayımı sağlanmadığında farklı deneysel koşullara göre iki ortalama vektörü arasındaki farkın anlamlı olup olmadığını belirlemek için

Yao (1965), Johansen (1980), Nel ve Van der Merwe (1986), Kim (1992), Krishnamoorthy ve Yu (2004), Yanagihara ve Yuan (2005) ile Kawasaki ve Seo (2015) tarafından önerilen test istatistiklerini I. tip hata olasılıkları bakımından karşılaştırmaktır. Böylece varyans-kovaryans matrislerinin eşit olmadığı problemler için veri setine ait istatistiksel özellikler dikkate alınarak daha yüksek performansa sahip test istatistiğinin kullanılması amaçlanmıştır. Bu amaçla çalışmanın ikinci bölümünde çok değişkenli iki örneklem Behrens- Fisher probleminin çözümü için önerilen bazı alternatif test istatistikleri açıklanmıştır. Üçüncü bölümde farklı deneysel koşullara göre önerilen test istatistiklerinin performansını belirlemek için bir simülasyon çalışması gerçekleştirilmiştir. Son bölümde ise test istatistikleri, gerçek bir veri setine uygulanarak test istatistikleri arasındaki anlamlılık sonuçlarına ait farklılıklar karşılaştırılmıştır.

2. Materyal ve Metot

Çok değişkenli iki grup ortalama vektörleri arasındaki farkın anlamlı olup olmadığını belirlemek için $i = 1{,}2$ ve p boyutlu iki grubun ortalamalar vektörü μ_i olmak üzere,

$$\begin{aligned} &H_0: \mu_1 = \mu_2 \\ &H_1: \mu_1 \neq \mu_2 \end{aligned} \tag{1}$$

eşitliğindeki sıfır hipotezi sınanmaktadır. Anakütle varyans-kovaryans matrislerinin eşit olduğu durumlarda bu hipotezi sınamak için Hotelling T^2 test istatistiğinden yararlanılmaktadır. Böylece $i = 1{,}2$ ve $j = 1{,}2, \dots, n_i$ olmak üzere $N_p(\mu_i, \Sigma_i)$ dağılımına sahip p boyutlu X_{ij} gözlem vektörleri için Hotelling T^2 test istatistiği,

$$T_H^2 = (\bar{x}_1 - \bar{x}_2)' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_p\right]^{-1} (\bar{x}_1 - \bar{x}_2) \sim \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{sd1,sd2,(1-\alpha)} \tag{2}$$

biçimindedir (Hotelling, 1931). Burada i. örneklemin $p \times 1$ boyutlu ortalamalar vektörü

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \tag{3}$$

ve ortak varyans-kovaryans matrisi ise

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2} \tag{4}$$

olarak ifade edilmektedir. Ayrıca i. örneklemin varyans-kovaryans matrisi

$$S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)' \tag{5}$$

eşitliğindeki gibi hesaplanmaktadır. Eşitlik 2'de verilen T_H^2 istatistiği $sd_1 = p$ ve $sd_2 = n_1 + n_2 - p - 1$ serbestlik dereceli F dağılımı ile karşılaştırılmaktadır. Eğer test istatistiği $F_{sd1,sd2,(1-\alpha)}$ kritik değerinden daha büyük ise sıfır hipotezi reddedilmekte ve gruplar arasındaki farkın anlamlı olduğuna karar verilmektedir (Hotelling, 1931). Ancak varyans-kovaryans matrislerinin homojen olmadığı durumlarda ise Hotelling T^2 istatistiği için yaklaşık bir dağılım belirlenememektedir. Bu durumda

$$T^2 = (\bar{x}_1 - \bar{x}_2)' \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1} (\bar{x}_1 - \bar{x}_2) \tag{6}$$

eşitliğindeki test istatistiği dikkate alınarak Yao (1965), Johansen (1980), Nel ve Van der Merwe (1986), Kim(1992), Krishnamoorthy ve Yu (2004), Yanagihara ve Yuan (2005) ile Kawasaki ve Seo (2015) tarafından yeni çözümleme yöntemleri önerilmiştir.

2.1. **Yao Testi**

Welch (1947), tek değişkenli Behrens-Fisher problemleri için t dağılımına dayanan yeni bir test istatistiği önermiş ve bu istatistiğin kritik değerini belirlemek için seri açılımı ve yaklaşık serbestlik derecesi olmak üzere iki farklı yöntem geliştirmiştir. Yao (1965) ise Welch (1947)'in yaklaşık serbestlik derecesi kavramını çok değişkenli iki örneklem Behrens-Fisher problemleri için genelleştirilmiştir. Bu durumda $W_i = {S_i}/{n_i}$, $W = \sum_{i=1}^{2} W_i$ ve

$$v_{Yao} = \left[\sum_{i=1}^{2} \frac{1}{n_i - 1}\left(\frac{(\bar{x}_1 - \bar{x}_2)'(W)^{-1}(W_i)(W)^{-1}(\bar{x}_1 - \bar{x}_2)}{(\bar{x}_1 - \bar{x}_2)'(W)^{-1}(\bar{x}_1 - \bar{x}_2)}\right)^2\right]^{-1} \tag{7}$$

olmak üzere Yao'nun (1965) test istatistiği, Eşitlik 6'da verilen T^2 istatistiğine dayalı olarak;

$$T_{Yao} = \frac{v_{Yao} - p + 1}{v_{Yao} * p} T^2 \sim F_{p, v_{Yao} - p + 1} \tag{8}$$

eşitliği yardımıyla hesaplanmaktadır. Eğer T_{Yao} test istatistiği $F_{p,v_{Yao}-p+1}$ kritik değerinden daha büyük ise Eşitlik 1'deki sıfır hipotezi reddedilmektedir (Yao, 1965).

2.2. **Johansen Testi**

Johansen (1980) çok değişkenli Behrens-Fisher problemleri için yaklaşık serbestlik derecesi kavramını kullanarak test istatistiğinin olasılık dağılımına yakınsamayı amaçlamıştır. Eşitlik 1'deki sıfır hipotezini sınamak için Johansen (1980) testinde,

$$c = p + 2D - \frac{6D}{p + 2} \tag{9}$$

olmak üzere T^2 test istatistiği c sabitine oranlamakta ve

$$T_{Joh} = \frac{T^2}{c} \tag{10}$$

eşitliği $sd_1 = p$ ve $sd_2 = \frac{p(p+2)}{3D}$ serbestlik dereceli F dağılımına yakınsamaktadır. Burada D değeri;

$$D = \frac{1}{2}\sum_{i=1}^{2}\frac{1}{n_i - 1}\left\{tr\left[\left(I - (W_1^{-1} + W_2^{-1})^{-1}W_i^{-1}\right)\right]^2 + \left[tr\left(I - (W_1^{-1} + W_2^{-1})^{-1}W_i^{-1}\right)\right]^2\right\} \tag{11}$$

eşitliği yardımıyla hesaplanmaktadır. Eğer Johansen test istatistiği $T_{Joh} > F_{sd_1,sd_2,1-\alpha}$ olacak şekilde $F_{sd_1,sd_2,1-\alpha}$ kritik değerinden daha büyük ise yokluk hipotezi reddedilmektedir.

2.3. **Nel ve Van der Merwe Testi**

Nel ve Van der Merwe (1986), Welch (1947)'in yaklaşık serbestlik derecesi yöntemine dayanan farklı bir test istatistiği geliştirmişlerdir. Nel ve Van Der Merwe (1986), Eşitlik 6'daki test istatistiğinin dağılımına yakınsamak için,

$$v_{NV} = \frac{iz((W)^2) + \left(iz(W)\right)^2}{\sum_{i=1}^{2}\frac{1}{n_i-1}\left[iz((W_i)^2) + \left(iz(W_i)\right)^2\right]} \tag{12}$$

olmak üzere yeni bir test istatistiği olarak,

$$T_{NV} = \frac{v_{NV} - p + 1}{v_{NV} * p}T^2 \sim F_{p,v_{NV}-p+1} \tag{13}$$

eşitliğini önermişlerdir. Eğer T_{NV} test istatistiği $F_{p,v_{NV}-p+1}$ kritik değerinden daha büyük ise yokluk hipotezi reddedilmektedir (Nel ve Van Der Merwe, 1986).

