



**YIĞIN VARYANSLARI HETEROJEN İKEN ORTALAMALARIN
EŞİTLİĞİ HİPOTEZİ İÇİN BAZI TESTLER VE KARŞILAŞTIRMALARI**

Ahmet Metin YAVUZ

**YÜKSEK LİSANS TEZİ
İSTATİSTİK ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

TEMMUZ 2019

Ahmet Metin YAVUZ tarafından hazırlanan “YIĞIN VARYANSLARI HETEROJEN İKEN ORTALAMALARIN EŞİTLİĞİ HİPOTEZİ İÇİN BAZI TESTLER VE KARŞILAŞTIRMALARI” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi İstatistik Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Prof. Dr. Hamza GAMGAM

İstatistik Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Başkan: Prof. Dr. Bülent ÇELİK

İstatistik Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Üye: Doç. Dr. Rukiye DAĞALP

İstatistik Ana Bilim Dalı, Ankara Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Tez Savunma Tarihi: 17 / 07 / 2019

Jüri tarafından kabul edilen bu çalışmanın Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Sena YAŞYERLİ

Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Ahmet Metin YAVUZ

17 / 07 /2019

YIĞIN VARYANSLARI HETEROJEN İKEN ORTALAMALARIN EŞİTLİĞİ HİPOTEZİ İÇİN BAZI TESTLER VE KARŞILAŞTIRMALARI

(Yüksek Lisans Tezi)

Ahmet Metin YAVUZ

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Temmuz 2019

ÖZET

Yığın ortalamalarının eşitliği hipotezinin testi problemlerinde, yığın varyanslarının homojenliği varsayımının sağlanmadığı durumla yaygın olarak karşılaşılır. Literatürde yığın varyansları homojen değilken normal dağılımlı yığınların ortalamalarının eşitliği hipotezinin testi için önerilen birçok test vardır. Bu çalışmada bu amaçla önerilen Cochran Testi, James Testi, Welch Testi, Brown- Forsythe Testi, Alexander ve Govern Testi, Düzenlenmiş (Modified) Brown-Forsythe Testi, Düzenlenmiş (Modified) Welch Testi, Tek Aşamalı Test, Güvenilir (Fiducial) Yaklaşım Testi ve Parametrik Bootstrap Testi kısaca tanıtılmış ve Monte Carlo Simülasyon Yöntemi ile karşılaştırılmışlardır. Bu karşılaştırmalarda yığın sayısı, örnek hacimleri, varyans değerleri ve ortalama değerleri gibi tasarım parametreleri ile çeşitli simülasyon senaryoları kullanılmıştır. Bu simülasyon çalışmasında testler deneysel I. tip hata oranı ve testin gücü bakımından karşılaştırılmıştır. Simülasyon sonuçlarına göre James Testi I. tip hata oranı bakımından diğer testlerden daha iyi sonuçlar vermiştir. Güç sonuçlarına göre de, Cochran Testi'nin genel olarak diğer testlerden daha güçlü olduğu görülmüştür.

Bilim Kodu : 20506

Anahtar Kelimeler : Simülasyon, ANOVA, James, Cochran, Bootstrap

Sayfa Adedi : 57

Danışman : Prof. Dr. Hamza GAMGAM

SOME TESTS AND COMPARISONS FOR HYPOTHESIS OF EQUAL MEANS WHEN
POPULATION VARIANCES ARE HETEROGENEOUS

(M. Sc. Thesis)

Ahmet Metin YAVUZ

GAZİ UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

July 2019

ABSTRACT

In the problems of the test of the hypothesis indicating the equality of population means, it is commonly encountered that the assumption of the homogeneity of population variances is not assured. In the literature, there are many tests proposed for the test of hypothesis indicating the equality of means for the populations with normal distributions when the population variances are not homogeneous. In this study, Cochran Test, James Test, Welch Test, Brown-Forsythe Test, Alexander and Govern Test, Brown-Forsythe Test, Modified Welch Test, One-Stage Test, Fiducial Approach Test and Parametric Bootstrap Test have been briefly introduced and compared to Monte Carlo Simulation Method. In these comparisons, various simulation scenarios have been used with design parameters such as population number, sample size, variance values and mean values. In this simulation study, these tests have been compared in terms of experimental type I error rate and the power of the test. According to the simulation results, the James Test has given better results than the other tests in terms of type I error rate. According to the power results, it has been seen that the Cochran Test is more powerful than the other tests in general.

Science Code : 20506

Key Words : Simulation, ANOVA, James, Cochran, Bootstrap

Page Number : 57

Supervisor : Prof. Dr. Hamza GAMGAM

TEŞEKKÜR

Çalışmanın düzgün bir şekilde ilerlemesi için zaman ve emek harcayarak bilgi ve deneyimleri ile yol gösteren hocam Prof. Dr. Hamza GAMGAM'a katkılardan dolayı; bilgi ve deneyimleri ile özellikle yazılım konusunda destek veren hocam Doç. Dr. Bülent ALTUNKAYNAK'a; bu çalışmayı yapmam için izin ve destek veren amirim "Ticaret Bakanlığı" Daire Başkanı Niyazi ALBAYRAK'a; tüm hayatım boyunca her türlü desteği veren aileme teşekkürlerimi sunarım.



İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT	v
TEŞEKKÜR	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ	viii
ŞEKİLLERİN LİSTESİ	xi
SİMGELER VE KISALTMALAR.....	xii
1. GİRİŞ	1
2. YIĞIN VARYANSLARI HETEROJEN İKEN ORTALAMALARIN EŞİTLİĞİ HİPOTEZİ İÇİN BAZI TESTLER	11
2.1. Cochran Testi	11
2.2. James Testi.....	12
2.3. Welch Testi	14
2.4. Brown-Forsythe Testi	15
2.5. Alexander ve Govern Testi	16
2.6. Düzenlenmiş (Modified) Brown-Forsythe Testi	17
2.7. Düzenlenmiş (Modified) Welch Testi	18
2.8. Tek Aşamalı Test.....	19
2.9. Güvenilir (Fiducial) Yaklaşım Testi.....	22
2.10. Parametrik Bootstrap Testi.....	26
3. GERÇEK VERİ UYGULAMASI	29
4. SİMÜLASYON ÇALIŞMASI	35
5. SONUÇ	51
KAYNAKLAR	53
ÖZGEÇMİŞ	57

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. Çiçek türleri için taç yaprak uzunlukları.....	29
Çizelge 3.2. Çiçek türleri için taç yaprak uzunluklarına ilişkin tanımlayıcı istatistikler	30
Çizelge 3.3. Normal dağılıma uygunluk için Shapiro-Wilk uyum iyiliği testi sonuçları	31
Çizelge 3.4. Levene testi sonuçları	31
Çizelge 3.5. Test sonuçları	31
Çizelge 3.6. Tohum türleri için mısır verim değerleri	32
Çizelge 3.7. Tohum türleri için mısır verimine ilişkin tanımlayıcı istatistikler.....	32
Çizelge 3.8. Normal dağılıma uygunluk için Shapiro-Wilk uyum iyiliği testi sonuçları	33
Çizelge 3.9. Levene testi sonuçları	33
Çizelge 3.10. Test sonuçları.....	34
Çizelge 4.1. $k = 3$, $n_1 = n_2 = n_3 = 5$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları.....	36
Çizelge 4.2. $k = 3$, $n_1 = n_2 = n_3 = 10$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları.....	36
Çizelge 4.3. $k = 3$, $n_1 = n_2 = n_3 = 20$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları.....	37
Çizelge 4.4. $k = 3$, $n_1 = 7$, $n_2 = 9$, $n_3 = 11$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları	38
Çizelge 4.5. $k = 4$, $n_1 = n_2 = n_3 = n_4 = 10$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları	38
Çizelge 4.6. $k = 4$, $n_1 = n_2 = n_3 = n_4 = 20$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları	39
Çizelge 4.7. $k = 4$, $n_1 = n_2 = n_3 = n_4 = 25$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları	39
Çizelge 4.8. $k = 4$, $n_1 = 5$, $n_2 = 10$, $n_3 = 15$, $n_4 = 20$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları	40

Çizelge	Sayfa
Çizelge 4.9. $k = 4$, $n_1 = 10$, $n_2 = 15$, $n_3 = 20$, $n_4 = 25$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları	40
Çizelge 4.10. $k = 5$, $n_1 = n_2 = n_3 = n_4 = n_5 = 10$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları	41
Çizelge 4.11. $k = 5$, $n_1 = n_2 = n_3 = n_4 = n_5 = 20$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları	42
Çizelge 4.12. $k = 6$, $n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = 10$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları	42
Çizelge 4.13. Nominal $\alpha = 0,05$, $k = 3$, $n_1 = n_2 = n_3 = 15$ ve $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$ iken testlerin güç değerleri	43
Çizelge 4.14. Nominal $\alpha = 0,05$, $k = 3$, $n_1 = n_2 = n_3 = 15$ ve $\sigma_1^2 = 1$, $\sigma_2^2 = 1.5$, $\sigma_3^2 = 2$ iken testlerin güç değerleri	43
Çizelge 4.15. Nominal $\alpha = 0,05$, $k = 3$, $n_1 = n_2 = n_3 = 15$ ve $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_3^2 = 3$ iken testlerin güç değerleri	44
Çizelge 4.16. Nominal $\alpha = 0,05$, $k = 3$, $n_1 = 7$, $n_2 = 9$, $n_3 = 11$ ve $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$ iken testlerin güç değerleri	44
Çizelge 4.17. Nominal $\alpha = 0,05$, $k = 3$, $n_1 = 7$, $n_2 = 9$, $n_3 = 11$ ve $\sigma_1^2 = 1$, $\sigma_2^2 = 1.5$, $\sigma_3^2 = 2$ iken testlerin güç değerleri	45
Çizelge 4.18. Nominal $\alpha = 0,05$, $k = 3$, $n_1 = 7$, $n_2 = 9$, $n_3 = 11$ ve $\sigma_1^2 = 3$, $\sigma_2^2 = 2$, $\sigma_3^2 = 1$ iken testlerin güç değerleri.....	45
Çizelge 4.19. Nominal $\alpha = 0,05$, $k = 4$, $n_1 = n_2 = n_3 = n_4 = 10$ ve $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$ iken testlerin güç değerleri	46
Çizelge 4.20. Nominal $\alpha = 0,05$, $k = 4$, $n_1 = n_2 = n_3 = n_4 = 10$ ve $\sigma_1^2 = 1$, $\sigma_2^2 = 1.5$, $\sigma_3^2 = 2$, $\sigma_4^2 = 2.5$ iken testlerin güç değerleri	46
Çizelge 4.21. Nominal $\alpha = 0,05$, $k = 4$, $n_1 = n_2 = n_3 = n_4 = 10$ ve $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_3^2 = 3$, $\sigma_4^2 = 4$ iken testlerin güç değerleri	47