2.4. **Kim Testi**

Welch (1947)'in yaklaşık serbestlik derecesinin farklı bir genellemesi de Kim (1992) tarafından gerçekleştirilmiştir. Kim (1992), iki ortalama vektör için güven elipsoidlerinin geometrisine dayanan yeni bir yaklaşım önermiştir. Kim (1992) tarafından sunulan test istatistiği;

$$T_{Kim} = \frac{v_K - p + 1}{c_K * f_K * v_K}(\bar{x}_1 - \bar{x}_2)W_K^{-1}(\bar{x}_1 - \bar{x}_2) \sim F_{f_K,v_K-p+1} \tag{14}$$

şeklindedir. Eşitlik 14'te yer alan Γ_K, Z_j, c_K, f_K ve W_K değerlerinin hesaplanması için sırasıyla aşağıdaki eşitliklerden yararlanılmaktadır.

$$\Gamma_K = \left(det\left(W_1 W_2{}^{-1}\right)\right)^{1/2p}, \tag{15}$$

$$Z_j = \frac{(e_j+1)}{\left(\mathrm{e_j}^{1/2}+\Gamma_K\right)^2} \ , \tag{16}$$

$$\mathrm{c_K} = \sum_{j=1}^{p} Z_j^2 \Big/ \sum_{j=1}^{p} Z_j \ , \tag{17}$$

$$\mathrm{f_K} = \left(\sum_{j=1}^{p} Z_j\right)^2 \Big/ \sum_{j=1}^{p} Z_j^2 \ , \tag{18}$$

$$\mathrm{W_K} = \mathrm{W_1} + \Gamma_K{}^2\mathrm{W_2} + 2\Gamma_K\mathrm{W_2}{}^{1/2}\left(\mathrm{W_2}{}^{-1/2}\mathrm{W_1}\mathrm{W_2}{}^{-1/2}\right)^{1/2}\mathrm{W_2}{}^{1/2} \ . \tag{19}$$

Burada $j = 1, \dots, p$ olmak üzere e_j değerleri $\mathrm{W_1}\mathrm{W_2}{}^{-1}$ ifadesinin j. özdeğerini göstermektedir. Ayrıca $\mathrm{v_K}$ ise Yao (1965) tarafından oluşturulan $\mathrm{v_{Yao}}$ değerine eşittir. Böylece $\mathrm{T_{Kim}}$ test istatistiği $\mathrm{F}_{f_K,\mathrm{v_K-p+1}}$ kritik değerinden daha büyük ise yokluk hipotezi reddedilmektedir (Kim, 1992; Coombs, Algina ve Oltman, 1996; Lix ve Keselman, 2004).

2.5. **Değiştirilmiş Nel ve Van der Merwe (MNV) Testi**

Krishnamoorthy ve Yu (2004), farklı bir serbestlik derecesi dikkate alarak Nel ve Van Der Merwe'nin (1986) test istatistiğini yeniden düzenlemiştir. Böylece;

$$\mathrm{v_{MNV}} = \frac{\mathrm{p} + \mathrm{p}^2}{\sum_{\mathrm{i}=1}^{2} \frac{1}{\mathrm{n_i}-1}\left[\mathrm{iz}((\mathrm{W_i}\mathrm{W}^{-1})^2) + \left(\mathrm{iz}(\mathrm{W_i}\mathrm{W}^{-1})\right)^2\right]} \tag{20}$$

olmak üzere, Krishnamoorthy ve Yu'nun (2004) test istatistiği, Eşitlik 6'daki T^2 istatistiğine dayalı olarak;

$$\mathrm{T_{MNV}} = \frac{\mathrm{v_{MNV}} - \mathrm{p} + 1}{\mathrm{v_{MNV}} * \mathrm{p}}\mathrm{T}^2 \sim \mathrm{F}_{\mathrm{p},\mathrm{v_{MNV}}-\mathrm{p}+1} \tag{21}$$

eşitliği yardımıyla hesaplanmaktadır. Eğer $\mathrm{T_{MNV}}$ test istatistiği $\mathrm{F}_{\mathrm{p},\mathrm{v_{MNV}}-\mathrm{p}+1}$ kritik değerinden daha büyük ise yokluk hipotezi reddedilmektedir (Krishnamoorthy ve Yu, 2004).

2.6. **F Yaklaşımı Testi**

Yanagihara ve Yuan (2005), çok değişkenli Behrens- Fisher problemlerinde T^2'nin olasılık dağılımına yakınsamak için bir F dağılımından yararlanmışlar ve Welch'in (1938) test yaklaşımını çok değişkenli örneklemler için genellemişlerdir. Bu durumda;

$$\hat{\theta}_1 = \frac{\mathrm{p}\hat{\eta}_1+(\mathrm{p}-2)\hat{\eta}_2}{\mathrm{p}(\mathrm{p}+2)} \quad , \quad \hat{\theta}_2 = \frac{\hat{\eta}_1+2\hat{\eta}_2}{\mathrm{p}(\mathrm{p}+2)} \tag{22}$$

ve

$$v_{YY} = \frac{\left(n_1 + n_2 - 2 - \hat{\theta}_1\right)^2}{(n_1 + n_2 - 2)\hat{\theta}_2 - \hat{\theta}_1} \tag{23}$$

olmak üzere Yanagira ve Yuan'ın (2005) F dağılımına bağlı test istatistiği;

$$T_{YY} = \frac{\left(n_1 + n_2 - 2 - \hat{\theta}_1\right)}{(n_1 + n_2 - 2) * p} T^2 \sim F_{p,v_{YY}} \tag{24}$$

şeklinde ifade edilmektedir. Burada $S_W = \frac{n_2}{n_1+n_2} S_1 + \frac{n_1}{n_1+n_2} S_2$ olmak üzere $\hat{\eta}_1$ ve $\hat{\eta}_2$ parametrelerinin tahmincileri sırasıyla;

$$\hat{\eta}_1 = \frac{n_2^2(n_1+n_2-2)}{n^2(n_1-1)} \left(iz\left(S_1 S_W{}^{-1}\right)\right)^2 + \frac{n_1^2(n_1+n_2-2)}{n^2(n_2-1)} \left(iz\left(S_2 S_W{}^{-1}\right)\right)^2, \tag{25}$$

$$\hat{\eta}_2 = \frac{n_2^2(n_1 + n_2 - 2)}{n^2(n_1 - 1)} \left(iz\left(S_1 S_W{}^{-1} S_1 S_W{}^{-1}\right)\right) + \frac{n_1^2(n_1 + n_2 - 2)}{n^2(n_2 - 1)} \left(iz\left(S_2 S_W{}^{-1} S_2 S_W{}^{-1}\right)\right) \tag{26}$$

eşitlikleri yardımıyla hesaplanmaktadır. Eğer T_{YY} test istatistiği $F_{p,v_{YY}}$ kritik değerinden daha büyük ise yokluk hipotezi reddedilmektedir (Yanagira ve Yuan, 2005; Kawasaki ve Seo, 2015).

2.7. **Düzeltilmiş Bartlett (MB) Testi**

Yanagira ve Yuan (2005) çalışmalarında, Fujiskoshi'nin (2000) Düzeltilmiş Bartlett (MB) yaklaşımını çok değişkenli durumlara genelleyerek çok değişkenli Behrens-Fisher problemleri için farklı bir test istatistiği daha önermişlerdir. Yanagira ve Yuan (2005) MB test istatistiğini;

$$T_{MB} = \left((n_1 + n_2 - 2)\hat{\beta}_1 + \hat{\beta}_2\right) \log\left(1 + \frac{T^2}{(n_1 + n_2 - 2)\hat{\beta}_1}\right) \sim X_p^2 \tag{27}$$

biçiminde tanımlamışlardır. Burada;

$$\hat{\xi}_1 = \frac{\hat{\eta}_1 + \hat{\eta}_2}{p} \tag{28}$$

ve

$$\hat{\xi}_2 = \frac{2(p + 3)\hat{\eta}_1 + 2(p + 4)\hat{\eta}_2}{p(p + 2)} \tag{29}$$

olmak üzere Eşitlik 27'deki test istatistiğinde yer alan $\hat{\beta}_1$ ve $\hat{\beta}_2$ parametrelerinin tahminleri ise sırasıyla;

$$\hat{\beta}_1 = \frac{2}{\hat{\xi}_2 - 2\hat{\xi}_1}\ , \tag{30}$$

$$\hat{\beta}_2 = \frac{(p+2)\hat{\xi}_2 - 2(p+4)\hat{\xi}_1}{2(\hat{\xi}_2 - 2\hat{\xi}_1)} \tag{31}$$

şeklinde hesaplanmaktadır. Eşitlik 27'de verilen T_{MB} test istatistiği p serbestlik dereceli X_p^2 kritik değerinden daha büyük ise yokluk hipotezi reddedilmektedir (Yanagira ve Yuan, 2005).

2.8. **Yanlılık Düzeltme (BC) ve İkinci Sıra (S) Testleri**

Kawasaki ve Seo (2015), F dağılımının serbestlik derecelerini düzelterek çok değişkenli Behrens-Fisher problemi için "Yanlılık Düzeltme Yöntemi (BC)" ve "İkinci Sıra Yöntemi (S)" olmak üzere iki farklı yaklaşık çözüm önermiştir. $N = n_1 + n_2 - 2$ olmak üzere

$$\hat{v}_{BC} = \frac{2(N^2 - N\theta_1 + \theta_2 - \theta_3 - \theta_1^*)^2}{N^2(N^2 - 2N\theta_1 + 2N\theta_4 + 2\theta_5 - \theta_6 - 2\theta_1^* + 2\theta_4^*) - (N^2 - N\theta_1 + \theta_2 - \theta_3 - \theta_1^*)^2} \tag{32}$$

ve

$$\hat{\phi}_{BC} = \frac{N^2 v_{BC}}{N^2 - N\theta_1 + \theta_2 - \theta_3 - \theta_1^*} \tag{33}$$

için BC yöntemine ait test istatistiği;

$$T_{BC} = \frac{\hat{v}_{BC}}{p\hat{\phi}_{BC}} T \sim F_{p,\hat{v}_{BC},(1-\alpha)} \tag{34}$$

biçimindedir. Benzer şekilde S yöntemine ait test istatistiği ise

$$\hat{v}_s = \frac{2(N^2 - N\theta_1 + \theta_2 - \theta_3)^2}{N^2(N^2 - 2N\theta_1 + 2N\theta_4 + 2\theta_5 - \theta_6) - (N^2 - N\theta_1 + \theta_2 - \theta_3)^2} \tag{35}$$

ve

$$\hat{\phi}_s = \frac{N^2 v_s}{N^2 - N\theta_1 + \theta_2 - \theta_3} \tag{36}$$

olmak üzere

$$T_s = \frac{\hat{v}_s}{p\hat{\phi}_s} T \sim F_{p,\hat{v}_s,(1-\alpha)} \tag{37}$$

biçiminde hesaplanmaktadır. Eğer T_{BC} ve T_S test istatistikleri sırasıyla $F_{p,\hat{v}_{BC},(1-\alpha)}$ ve $F_{p,\hat{v}_s,(1-\alpha)}$ kritik değerlerinden daha büyük ise yokluk hipotezi reddedilmektedir. Ayrıca $r = 1,\ldots,6$ için θ_r ve θ_r^* değerleri Kawasaki ve Seo (2015)'da görüldüğü gibi hesaplanmaktadır.

3. Bulgular ve Tartışma

Bu bölümde çok değişkenli iki örneklem Behrens-Fisher probleminin çözümü için önerilen test istatistiklerinin bir simülasyon çalışması ile karşılaştırılması amaçlanmıştır. Bu amaçla R (sürüm 4.1.0) programlama dili ve RStudio geliştirme ortamından yararlanılarak Yao (1965), Johansen (1980), Nel ve Van der Merwe (1986), Kim (1992), Krishnamoorthy ve Yu (2004), Yanagihara ve Yuan (2005) ile Kawasaki ve Seo (2015) tarafından önerilen test istatistiklerinin R fonksiyonları kodlanmıştır. Ayrıca çok değişkenli iki örneklem Behrens-Fisher problemlerinin çözümü için önerilen test istatistikleri, RStudio ortamındaki "HSAUR2" kütüphanesinde yer alan "skulls" veri seti kullanılarak karşılaştırılmıştır.

Simülasyon çalışmasında ilk olarak normal dağılıma sahip rassal örneklemler türetilmiştir. Ayrıca ikinci aşamada farklı dağılıma sahip veri setlerine göre test istatistiklerinin performanslarını görmek için Gamma $(\Gamma(3,2))$ dağılımına göre standartlaştırılmış tesadüfi örneklemler elde edilmiştir. Varyans-kovaryans matrislerinin heterojen olmasını sağlamak için $i = 1,2$ ve $j = 1, \dots, p$ olmak üzere

$$\Sigma_i = \begin{cases} 1 & i = j \\ (\rho_i)^{|i-j|} & i \neq j \end{cases} \tag{38}$$

eşitliğindeki kovaryans modeli kullanılmıştır. Böylece varyans-kovaryans matrisleri heterojen olan veriler için önerilen test istatistiklerinin kritik değerleri elde edilmiştir. Test istatistiğinin değeri ile kritik değerler karşılaştırılmış ve yokluk hipotezinin reddedilip reddedilemediğine karar verilmiştir. Çok sayıda tekrar edilen bu işlemler sonucunda her bir test istatistiği için elde edilen red sayıları, yapılan tekrar sayısına oranlanarak I. tip hata olasılıkları elde edilmiştir. Test istatistiklerine ait I. Tip hata olasılıklarını belirlemek için 50000 tekrar gerçekleştirilmiştir. Ayrıca iki örneklem durumunda bağımlı değişken sayısının test istatistikleri üzerindeki etkisini görmek için $p = 3$ ve $p = 6$ olmak üzere iki farklı durum ele alınmıştır.

Simülasyon çalışması sonucunda test istatistiklerinin performanslarını elde edilen deneysel I. Tip hata olasılıklarına $(\hat{\alpha})$ göre belirlemek için Zhang ve Liu (2011) ile Zhang (2012) tarafından kullanılan "Ortalama Göreceli Hata (ARE)" değeri hesaplanmıştır. Test istatistikleri için farklı koşullar altında elde edilen olasılık değerlerinin sayısı $\widetilde{N}$ olmak üzere her bir test istatistiğine göre ARE değeri

$$ARE = \frac{100}{\widetilde{N}} \sum_{t=1}^{\widetilde{N}} \frac{|\hat{\alpha}_t - \alpha|}{\alpha} \tag{39}$$

eşitliğinden yararlanılarak hesaplanmaktadır. Böylece ARE değeri daha düşük olan test istatistiklerinin anlamlılık düzeyine daha yakın değerler ortaya koyduğu söylenebilmektedir.