Çizelge	Sayfa
Çizelge 4.22. Nominal $\alpha = 0,05$, $k = 5$, $n_1 = n_2 = n_3 = n_4 = n_5 = 5$ ve $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = 1$ iken testlerin güç değerleri	47
Çizelge 4.23. Nominal $\alpha = 0,05$, $k = 5$, $n_1 = n_2 = n_3 = n_4 = n_5 = 5$ ve $\sigma_1^2 = 1$, $\sigma_2^2 = 1.5$, $\sigma_3^2 = 2$, $\sigma_4^2 = 2.5$, $\sigma_5^2 = 3$ iken testlerin güç değerleri	48
Çizelge 4.24. Nominal $\alpha = 0,05$, $k = 5$, $n_1 = n_2 = n_3 = n_4 = n_5 = 5$ ve $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_3^2 = 3$, $\sigma_4^2 = 4$, $\sigma_5^2 = 5$ iken testlerin güç değerleri	48
Çizelge 5.1. Simülasyon sonuçlarına göre iyi sonuç veren testler.....	52



ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 3.1. Çiçek türleri için taç yaprak uzunluklarına ilişkin kutu grafiği	30
Şekil 3.2. Tohum türleri için mısır verimine ilişkin kutu grafiği.....	33



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	Açıklamalar
A	Alexander-Govern test istatistiği
B	Brown-Forsythe test istatistiği
C	Cochran test istatistiği
F	F dağılımı
n	Örnek hacmi
w	Ağırlık katsayısı
W	Welch test istatistiği
φ	Ağırlık fonksiyonu
$\tilde{F}^1$	Tek aşamalı test istatistiği
H_0	Yokluk hipotezi
H_1	Alternatif hipotez
J_α	James testi kritik değeri
$S^2_{[b]}$	En büyük örnek varyansı
S^2	Örnek varyansı
T_F	Tek aşamalı test için kritik değeri
W^*	Düzeltilmiş Welch test istatistiği
$\bar{X}$	Örnek ortalaması
σ^2	Yığın varyansı
χ^2	Ki-kare dağılımı
μ	Yığın ortalaması

Kısaltmalar	Açıklamalar
AG	Alexander ve Govern Testi
BF	Brown-Forsythe Testi
CF	Cochran Testi

Kısaltmalar**Açıklamalar**

FA	Güvenilir (Fiducial) Yaklaşım Testi
J	James Testi
MBF	Düzenlenmiş (modified) Brown-Forsythe Testi
MW	Düzenlenmiş (modified) Welch Testi
OS	Tek Aşamalı Test
PB	Parametrik Bootstrap Testi
W	Welch Testi



1. GİRİŞ

Araştırmalarda iki veya daha fazla sayıda yığın ortalamalarının eşitliği hipotezinin test edilmesi problemi ile yaygın olarak karşılaşılır. Yığın ortalamalarının eşitliği hipotezinin test edilmesi problemi literatürde varyans analizi olarak bilinir. Kullanılan varyans analizi modeli faktör sayısına göre isimlendirilir. İlgilenilen problemde sadece bir faktörün bağımlı değişken üzerindeki etkisi araştırılıyorsa, buna “tek faktörlü varyans analizi” adı verilir.

İlgilenilen tek faktörlü varyans analizi modeli, i . yığının ortalaması μ_i , hata terimi ε_{ij} olmak üzere,

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma_i^2), \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \quad (1.1)$$

şeklinde ifade edilir.

Tek faktörlü varyans analizi problemlerinde, her biri normal dağılıma sahip yığın sayısı k , bunlardan i . yığının bilinmeyen ortalaması μ_i ve genel ortalama da μ olmak üzere, yokluk ve alternatif hipotezler

$$H_0: \mu_1 = \dots = \mu_k = \mu \quad (1.2)$$

ve

$$H_1: \mu_i \text{ parametrelerinin en az biri farklıdır} \quad (1.3)$$

biçiminde ifade edilir.

Örnek sayısı k , i . örnek hacmi n_i , $i = 1, \dots, k$, olmak üzere i . örnekteki j . birimin açıklanan değişken ölçüm değeri X_{ij} , $j = 1, \dots, n_i$ ile gösterilsin. i . örnek için ortalama ve varyans, sırasıyla,

$$\bar{X}_i = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i} \quad (1.4)$$

ve

$$S_i^2 = \frac{\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{n_i - 1} \quad (1.5)$$

olarak tanımlansın. Genel örnek ortalaması $\bar{X}$ ve toplam örnek hacmi n , sırasıyla,

$$\bar{X} = \frac{\sum_{i=1}^k \bar{X}_i n_i}{n} \quad (1.6)$$

ve

$$n = \sum_{i=1}^k n_i \quad (1.7)$$

olmak üzere, Eş. 1.3'deki alternatif hipoteze karşı Eş. 1.2'deki yokluk hipotezinin testi için F istatistiği,

$$F = \frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2 / (k-1)}{\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 / (n-k)} \quad (1.8)$$

olarak tanımlanır (Scheffe,1959:25,87). Yokluk hipotezinin doğruluğu altında bu istatistik bilinen $k - 1$ ve $n - k$ serbestlik dereceli F dağılımına sahiptir. Buna göre bu istatistik için hesaplanan değer F_h olmak üzere, $P(F_{k-1,n-k} > F_h) < \alpha$ ise $1 - \alpha$ güven düzeyinde yığın ortalamalarının eşitliği yokluk hipotezi reddedilir.

Tek faktörlü varyans analizi problemlerinde ortalamaların eşitliği hipotezinin testi için sağlanması gereken bazı varsayımlar vardır. Tek faktörlü varyans analizinde ortalamaların eşitliği hipotezinin test edilmesi amacıyla klasik bir yöntem olan F testinin kullanılabilmesi için,

- Örnekler bağımsız ve rassal olarak seçilmiştir
- Örneklerin her biri normal dağılımdan gelmiştir
- Örneklerin geldikleri yığınların varyansları homojendir

varsayımlarının sağlanıyor olması gerekir.

Yukarıdaki varsayımların sağlanması durumunda yığın ortalamalarının eşitliği hipotezini test etmek için klasik F testi kullanılır. Çünkü bu varsayımların sağlanması durumunda klasik F testi diğer testlere göre testin gücü bakımından daha iyi sonuçlar verir. Bir başka ifade ile gerçekte yanlış olan yokluk hipotezini reddetme olasılığı olan testin gücü, diğer testlere göre klasik F testi için daha büyüktür. Klasik F testinin kullanılabilmesi için gerekli olan varsayımların sağlanmaması durumunda alternatif olan testlerden biri parametrik olmayan testlerdir. Bu testlerin kullanılabilmesi için bağımsızlık ve rassallık varsayımlarının sağlanması yeterlidir. Fakat parametrik olmayan testlerin uygulanması durumunda oluşan bilgi kaybı bu testlerin gücünü azaltmaktadır. Buna ek olarak, yapılan birçok çalışma varsayımların sağlandığı durumda parametrik testlerin parametrik olmayan testlere göre hem I. tip hata oranı hem de II. tip hata oranı açısından daha iyi sonuçlar verdiğini göstermiştir (Keselman, Rogan ve Feir-Walsh 1977).

Varyans analizi problemlerinde örneklerin her birinin normal dağılımlı yığından gelip gelmediğinin araştırılması gerekir. Öngörülen dağılıma uygunluk için kullanılan testler “uyum iyiliği testleri” olarak bilinir. Normal dağılıma uygunluk için literatürde çok sayıda test bulunmaktadır. “Örneklerin her biri normal dağılımlı yığından gelmiştir” hipotezini test etmek için Ki-Kare, Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk gibi testler kullanılabilir (Gamgam ve Altunkaynak, 2012:35,72). Yapılan uyum iyiliği testleri sonucunda yığınlardan biri bile normal dağılımlı değilse varyans analizi için F testinin kullanılması uygun değildir. Literatürde, normal olmama durumunun F testinin gücü üzerindeki etkileri için David ve Johnson (1951), Srivastava (1959) ve Tiku (1971) tarafından çalışmalar yapılmıştır. Bu çalışmaların sonuçları, yığının dağılımının normal dağılımdan az farklılaşması durumunun F testinin gücü üzerinde çok az etkisinin olduğunu göstermektedir. Normal dağılımdan kısmen uzaklaştığında örneğin, çarpıklık (skewness) değerinin +0,5 olan yığınlar için testin gücü 0,8’e kadar yükselebildiğini tespit etmişlerdir. Ayrıca, çarpıklık değerinin +0,5’ten büyük olan yığınlar için ise testin gücünün azaldığını gözlemlemişlerdir. Bunun yanında, yığın için basıklık değeri negatif iken bu testin gücünün azalma yönünde bir eğilime sahip olduğu anlaşılmaktadır (Harwell, Rubinstein, Hayes ve Olds, 1992).

Her bir örnek için “Örnek normal dağılımlı yığından gelmiştir” hipotezlerinin hepsinin de kabul edilmediği durumda yığın ortalamalarının eşitliği hipotezinin testi için çok sayıda çalışma yapılmıştır. Clinch ve Keselman (1982) tarafından yapılan çalışmada, yığınların dağılımları normal değil, varyansların heterojen ve örnek hacimleri de farklı iken Brown-Forsthye (1974b) tarafından önerilen Brown-Forsthye testinin parametrik testlere iyi bir alternatif olduğu Monte Carlo simülasyon çalışmasıyla gösterilmiştir. Hatta Brown-Forsthye testinin, Klasik F testi ve Welch testi arasında iyi sonuç veren test olduğunu da ortaya koymuştur. Bunun yanı sıra Welch (1951) tarafından önerilen Welch testinin de alternatif olabileceği sonucuna da varmıştır. Bu çalışmada ayrıca, Klasik F testinin varsayımlarının sağlanmamasından oldukça etkilendiği de gösterilmiştir.

Yığınlar normal dağılımlı olsalar bile varyanslarının homojen olmadığı durumla sıkça karşılaşilmektedir. Bu nedenle, varyansların homojenliği varsayımının sağlanıp sağlanmadığı da araştırılmalıdır. Bu amaçla, yığınların dağılımlarının normal olduğu varsayımına dayanan Bartlett (1946) tarafından önerilen Bartlett testi bu varsayıma dayanan diğer testlere göre daha iyi sonuçlar verdiği için daha çok tercih edilmektedir. Bu testin yanı sıra normallik varsayımının sağlanmamasından etkilenmeyen testler de vardır. Levene (1960), Brown-Forsthe (1974a) ve Layard (1973) tarafından önerilen testler de varyansların homojenlik testi için normal dağılım varsayımı gerektirmeyen ve yaygın olarak kullanılan testlerdir. Layard χ^2 testi, Jackknife testine dayalı olarak geliştirilmiş bir testtir. Brown ve Forsythe (1974a) tarafından yapılan çalışmada normallik varsayımı sağlanmadığında özellikle Levene testinin sağlam (robust) sonuçlar sağladığını göstermiştir.