Simülasyon çalışması ile normal dağılımdan türetilen veri seti için farklı gözlem büyüklükleri ile varyans-kovaryans matrislerinin heterojenlik koşulu altında $p = 3$ olduğunda %5 $(\alpha = 0{,}05)$ anlamlılık düzeyinde elde edilen I. Tip hata olasılıkları Tablo 1'de gösterilmiştir.

Tablo 1. Normal dağılıma göre $p = 3$ için elde edilen I. tip hata olasılıkları

$(\boldsymbol{n_1}; \boldsymbol{n_2})$	$(\boldsymbol{\rho_1}; \boldsymbol{\rho_2})$	$\boldsymbol{T_{Yao}}$	$\boldsymbol{T_{Joh}}$	$\boldsymbol{T_{NV}}$	$\boldsymbol{T_{MNV}}$	$\boldsymbol{T_{Hot}}$	$\boldsymbol{T_{BC}}$	$\boldsymbol{T_S}$	$\boldsymbol{T_{YY}}$	$\boldsymbol{T_{MB}}$	$\boldsymbol{T_{Kim}}$
(12; 12)	(0,1; 0,9)	0,0478	0,0508	0,0520	0,0484	0,0580	0,0480	0,0484	0,0471	0,0484	0,0401
	(0,4; 0,6)	0,0457	0,0477	0,0462	0,0464	0,0490	0,0443	0,0461	0,0461	0,0465	0,0450
(24; 24)	(0,1; 0,9)	0,0485	0,0491	0,0515	0,0491	0,0532	0,0490	0,0490	0,0489	0,0490	0,0441
	(0,4; 0,6)	0,0497	0,0498	0,0499	0,0499	0,0508	0,0493	0,0498	0,0498	0,0499	0,0491
(48; 48)	(0,1; 0,9)	0,0490	0,0490	0,0507	0,0494	0,0516	0,0492	0,0493	0,0491	0,0493	0,0464
	(0,4; 0,6)	0,0484	0,0482	0,0486	0,0485	0,0487	0,0483	0,0485	0,0485	0,0485	0,0481
(12; 24)	(0,1; 0,9)	0,0537	0,0553	0,0619	0,0527	0,1434	0,0552	0,0534	0,0504	0,0529	0,0389
	(0,4; 0,6)	0,0539	0,0517	0,0500	0,0502	0,0611	0,0524	0,0513	0,0493	0,0503	0,0473
(24; 12)	(0,1; 0,9)	0,0493	0,0500	0,0381	0,0497	0,0341	0,0491	0,0495	0,0491	0,0495	0,0465
	(0,4; 0,6)	0,0536	0,0516	0,0453	0,0509	0,0431	0,0518	0,0514	0,0503	0,0509	0,0502
(24; 36)	(0,1; 0,9)	0,0500	0,0504	0,0557	0,0504	0,0941	0,0507	0,0504	0,0498	0,0504	0,0444
	(0,4; 0,6)	0,0497	0,0490	0,0503	0,0493	0,0560	0,0494	0,0494	0,0491	0,0493	0,0473
(36; 24)	(0,1; 0,9)	0,0496	0,0501	0,0468	0,0504	0,0360	0,0501	0,0503	0,0501	0,0503	0,0464
	(0,4; 0,6)	0,0509	0,0506	0,0495	0,0509	0,0454	0,0506	0,0509	0,0508	0,0509	0,0506
(12; 48)	(0,1; 0,9)	0,0580	0,0575	0,0583	0,0536	0,2827	0,0578	0,0551	0,0499	0,0539	0,0354
	(0,4; 0,6)	0,0622	0,0567	0,0514	0,0542	0,0738	0,0597	0,0564	0,0513	0,0545	0,0443
(48; 12)	(0,1; 0,9)	0,0514	0,0512	0,0289	0,0511	0,0406	0,0522	0,0513	0,0504	0,0510	0,0484
	(0,4; 0,6)	0,0577	0,0528	0,0418	0,0511	0,0374	0,0563	0,0531	0,0493	0,0512	0,0462
	ARE	***5,967***	***3,989***	***9,633***	***2,756***	***54,489***	***5,578***	***3,729***	***1,922***	***2,878***	***9,211***

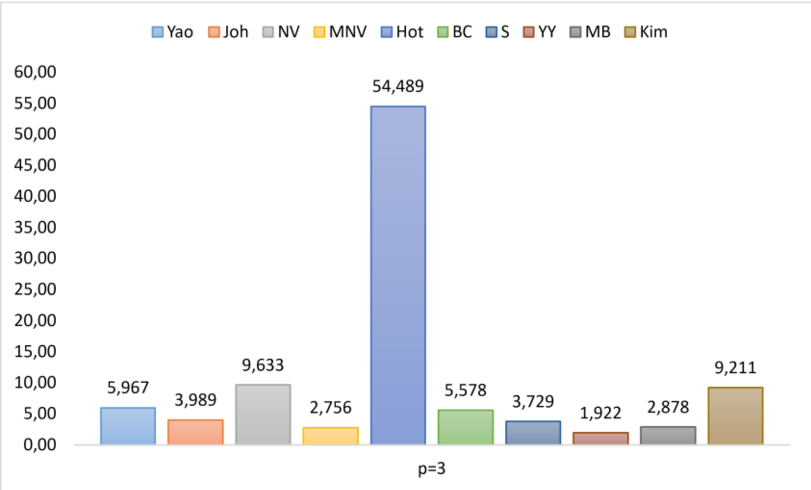
Tablo 1'e göre $p = 3$ durumunda anlamlılık düzeyine en yakın değerler ortaya koyan test istatistiklerinin T_{YY}, T_{MNV}, T_{MB} ve T_S olduğu görülmektedir. Ancak T_{Hot}, T_{NV} ve T_{Kim} test istatistikleri genel olarak daha düşük performans ortaya koymaktadır. Üstelik tüm koşullar altında en kötü performans $n_i = (12; 48)$ ve $\rho = (0,1; 0,9)$ koşulunda 0,2827 değeri ile T_{Hot} istatistiğinden elde edilmiştir. Dolayısıyla T_{Hot} istatistiği, eşit olmayan gözlem büyüklüklerinden ve gözlem sayıları arasındaki farkın artışından etkilenmektedir. T_{NV} istatistiği, T_{Hot} istatistiğine göre daha iyi performans göstermiştir. Ancak T_{Yao}, T_{MNV}, T_{BC}, T_S, T_{YY} ve T_{MB} test istatistiklerine göre T_{NV} ve T_{Kim} için I. tip hata olasılıklarının anlamlılık düzeyinden daha uzak olduğu belirlenmiştir. Normal dağılımdan türetilen veri seti için farklı gözlem büyüklükleri ile varyans-kovaryans matrislerinin heterojenlik koşulu altında $p = 6$ için %5 anlamlılık düzeyindeki I. Tip hata olasılıkları Tablo 2'de gösterilmiştir.