Varyansların homojenliği varsayımı sağlanmıyor ve heterojenlik durumu düşük ise klasik F testinin gücünün ciddi bir şekilde etkilenmediği bulgusuna ulaşılmıştır (Box, 1954). Ancak örnek hacimleri eşit olmadığında ve varyansların heterojenliği yüksek olduğunda, klasik F testinin sağlam (robust) olmadığı görülmüştür (Boneau, 1960; Bishop, 1976; Reed ve Stark, 1995). Bir diğer deyişle; varyansların homojenliği sağlanmıyor ve özellikle örnek hacimleri farklı iken klasik F testinin gücü, yani gerçekte yanlış olan yokluk hipotezini reddetme olasılığı küçülmektedir. Bu sebeple varyansların homojenliği varsayımının sağlanması en kritik varsayım olarak bilinir. Diğer yandan, Rogan ve Keselman (1977) örnek hacimleri eşit iken varyansların heterojenliğinin tüm durumları için klasik F testinin sağlam bir istatistik olmadığı sonucuna ulaşmıştır. Ayrıca, varyansların heterojenliği sorununu çözmek amacıyla dönüşüm yöntemleri de uygulanmaktadır. Dönüşüm yöntemleri uygulandıktan sonra

varyansların homojenliği varsayımı sağlanabileceği ve böylelikle parametrik test kullanılabileceği yönünde görüşler mevcuttur. Diğer yandan Budescu ve Applebaum (1981) tarafından yapılan çalışma, dönüşümlerin I. tip hata oranları ve testin gücü üzerinde etkili olmadığı sonucunu vermiştir. Bu sebepten ötürü dönüşüm tekniklerinin kullanılmamasının daha uygun olacağı sonucunu varılabilir. Clinch ve Keselman (1982), dönüşüm yöntemleri yerine, Hays (1973)'in çalışmasındaki eşit örnek tavsiyesinin dikkate alınması gerektiği sonucuna ulaşmışlardır.

Yığın varyanslarının homojen olmadığı durumlarda, normal dağılıma sahip yığınların ortalamalarının eşitliği hipotezini test etmek için çok sayıda test geliştirilmiştir. İlk yapılan çalışmalarda iki yığın ortalamasının eşitliği hipotezinin test edilmesine odaklanılmıştır. Bu konu literatürde Behrens-Fisher problemi olarak adlandırılmaktadır. Cochran (1937) tarafından yapılan çalışmada birkaç ortalamanın eşitliği hipotezinin testi için kesin sonuçlar verecek bir test geliştirememiştir ve kesin sonuçlar verecek test için daha fazla çalışma yapmaya ihtiyaç duyulduğunu belirtmiştir. Behrens-Fisher probleminin çözümü için birçok yazar, varyansların bilinmediği ve heterojen olduğu durumlar için asimptotik yöntemler önermiştir. Welch (1938), iki yığın ortalamasının eşitliğinin testi için t dağılımına dayalı kritik değerleri elde etmiştir. Bu kritik değerleri üretirken yığınların normal dağılımlı olduğu ve varyansların homojen olmadığı durumu göz önünde bulundurmuştur. Hsu (1938) ile Grunow (1951) tarafından yapılan çalışmalar, Welch (1938)'in çalışmasının bulgularını destekler niteliktedir (Box, 1954). Welch (1938) çalışmasında elde ettiği sonuçlardan yararlanarak iki yığın ortalamasının eşitliliği hipotezini test etmek için bir test geliştirmiştir (Welch, 1947). Aspin (1948) tarafından yapılan çalışmayla Welch testine dayalı olarak iki yığın ortalamasının eşitliği hipotezinin testi için daha iyi sonuçlar verecek bir test geliştirmeye çalışmıştır. Geliştirdiği test Welch-Aspin testi olarak adlandırılmaktadır ve yaygın olarak kullanılmamaktadır. İki yığın ortalamasının eşitliği hipotezinin testinde iki örnek t-testi ve Welch testi arasında seçim yapmak için, önce varyansların homojenliği test edilmelidir. Geleneksel yaklaşım olarak eğer varyansların homojenliği kabul edilir ise t-testi kullanılır, aksi durumda Welch testi kullanılır. Welch (1951), ikiden fazla yığın ortalamalarının eşitliği hipotezini test etmek için daha önceki çalışmalarını geliştirerek yeni bir yaklaşım önermiştir. Daha önce bahsedilen geleneksel yaklaşım için Moser, Stevens ve Watts (1989) tarafından yapılan I. tip hata ve güç çalışmaları, varyans homojenliği hakkındaki ön testin yüzeysel olduğunu ve Welch testinin de tüm parametre yapılandırmaları

için yeterli olduğunu göstermiştir. Bu sonuçları, Rasch, Kubinger ve Moder (2011) tarafından yapılan çalışma da destekler niteliktedir.

Literatürde ikiden fazla yığın ortalamasının eşitliği hipotezinin testi için, Behrens–Fisher probleminin çözümünü genelleştiren Welch testi önerilen ilk test olarak bilinir. Bu problem üzerine diğer bir çalışma da James (1951), Welch testinden yararlanarak iki ve daha fazla yığın ortalamasının eşitliği hipotezini test etmek için James testini geliştirmiştir. James testi literatürde “James ikinci sıra istatistiği” olarak da bilinir. Küçük örnek hacimleri için Welch (1951) tarafından önerilen kritik değerler, James tarafından önerilenlere göre daha iyi bir yaklaşımdır (Brown-Forsythe, 1974b). Welch testine alternatif testler de geliştirilmiştir. Bu testler arasında Brown ve Forsythe (1974b) tarafından varyansların homojen olmadığı durumda yığın ortalamalarının eşitliği hipotezini test etmek için geliştirilmiş olan alternatif bir yaklaşım olarak iyi sonuçlar vermektedir.