Tablo 2. Normal dağılıma göre $p = 6$ için elde edilen I. tip hata olasılıkları

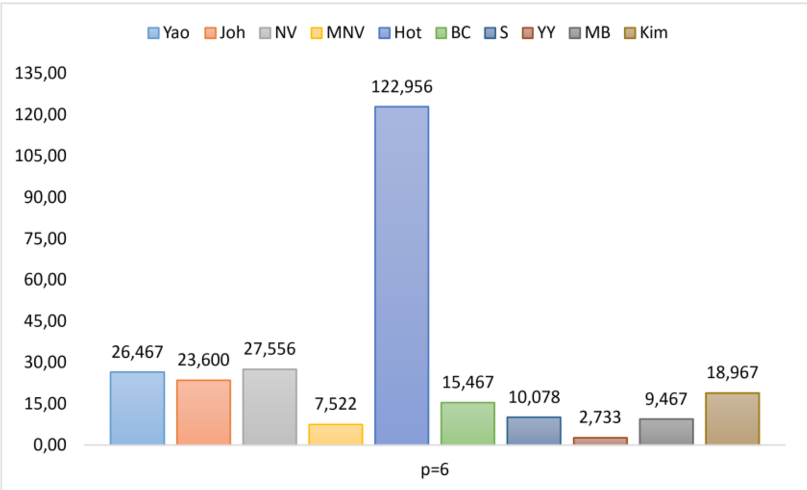
$(\boldsymbol{n_1}; \boldsymbol{n_2})$	$(\boldsymbol{\rho_1}; \boldsymbol{\rho_2})$	$\boldsymbol{T_{Yao}}$	$\boldsymbol{T_{Joh}}$	$\boldsymbol{T_{NV}}$	$\boldsymbol{T_{MNV}}$	$\boldsymbol{T_{Hot}}$	$\boldsymbol{T_{BC}}$	$\boldsymbol{T_S}$	$\boldsymbol{T_{YY}}$	$\boldsymbol{T_{MB}}$	$\boldsymbol{T_{Kim}}$
(12; 12)	(0,1; 0,9)	0,0576	0,0736	0,0582	0,0514	0,0768	0,0531	0,0522	0,0472	0,0543	0,0304
	(0,4; 0,6)	0,0420	0,0589	0,0435	0,0440	0,0506	0,0432	0,0454	0,0482	0,0465	0,0400
(24; 24)	(0,1; 0,9)	0,0501	0,0533	0,0562	0,0505	0,0625	0,0511	0,0506	0,0491	0,0508	0,0361
	(0,4; 0,6)	0,0479	0,0497	0,0487	0,0486	0,0500	0,0481	0,0487	0,0484	0,0489	0,0471
(48; 48)	(0,1; 0,9)	0,0498	0,0497	0,0540	0,0500	0,0563	0,0502	0,0500	0,0496	0,0500	0,0425
	(0,4; 0,6)	0,0494	0,0490	0,0496	0,0495	0,0499	0,0493	0,0495	0,0494	0,0496	0,0486
(12; 24)	(0,1; 0,9)	0,0830	0,0867	0,0886	0,0615	0,2448	0,0688	0,0634	0,0517	0,0654	0,0334
	(0,4; 0,6)	0,0735	0,0673	0,0478	0,0549	0,0704	0,0631	0,0584	0,0508	0,0569	0,0567
(24; 12)	(0,1; 0,9)	0,0469	0,0528	0,0203	0,0479	0,2470	0,0473	0,0479	0,0469	0,0483	0,0424
	(0,4; 0,6)	0,0595	0,0596	0,0330	0,0514	0,0396	0,0553	0,0539	0,0493	0,0527	0,0572
(24; 36)	(0,1; 0,9)	0,0543	0,0565	0,0680	0,0530	0,1390	0,0550	0,0534	0,0510	0,0536	0,0344
	(0,4; 0,6)	0,0524	0,0514	0,0520	0,0502	0,0622	0,0517	0,0510	0,0495	0,0505	0,0468
(36; 24)	(0,1; 0,9)	0,0494	0,0505	0,0408	0,0500	0,0297	0,0497	0,0499	0,0495	0,0500	0,0426
	(0,4; 0,6)	0,0512	0,0511	0,0458	0,0502	0,0433	0,0504	0,0507	0,0498	0,0506	0,0501
(12; 48)	(0,1; 0,9)	0,0921	0,0948	0,0711	0,0634	0,4828	0,0718	0,0651	0,0462	0,0676	0,0284
	(0,4; 0,6)	0,1037	0,0788	0,0393	0,0598	0,0912	0,0773	0,0657	0,0490	0,0626	0,0572
(48; 12)	(0,1; 0,9)	0,0558	0,0557	0,0076	0,0535	0,0316	0,0556	0,0542	0,0516	0,0536	0,0564
	(0,4; 0,6)	0,0904	0,0698	0,0237	0,0579	0,0329	0,0734	0,0635	0,0516	0,0599	0,0658
	ARE	***26,467***	***23,600***	***27,556***	***7,522***	***122,956***	***15,467***	***10,078***	***2,733***	***9,467***	***18,967***

Tablo 2 incelendiğinde bağımlı değişken sayısı arttıkça test istatistikleri için elde edilen deneysel hata olasılıklarının da %5 anlamlılık düzeyinden uzaklaştığı gözlemlenmiştir. Ancak T_{YY} test istatistiğinin, diğer test istatistiklere göre anlamlılık düzeyine daha yakın değerler ortaya koyduğu görülmektedir. T_{MNV}, T_{MB}, T_S ve T_{BC} test istatistikleri birbirlerine daha yakın değerler ortaya koyarken T_{Yao} ve T_{Joh} istatistiklerinin ise gözlem sayıları arasındaki farkın artışından oldukça fazla etkilendiği sonucuna ulaşılmıştır.

Tablo 1 ve Tablo 2 'de yer alan olasılık değerlerine göre elde edilen ARE değerleri Şekil 1'de gösterilmiştir.



Şekil 1. Normal dağılıma göre p = 3 için ARE değerleri



Şekil 2. Normal dağılıma göre p = 6 için ARE değerleri

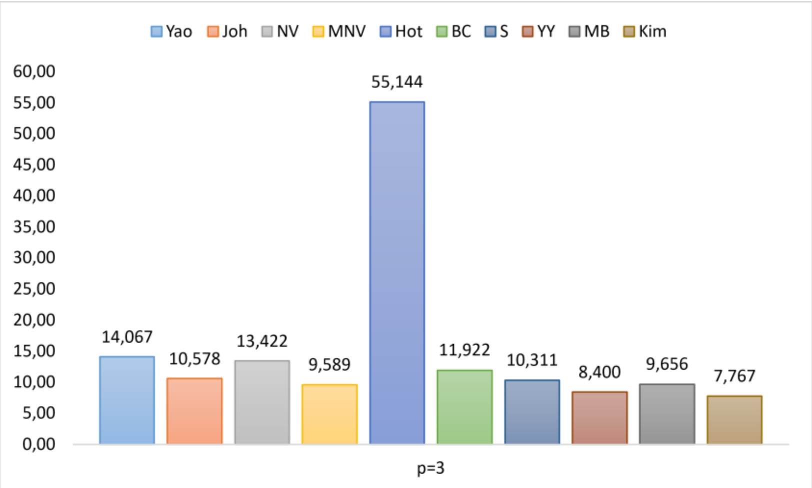
Şekil 1 ve Şekil 2'ye göre $p = 3$ için en düşük ARE değerleri sırasıyla T_{YY}, T_{MNV} ve T_{MB} test istatistiklerinden elde edilmiştir. Ayrıca T_s ve T_{Joh} ile T_{BC} ve T_{Yao} test istatistikleri ise birbirlerine yakın değerler ortaya koymaktadır. Ancak bağımlı değişken sayısı $p = 6$ olduğunda bütün test istatistiklerine ait ARE değerlerinin yükseldiği gözlemlenmiştir. En düşük ARE değeri T_{YY} istatistiğinden elde edilirken bu test istatistiğinin performansını sırasıyla T_{MNV}, T_{MB}, T_S ve T_{BC} test istatistikleri takip etmektedir.

Simülasyon çalışmasının bu aşamasında ise Gamma $(\Gamma(3{,}2))$ dağılımına göre türetilen veri seti için farklı gözlem büyüklükleri ile varyans-kovaryans matrislerinin heterojenlik koşulu altında $p = 3$ olduğunda %5 anlamlılık düzeyinde elde edilen I. Tip hata olasılıkları Tablo 3'te gösterilmiştir.