Dudewicz ve Dalal (1975) tarafından yapılan çalışma, homojen olmayan varyanslar altında k sayıda normal dağılımlı yığın ortalamalarının karşılaştırılmasında, tüm örnek hacimleri ve farklı parametre kombinasyonları için tatmin edici performans gösteren tek aşamalı bir yöntem olmadığını göstermiştir. Bishop ve Dudewicz (1978) ise ikiden fazla yığın ortalamasının eşitliği hipotezini test etmek için tam olasılık dağılımlı iki aşamalı bir örnekleme yöntemi önermişlerdir. Bu yöntem, Dudewicz ve Dalal (1975) tarafından yapılan çalışmaya dayalı olarak geliştirmişlerdir. Geliştirilen iki aşamalı yöntem ilk aşamada kullanılan örnek hacminden ikinci aşamada daha fazla hacimli örnek gerektirmektedir. Bu ilave örnekler gerçek veri uygulamalarında daha masraflı ve daha fazla zaman gerektirmekte olduğundan dezavantajlı bir durum yaratmaktadır. Bu testin yaratmış olduğu dezavantajlı durumdan kurtulmak için S. Chen ve H. Chen (1998) tek aşamalı yöntem önermişlerdir. Yapılan çalışmalarda, tek aşamalı yöntemin iki aşamalı yönteme göre testin gücü bakımından daha iyi sonuçlar verdiği gözlenmiştir. Özellikle küçük hacimli örneklerde gerçekleştirilen simülasyon çalışmaları tek aşamalı yöntemin yeterli bir istatistik olduğunu göstermiştir. Tek aşamalı yöntem, iki aşamalı yöntem gibi ilave örnek gerektirmemektedir. Özellikle bu yönüyle tek aşamalı yöntem iki aşamalı yönteme göre daha avantajlıdır. Bu sebeple büyük hacimli örnekleri elde etmenin kolay olmadığı alanlarda iki aşamalı yöntem yerine tek aşamalı yöntem tercih edilmektedir. Varyansların homojen olmadığı durumda yığın ortalamalarının eşitliği hipotezini test ederken, tek aşamalı yöntemin kullanılabilmesi için örnek hacmi 6 veya daha fazla olması gerekmektedir. Bu koşulun sağlanması

durumunda tek aşamalı yöntemin testin güç değerleri bakımından iyi sonuçlara sahip olduğu tespit edilmiştir (S. Chen ve H. Chen, 1998).

Wilcox (1988) tarafından yapılan çalışmada basitleştirilmiş bir yaklaşım önerilmiştir. Bu yaklaşım birçok açıdan tatmin edici bir istatistiktir. Fakat Wilcox (1988), James testinin, kendi yaklaşımından genel olarak daha iyi sonuçlar verdiğini tespit etmiştir. Daha sonra Wilcox (1989) bir önceki çalışmasında ortaya koyduğu yaklaşımı geliştirmiştir. Ancak yapılan Monte Carlo simülasyon karşılaştırmalarında yine James yönteminin genellikle daha yüksek güç değerlerine sahip sonuçlar ürettiği tespit edilmiştir.

Rice ve Gaines (1989), yığın ortalamalarının eşitliği hipotezinin testi için Behren-Fisher'in T testini geliştirerek yeniden uyarlanmış (modified) F testini önermişlerdir. Bu testte, standart deneme-yanılma dönüşüm tekniği tercih edilmiştir. Bu testin, özellikle biyoloji alanında yapılan çalışmalarda karşılaşılan tek faktörlü varyans analizi problemlerinde kullanılması önerisinde bulunmuşlardır.

Tsui ve Weeranhandi (1989) tarafından yapılan çalışma ile diğer çalışmalarda yığın varyanslarının bilinmediği veya dağılım bilgisi bulunmadığı durumların incelenmediğini tespit etmişlerdir. Bu konu üzerine Tsui ve Weeranhandi (1989) tarafından literatürdeki çalışmalardan farklı olarak yeni bir çalışma yapmışlardır. Weerahandi (1995a), yığın ortalamalarının eşitliği hipotezini test etmek amacıyla kullanılacak “genelleştirilmiş p-değeri yaklaşımı” olarak bilinen bir yaklaşım önermiştir. Varyansların bilinmediği durum için F dağılımından yararlanarak genelleştirilmiş p-değeri yaklaşımıyla Behrens-Fisher problemleri için kritik değerleri elde etmiştir. Gamage ve Weerahandi (1998) tarafından yapılan çalışmada genelleştirilmiş p-değeri yaklaşımı ile Welch testinin birbirine yakın ve testin gücü bakımından iyi sonuçlar verdiğini saptamışlardır (Xu ve Wang, 2008). Gamage, Mathew ve Weerahandi (2004) genelleştirilmiş p-değeri yaklaşımını çok değişkenli problemler için kullanılacak şekilde geliştirmişlerdir.

Alexander ve Govern (1994), t-dağılımından yararlanarak yığın ortalamalarının eşitliği hipotezini test etmek için yeni bir yaklaşımla probleme çözüm üretmişlerdir. Elde edilen Monte Carlo simülasyon sonuçlarına göre, I. tip hata oranı bakımından bu test ile James testinin yakın sonuçlar verdiği görülmüştür.