Tablo 3. Gamma $(\Gamma(3{,}2))$ dağılıma göre $p = 3$ için elde edilen I. tip hata olasılıkları

$(n_1; n_2)$	$(\rho_1; \rho_2)$	T_{Yao}	T_{Joh}	T_{NV}	T_{MNV}	T_{Hot}	T_{BC}	T_S	T_{YY}	T_{MB}	T_{Kim}
$(12; 12)$	$(0{,}1; 0{,}9)$	0,0495	0,0516	0,0510	0,0491	0,0588	0,0485	0,0491	0,0479	0,0491	0,0395
	$(0{,}4; 0{,}6)$	0,0417	0,0446	0,0422	0,0430	0,0472	0,0410	0,0428	0,0424	0,0430	0,0424
$(24; 24)$	$(0{,}1; 0{,}9)$	0,0520	0,0524	0,0542	0,0525	0,0567	0,0522	0,0523	0,0520	0,0523	0,0461
	$(0{,}4; 0{,}6)$	0,0470	0,0474	0,0475	0,0476	0,0490	0,0471	0,0475	0,0475	0,0475	0,0481
$(48; 48)$	$(0{,}1; 0{,}9)$	0,0515	0,0521	0,0535	0,0524	0,0546	0,0523	0,0523	0,0523	0,0523	0,0495
	$(0{,}4; 0{,}6)$	0,0506	0,0505	0,0508	0,0508	0,0514	0,0506	0,0507	0,0507	0,0508	0,0523
$(12; 24)$	$(0{,}1; 0{,}9)$	0,0630	0,0614	0,0675	0,0590	0,1440	0,0611	0,0594	0,0567	0,0590	0,0456
	$(0{,}4; 0{,}6)$	0,0587	0,0557	0,0533	0,0543	0,0594	0,0559	0,0550	0,0534	0,0544	0,0529
$(24; 12)$	$(0{,}1; 0{,}9)$	0,0487	0,0499	0,0382	0,0496	0,0352	0,0491	0,0494	0,0489	0,0494	0,0450
	$(0{,}4; 0{,}6)$	0,0532	0,0511	0,0449	0,0505	0,0412	0,0509	0,0508	0,0498	0,0505	0,0517
$(24; 36)$	$(0{,}1; 0{,}9)$	0,0536	0,0531	0,0578	0,0531	0,0957	0,0534	0,0532	0,0527	0,0531	0,0459
	$(0{,}4; 0{,}6)$	0,0512	0,0504	0,0511	0,0506	0,0557	0,0507	0,0506	0,0504	0,0506	0,0497
$(36; 24)$	$(0{,}1; 0{,}9)$	0,0504	0,0511	0,0480	0,0514	0,0383	0,0510	0,0512	0,0511	0,0513	0,0474
	$(0{,}4; 0{,}6)$	0,0490	0,0489	0,0479	0,0491	0,0439	0,0489	0,0490	0,0490	0,0490	0,0503
$(12; 48)$	$(0{,}1; 0{,}9)$	0,0745	0,0695	0,0703	0,0664	0,2832	0,0702	0,0676	0,0630	0,0668	0,0424
	$(0{,}4; 0{,}6)$	0,0784	0,0688	0,0623	0,0667	0,0732	0,0720	0,0686	0,0644	0,0669	0,0567
$(48; 12)$	$(0{,}1; 0{,}9)$	0,0545	0,0543	0,0323	0,0542	0,0432	0,0549	0,0545	0,0533	0,0541	0,0505
	$(0{,}4; 0{,}6)$	0,0709	0,0640	0,0500	0,0628	0,0384	0,0667	0,0644	0,0611	0,0628	0,0571
	ARE	***14,067***	***10,578***	***13,422***	***9,589***	***55,144***	***11,922***	***10,311***	***8,400***	***9,656***	***7,767***

Ayrıca $p = 3$ ve n_i gözlem sayıları için Tablo 3'te yer alan olasılık değerlerine göre elde edilen ARE değerleri Şekil 3'te gösterilmiştir.



Şekil 3. Gamma $(\Gamma(3{,}2))$ dağılıma göre p = 3 için ARE değerleri

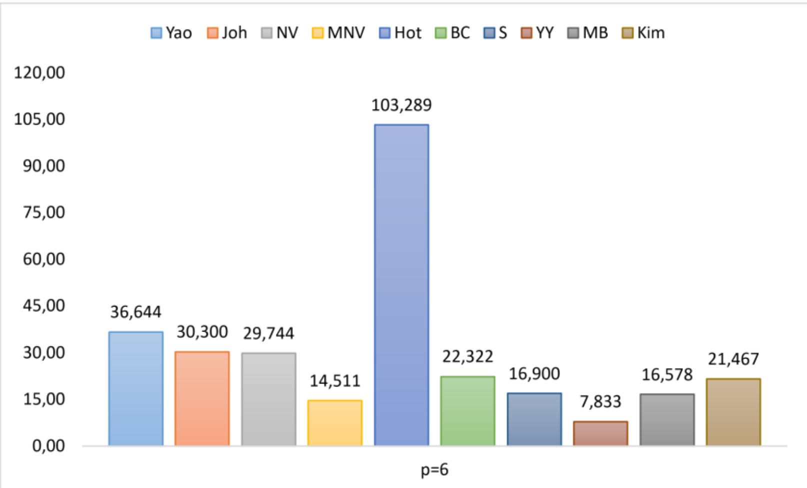
Tablo 3 ve Şekil 3 incelendiğinde $p = 3$ durumu için Gamma $(\Gamma(3{,}2))$ dağılımına göre anlamlılık düzeyine en yakın değerler ortaya koyan test istatistiklerinin T_{Kim}, T_{YY}, T_{MNV}, T_{MB} ve T_S olduğu görülmektedir. Bu sonuçlar, normal dağılıma göre türetilen veri setine göre elde edilen bulgular ile paralellik göstermektedir. Ancak Gamma $(\Gamma(3{,}2))$ dağılıma göre elde edilen I. tip hata olasılıkları ile genel olarak ARE değerlerinin daha yüksek olduğu sonucuna ulaşılmıştır. Bütün deneysel koşullar altında test istatistikleri iyi bir performans ortaya koyarken T_{Hot} istatistiği gözlem büyüklüklerinin farklılığından çok fazla etkilenmektedir. ARE değerlerine göre en düşük ARE değerleri sırasıyla T_{Kim}, T_{YY}, T_{MNV} ve T_{MB} test istatistiklerinden elde edilmiştir.

Gamma $\left(\Gamma(3,2)\right)$ dağılımına göre türetilen veri seti için farklı gözlem büyüklükleri ile varyans-kovaryans matrislerinin heterojenlik koşulu altında $p = 6$ olduğunda %5 anlamlılık düzeyinde elde edilen I. Tip hata olasılıkları Tablo 4'te gösterilmiştir.

Tablo 4. Gamma $\left(\Gamma(3,2)\right)$ dağılıma göre $p = 6$ için elde edilen I. tip hata olasılıkları