Literatürde homojen olmayan varyans durumunda yığın ortalamaların eşitliği hipotezinin testi için yapılan çalışmalar, genel olarak, yığın sayısının az olduğu durumları içermektedir. İlk olarak iki yığın olduğu durum göz önünde bulundurulmuştur. Sonrasında yapılan çalışmalarda yığın sayısının ikiden fazla olduğu durumlar çalışılmıştır. “Dajani (2002) tarafından yapılan çalışma, yığın sayısının fazla olduğu ve örnek hacimlerinin küçük olduğu durumda hiçbir test istatistiğinin I. tip hata oranı açısından tatmin edici sonuçlar vermediği sonucunu ortaya koymuştur” (Krishnamoorthy, Lu ve Mathew, 2007).

Literatürde bu amaçla kullanılan Parametrik ve parametrik olmayan bootstrap yöntemleri de mevcuttur. Krishnamoorthy, Lu ve Mathew (2007), tek faktörlü varyans analizinde normal dağılımlı yığın ortalamalarının eşitliği hipotezinin testi için Parametrik Bootstrap yöntemini geliştirmişlerdir. Bu test örnek hacimlerinin küçük olduğu durumlarda I. tip hata ve testin gücü bakımlarından iyi sonuçlar vermektedir. Küçük hacimli örnekler ve homojen olmayan varyanslar durumunda, Parametrik Bootstrap testinin yığın ortalamalarının eşitliği hipotezinin testi için verdiği sonuçların diğer testlerden daha iyi olmasından dolayı tercihi gitgide artmaktadır.

Weerahandi ve Krishnamoorthy (2017) tarafından yapılan çalışma daha önce önerdikleri testlerin karşılaştırmaları ve literatürdeki yaklaşımların uygunluğu üzerinedir. Bu çalışma sonucunda yığın ortalamalarının eşitliği hipotezinin testi için alternatif test olarak klasik yaklaşım, Bootstrap yaklaşımı, Fiducial yaklaşımı, Bayes yaklaşımı ve genelleştirilmiş p-değeri yaklaşımının kullanılabileceğini öne sürmüşlerdir. Li, Wang ve Liang (2011) tarafından yapılan çalışmada bu görüşü destekler niteliktedir. Hatta karşılaştırılan yığın sayısının az olduğu durumlarda önermiş oldukları ve genelleştirilmiş Fiducial yaklaşımı olarak adlandırdıkları yaklaşımın Parametrik Bootstrap yaklaşımından daha iyi sonuçlar verdiğini savunmuşlardır. Ayrıca varyans analizi problemlerinde modeldeki faktör sayısının birden fazla olduğu durumlar da mevcuttur. Karışık modellerde, Parametrik Bootstrap yaklaşımının başarısız olduğu durumlar için genelleştirilmiş p-değerine dayalı testlerin kullanılmasının daha uygun olacağı yönünde araştırmacılara öneride bulunmuşlardır (Weerahandi ve Krishnamoorthy, 2017).

İkinci bölümde, varyansları homojen olmayan normal dağılımlı yığınların ortalamalarının eşitliği hipotezi için Cochran (1937) tarafından önerilen Cochran testi, James (1951) tarafından önerilen James testi, Welch (1951) tarafından önerilen Welch testi, Brown-

Forsythe (1974) tarafından önerilen Brown- Forsythe testi, Alexander ve Govern (1994) tarafından önerilen Alexander ve Govern testi, Mehrotra (1997) tarafından önerilen Düzenlenmiş (modified) Brown-Forsythe testi, Hartung, Argac ve Makambi (2002) tarafından önerilen Düzenlenmiş (modified) Welch testi, S. Chen ve H. Chen (1998) tarafından önerilen Tek Aşamalı test, Li ve diğerleri (2011) tarafından önerilen Güvenilir (Fiducial) yaklaşım testi ve Krishnamoorthy ve diğerleri (2007) tarafından önerilen Parametrik Bootstrap testi kısaca tanıtılmıştır.

Üçüncü bölümde, gerçek veri seti üzerinde bu testlerin uygulamaları yapılmıştır.

Bu testlerin karşılaştırılmaları amacıyla dördüncü bölümde çeşitli yığın sayıları, örnek hacimleri, varyans değerleri ve ortalama değerleri gibi tasarım parametreleri ile simülasyon çalışması yapılmıştır. Bu simülasyon çalışması ile deneysel I. tip hata oranı ve testin gücü bakımından bu testlerin karşılaştırılmaları yapılmıştır.

Beşinci ve son bölümde de bu simülasyon çalışmasına dayalı olan genel değerlendirmelere yer verilmiştir.



2. YIĞIN VARYANSLARI HETEROJEN İKEN ORTALAMALARIN EŞİTLİĞİ HİPOTEZİ İÇİN BAZI TESTLER

Yığın varyansları homojen olmadığında klasik F testinin yığın ortalamalarının eşitliği hipotezi için istatistiksel olarak iyi sonuçlar vermediği daha önce vurgulanmıştı. Bu nedenle yığın ortalamalarının eşitliği hipotezinin testinde bilinmeyen yığın varyansları heterojen iken birçok test istatistiği önerildiği giriş bölümünde belirtilmişti. Bu bölümde normal dağılımlı ancak homojen varyanslı olmayan yığınlar için Eş. 1.3'deki alternatif hipoteze karşı Eş. 1.2'deki yokluk hipotezinin testi için literatürde yaygın olarak kullanılan Cochran testi, James testi, Welch testi, Brown-Forsythe testi, Alexander ve Govern testi, Düzenlenmiş (modified) Brown-Forsythe testi, Düzenlenmiş (modified) Welch testi, Tek Aşamalı test, Fiducial Yaklaşım testi ve Parametrik Bootstrap testi kısaca tanıtılmıştır. Bu testler aşağıda sırasıyla ele alınmıştır.