$(n_1; n_2)$	$(\rho_1; \rho_2)$	T_{Yao}	T_{Joh}	T_{NV}	T_{MNV}	T_{Hot}	T_{BC}	T_S	T_{YY}	T_{MB}	T_{Kim}
(12; 12)	(0,1; 0,9)	0,0608	0,0777	0,0585	0,0544	0,0798	0,0559	0,0552	0,0499	0,0574	0,0322
	(0,4; 0,6)	0,0394	0,0561	0,0394	0,0408	0,0482	0,0401	0,0417	0,0393	0,0430	0,0376
(24; 24)	(0,1; 0,9)	0,0557	0,0575	0,0599	0,0541	0,0670	0,0550	0,0543	0,0527	0,0545	0,0377
	(0,4; 0,6)	0,0460	0,0486	0,0467	0,0470	0,0493	0,0465	0,0471	0,0466	0,0473	0,0468
(48; 48)	(0,1; 0,9)	0,0507	0,0504	0,0544	0,0507	0,0571	0,0508	0,0507	0,0503	0,0507	0,0407
	(0,4; 0,6)	0,0477	0,0476	0,0482	0,0481	0,0490	0,0478	0,0481	0,0480	0,0481	0,0487
(12; 24)	(0,1; 0,9)	0,0912	0,0917	0,0919	0,0676	0,2398	0,0746	0,0691	0,0575	0,0714	0,0387
	(0,4; 0,6)	0,0736	0,0673	0,0459	0,0554	0,0654	0,0618	0,0583	0,0517	0,0569	0,0602
(24; 12)	(0,1; 0,9)	0,0487	0,0551	0,0214	0,0505	0,0267	0,0501	0,0506	0,0496	0,0509	0,0406
	(0,4; 0,6)	0,0590	0,0586	0,0330	0,0509	0,0385	0,0540	0,0526	0,0494	0,0520	0,0585
(24; 36)	(0,1; 0,9)	0,0599	0,0604	0,0719	0,0570	0,1415	0,0587	0,0574	0,0552	0,0574	0,0370
	(0,4; 0,6)	0,0527	0,0515	0,0512	0,0504	0,0586	0,0514	0,0509	0,0498	0,0507	0,0482
(36; 24)	(0,1; 0,9)	0,0502	0,0515	0,0416	0,0509	0,0303	0,0507	0,0507	0,0503	0,0509	0,0426
	(0,4; 0,6)	0,0488	0,0492	0,0447	0,0485	0,0417	0,0486	0,0489	0,0481	0,0488	0,0498
(12; 48)	(0,1; 0,9)	0,1163	0,1095	0,0822	0,0768	0,4821	0,0877	0,0787	0,0570	0,0824	0,0368
	(0,4; 0,6)	0,1273	0,0933	0,0450	0,0736	0,0892	0,0926	0,0799	0,0621	0,0766	0,0741
(48; 12)	(0,1; 0,9)	0,0572	0,0580	0,0093	0,0557	0,0331	0,0579	0,0565	0,0541	0,0558	0,0584
	(0,4; 0,6)	0,1058	0,0795	0,0271	0,0670	0,0341	0,0827	0,0730	0,0603	0,0688	0,0794
	ARE	***36,644***	***30,300***	***29,744***	***14,511***	***103,289***	***22,322***	***16,900***	***7,833***	***16,578***	***21,467***

Ayrıca $p = 6$ ve n_i gözlem sayıları için Tablo 4'te yer alan olasılık değerlerine göre elde edilen ARE değerleri Şekil 4'te gösterilmiştir.



Şekil 4. Gamma $\left(\Gamma(3,2)\right)$ dağılıma göre $\mathrm{p} = 6$ için ARE değerleri

Tablo 4 ve Şekil 4 incelendiğinde ise Gamma $\left(\Gamma(3,2)\right)$ dağılımına göre türetilen veri seti için bağımlı değişken sayısı arttıkça test istatistiklerinden elde edilen deneysel hata olasılıklarının da nominal anlamlılık düzeyinden uzaklaştığı gözlemlenmiştir. Ancak T_{YY} test istatistiğinin, diğer test istatistiklere göre anlamlılık düzeyine daha yakın değerler ortaya koyduğu görülmektedir. Şekil 4'e göre en düşük ARE değerleri sırasıyla T_{YY}, T_{MNV} ve T_{MB} test istatistiklerinden elde edilmiştir.

Son bölümde ise önerilen test istatistiklerinin gerçek bir veri örneği üzerinde karşılaştırılması için RStudio ortamındaki "HSAUR2" kütüphanesinde yer alan "skulls" veri seti kullanılmıştır. Bu veri seti; hanedanlık (firavunlar) öncesi ilk dönemler (yaklaşık MÖ 4000); hanedanlık öncesi son dönemler (yaklaşık MÖ 3300); 12. ve 13. hanedanlar (yaklaşık MÖ 1850); Ptolemaios hanedanlık dönemi (yaklaşık MÖ 200) ve Roma dönemi (yaklaşık MS 150) olmak üzere Mısır'da beş farklı döneme ait erkek kafatası örneklerine ilişkin gözlem verileri içermektedir. Ayrıca kafatası örneklerinin her biri için dört farklı özellik bilinmekte ve her bir dönemden 30'ar tane olmak üzere 150 gözlemden oluşmaktadır. Ayrıca bu veri seti Thomson ve Randall-Maciver (1905) tarafından sunulmuş ve Krishnamoorthy ve Lu (2010), Zhang ve Liu (2011) ile Zhang (2012) tarafından k örneklem Behrens-Fisher problemleri için uygulanmıştır.

Çalışmada MÖ 4000-MÖ 3300 (I), MÖ 3300-MÖ 1850 (II) ve MÖ 1850-MÖ 200 (III) olmak üzere üç farklı karşılaştırma yapılmıştır. Böylece grup ortalamaları arasındaki farkların zamanla değişip değişmediğinin tespit edilmesi amaçlanmıştır. Ayrıca dengeli ve dengesiz gözlem büyüklüklerinin test istatistiklerinin performansı üzerindeki etkisini göstermek amacıyla MÖ 4000, MÖ 3300, MÖ 1850 ve MÖ 200 dönemleri için sırasıyla veri setine ait ilk $n_1 = 21$, $n_2 = 21$, $n_3 = 10$ ve $n_4 = 6$ adet gözlem kullanılmıştır. Bu durumda iki örneklem Behrens-Fisher problemleri için önerilen test istatistikleri kullanılarak %5 anlamlılık düzeyinde dört farklı karşılaştırma için elde edilen olasılık değerleri Tablo 5'te yer almaktadır.

Tablo 5. Kafatası örneklerine göre üç farklı karşılaştırma için elde edilen olasılık değerleri

	T_{Yao}	T_{Joh}	T_{NV}	T_{MNV}	T_{Hot}	T_{BC}	T_{S}	T_{YY}	T_{MB}	T_{Kim}
I	0,62278	0,61620	0,62333	0,62358	0,62256	0,62418	0,62333	0,62325	0,62299	0,61717
II	0,13011	0,09789	0,11270	0,10773	***0,04708***	0,09719	0,10301	0,11091	0,10651	0,11241
III	0,13106	0,06520	0,13442	0,09217	***0,03526***	0,08011	0,08507	0,09964	0,08786	0,12050

Tablo 5 incelendiğinde I. durum için bütün test istatistiklerine göre elde edilen olasılık değerlerinin %5'ten daha büyük olduğu görülmüştür. Bu durumda MÖ 4000-MÖ 3300 (I) karşılaştırılması için "Kafatası örnekleri arasında anlamlı bir farklılık yoktur" biçimindeki sıfır hipotezi reddedilememiştir. Böylece MÖ 4000 ve MÖ 3300 dönemlerine ait kafatası örnekleri arasında dört farklı ölçüm bakımından anlamlı bir farklılık olmadığı söylenebilmektedir. MÖ 3300-MÖ 1850 (II) ile MÖ 1850-MÖ 200 (III) için gerçekleştirilen analiz sonuçları incelendiğinde ise T_{Hot} test istatistiği için %5 anlamlılık düzeyinde sıfır hipotezinin reddedildiği ve ilgili dönemler arasında anlamlı bir farklılık olduğu sonucuna ulaşılmıştır. Ancak diğer bütün test istatistiklerine göre %5 anlamlılık düzeyinde ilgili dönemler arasında kafatası örnekleri için dört farklı ölçüm bakımından anlamlı bir farklılık olmadığı görülmektedir.

4. Sonuç

Bu çalışmanın amacı, iki örneklem Behrens-Fisher problemleri için önerilen test istatistiklerini I. tip hata olasılıkları bakımından karşılaştırmaktır. Bu amaçla varsayımların ihlal edilmesi durumunda iki grup ortalama vektörünün eşitliğini test etmek için Yao (1965), Johansen (1980), Nel ve Van der Merwe (1986), Kim (1992), Krishnamoorthy ve Yu (2004), Yanagihara ve Yuan (2005) ile Kawasaki ve Seo (2015) tarafından önerilen test istatistikleri bir simülasyon çalışması ile karşılaştırılmıştır. Ayrıca çalışmada bağımlı değişken sayısına ve gözlem büyüklükleri için farklı koşullar altında test istatistiklerinin performansını incelenmiştir.

Bağımlı değişken sayısının küçük olduğu durumda T_{Yao}, T_{Joh}, T_{MNV}, T_{BC}, T_S, T_{YY} ve T_{MB} test istatistikleri için elde edilen olasılık değerlerinin birbirine yakın olduğu görülmüştür. Özellikle gözlem sayılarının dengeli olduğu durumlarda anlamlılık düzeyine oldukça yakın değerler elde edilmiştir. Ancak iki örneklemin gözlem sayıları arasındaki fark yükseldikçe elde edilen I. tip hata olasılıklarının da anlamlılık düzeyinden uzaklaştığı gözlemlenmiştir. Bağımlı değişken sayısı yükseldikçe test istatistikleri için elde edilen deneysel hata olasılıklarının da nominal anlamlılık düzeyinden uzaklaştığı gözlemlenmiştir. Bütün sonuçlar incelendiğinde genel olarak en iyi performansın sırasıyla T_{YY}, T_{MNV}, T_{MB} ve T_S istatistikleri göstermiştir. Hotelling T^2 istatistiğinin ise varsayım ihlallerinden, bağımlı değişken sayısından ve dengesiz gözlem büyüklüğünden çok fazla etkilendiği görülmektedir. Bu durumda çalışmanın sonuçları incelenen test istatistiklerinin genel olarak bağımlı değişken sayısına ve gözlem büyüklüklerine karşı oldukça duyarlı olduğunu göstermiştir.

Bu çalışma ile varyans-kovaryans matrisleri heterojen olan iki örneklemin ortalama vektörlerini karşılaştırmak için ideal test istatistiklerinin ortaya konulması hedeflenmiştir. Ayrıca farklı deneysel koşullar altında test istatistiklerinin performansı araştırılmıştır. Ancak bağımlı değişken sayısına ve gözlem büyüklüklerine göre test istatistiklerinin performanslarının farklılaştığı gözlemlenmiştir. Bu nedenle araştırma problemi için elde edilen veri setinin özellikleri dikkate alınarak en yüksek performansa sahip test istatistiği seçilebilir. Bundan sonraki çalışmalarda çeşitli istatistiksel dağılımlar ya da gözlem büyüklüklerine göre test istatistiklerinin performansları araştırılabilir. Ayrıca bağımlı değişken sayısının gözlem sayısından daha büyük olduğu yüksek boyutlu veriler için iki örneklem Behrens-Fisher problemlerinin çözümüne yönelik alternatif test istatistikleri geliştirebilir.

Çıkar Çatışması Beyanı

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

Araştırmacıların Katkı Oranı Beyan Özeti

Bu çalışma Esin Budak'ın Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü'nde Zeki Yıldız danışmanlığında hazırladığı yüksek lisans tez konusundan üretilmiştir. Tüm yazarlar makaleye eşit oranda katkı sağlamış olduklarını beyan ederler.

Kaynakça

Behrens WV. Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen (A contribution to error estimation with few observations), Landwirtschaftliches Jahrbuch 1929; 68: 807-837.

Bennett BM. Note on a solution of the generalized Behrens–Fisher problem. Annals of the Institute Statistical Mathematics 1951; 2: 87–90.

Budak E. İki örneklem Behrens-Fisher problem. Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, sayfa no: 35, Eskişehir, Türkiye, 2021.

Christensen WF., Rencher AC. A comparison of type I error rates and power levels for seven solutions to the multivariate Behrens-Fisher problem. Communications in Statistics - Simulation and Computation 1997; 26: 1251-1273.

Coombs WT., Algina J., Oltman DO. Univariate and multivariate omnibus hypothesis tests selected to control type I error rates when population variances are not necessarily equal. Review of Educational Research 1996; 66(2): 137–179.

Erdoğan S. Heterojenlik altında iki grup ortalama vektörlerinin karşılaştırılması için önerilen yeni bir hesaplamalı yaklaşım testi. Gazi Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, sayfa no: 78, Ankara, Türkiye, 2018.

Fisher RA. The fiducial argument in statistical inference. Annals of Eugenics 1935; 6(4): 391-398.

Fujikoshi Y. Transformations with improved chi-squared approximations. Journal of Multivariate Analysis 2000; 72(2): 249-263.

Hotelling H. The generalization of Student’s ratio. The Annals of Mathematical Statistics 1931; 2(3): 360–378.

James GS. Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. Biometrika 1954; 41(1/2): 19-43.

Johansen S. The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. Biometrika 1980; 67(1): 85-92.

Kawasaki T., Seo T. A two sample test for mean vectors with unequal covariance matrices. Communications in Statistics: Simulation and Computation 2015; 44(7): 1850-1866.

Kim S. A practical solution to the multivariate Behrens- Fisher problem. Biometrika 1992; 79(1): 171-176.

Krishnamoorthy K., Lu F. A parametric bootstrap solution to the MANOVA under heteroscedasticity. Journal of Statistical Computation and Simulation 2010; 80(8): 873-887.

Krishnamoorthy K., Yu J. Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem. Statistics and Probability Letters 2004; 66: 161–169.

Krishnamoorthy K., Yu J. Multivariate Behrens-Fisher problem with missing data. Journal of Multivariate Analysis 2012; 105(1): 141–150.

Lix LM., Keselman HJ. Multivariate tests of means in independent groups designs: Effects of covariance heterogeneity and nonnormality. Evaluation & The Health Professions 2004; 27(1): 45-69.

Nel DG., Van der Merwe, CA. A solution to the multivariate Behrens-Fisher problem. Communication Statistics-Theory and Methods 1986; 15(12): 3719-3735.

Pfanzagl J. On the Behrens-Fisher problem. Biometrika 1974; 61(1): 39–47.

Sandal M. Kovaryans matrislerinin homojenliği varsayımı sağlanmadığında istatistiksel çözümleme yaklaşımları. Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi, sayfa no: 155, Eskişehir, Türkiye, 2020.

Scheffé H. On solutions of the Behrens–Fisher problem, based on the t-distribution. Annals of Mathematical Statistics 1943; 14(1): 35–44.

Subrahmaniam K., Subrahmaniam K. On the multivariate Behrens-Fisher problem. Biometrika 1973; 60(1): 107–111.

Thomson A. Randall-Maciver R. Ancient races of the Thebaid. Oxford University Press 1905.

Wald A. Testing the difference between means of two normal populations with unknown standard deviations. Selected Papers in Probability Statistics 1955; 669-695.

Welch BL. The significance of the difference between two means when the population variances are unequal. Biometrika 1938; 29(3/4): 350–362.

Welch BL. The generalization of Student's problem when several different population variances are involved. Biometrika 1947; 34(1-2): 28-35.

Yanagihara H., Yuan KH. Three approximate solutions to the multivariate Behrens-Fisher problem. Communication in Statistics - Simulation and Computation 2005; 34(4): 975-988.

Yao Y. An approximate degrees of freedom solution to the Multivariate Behrens-Fisher problem. Biometrika 1965; 52(1-2): 139-147.

Zezula I. Implementation of a new solution to the multivariate Behrens-Fisher problem. Stata Journal 2009; 9(4): 593-598.

Zhang JT., Liu X. A modified Bartlett test for heteroscedastic one-way MANOVA. Metrika 2011; 76: 135–152.

Zhang JT. An approximate Hotelling T^2-test for heteroscedastic one-way MANOVA. Open Journal of Statistics 2012; 2(1): 1-11.