2.1. Cochran Testi

Yığın sayısı k ve i . yığından seçilen örnek hacmi n_i olmak üzere, i . örnek

$$X_{i1}, X_{i2}, \dots, X_{in_i}, i = 1, \dots, k$$

ile gösterilsin. Ayrıca i . örnek için ortalama $\bar{X}_i$ ve genel örnek ortalaması da $\bar{X}_.$ ile gösterilsin ve bunlar aşağıdaki gibi tanımlansın.

$$\bar{X}_i = \sum_{j=1}^{n_i} X_{ij} / n_i, i = 1, \dots, k, j = 1, \dots, n_i$$

$$\text{ve} \tag{2.1}$$

$$\bar{X}_. = \sum_{i=1}^k n_i \bar{X}_i / n, k \geq 2 \text{ ve } n_i \geq 2$$

Burada n toplam örnek hacmi olup

$$n = \sum_{i=1}^k n_i \tag{2.2}$$

olarak tanımlanır. Örnek varyansları

$$S_i^2 = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 / (n_i - 1)$$

ve ağırlıklar (2.3)

$$w_i = n_i / S_i^2 ,$$

$$h_i = w_i / \sum_{i=1}^k w_i$$

olmak üzere, Eş. 1.3'deki alternatif hipoteze karşı Eş. 1.2'deki ortalamaların eşitliği yokluk hipotezinin testi için Cochran (1937)

$$C = \sum_{i=1}^k w_i (\bar{X}_i - \sum_{i=1}^k h_i \bar{X}_i)^2 \quad (2.4)$$

istatistiğini önermiştir. Yokluk hipotezinin doğruluğu altında C istatistiği yaklaşık olarak $k - 1$ serbestlik dereceli χ^2 dağılımına sahiptir. Buna göre,

$$P(\chi_{k-1}^2 > C) < \alpha \quad (2.5)$$

ise H_0 hipotezi reddedilir.

2.2. James Testi

James (1951), Welch testine dayalı olarak bir test geliştirmiştir. Normal dağılımlı k sayıda yığın olsun. Bunlardan i . yığından, $i = 1, \dots, k$, n_i hacimli örnek $X_{i1}, \dots, X_{in_i}$, örnek ortalaması $\bar{X}_i$ ve örnek varyansı da S_i^2 ile gösterilsin. Bu istatistikler, sırasıyla,

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$$

ve (2.6)

$$S_i^2 = \frac{1}{n_i-1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2, i = 1, \dots, k$$

olarak tanımlanır. Eğer σ^2 biliniyorsa, Eş. 1.2'deki yokluk hipotezinin Eş. 1.3'deki alternatif hipoteze karşı testi için

$$\begin{aligned} T_N(\bar{X}_1, \dots, \bar{X}_k; \sigma_1^2, \dots, \sigma_k^2) &= \sum_{i=1}^k \frac{n_i}{\sigma_i^2} \left[\bar{X}_i - \frac{\sum_{i=1}^k n_i \bar{X}_i / \sigma_i^2}{\sum_{i=1}^k n_i / \sigma_i^2} \right]^2 \\ &= \sum_{i=1}^k \frac{n_i}{\sigma_i^2} \bar{X}_i^2 - \frac{\left[\sum_{i=1}^k n_i \bar{X}_i / \sigma_i^2 \right]^2}{\sum_{i=1}^k n_i / \sigma_i^2} \end{aligned} \quad (2.7)$$

istatistiği önerilmiştir (James, 1951). Genel olarak yığın varyansları olan σ_i^2 'ler bilinmezler. Bu durumda S_i^2 , $i = 1, \dots, k$, istatistiği ile Eş. 2.7'deki σ_i^2 'lerin yer değiştirmesi ile test istatistiği,

$$T_N(\bar{X}_1, \dots, \bar{X}_k; S_1^2, \dots, S_k^2) = \sum_{i=1}^k \frac{n_i}{S_i^2} \bar{X}_i^2 - \frac{\left[\sum_{i=1}^k n_i \bar{X}_i / S_i^2 \right]^2}{\sum_{i=1}^k n_i / S_i^2} \quad (2.8)$$

olarak elde edilir. James (1951) , $T_N(\bar{X}_1, \dots, \bar{X}_k; S_1^2, \dots, S_k^2)$ istatistiği için kritik değer bulmak amacıyla bir yaklaşım önermiştir. Test istatistiği için

$$w_i = n_i / S_i^2, i = 1, \dots, k,$$

$$Q = \sum_{i=1}^k \frac{1}{n_i-1} \left(1 - \frac{w_i}{\sum_{j=1}^k w_j} \right)^2,$$

$$c_s = \frac{(\chi_{k-1}^2)^s}{(k-1)(k+1) \dots (k+2s-3)}, s = 1, 2, 3, 4$$

ve (2.9)

$$R_{st} = \sum_{i=1}^k \frac{1}{(n_i-1)^s} \left(\frac{w_i}{\sum_{j=1}^k w_j} \right)^t, s = 1, 2, t = 0, 1, 2, 3$$

olmak üzere, James (1951), $T_N(\bar{X}_1, \dots, \bar{X}_k; S_1^2, \dots, S_k^2)$ istatistiği için kritik değeri

$$\begin{aligned}
J_\alpha = & \chi_{k-1, \alpha}^2 + \frac{1}{2}(3c_2 + c_1)Q + \left\{ \frac{1}{16}(3c_2 + c_1)^2 \left(1 - \frac{k-3}{\chi_{k-1, \alpha}^2} \right) Q^2 \right. \\
& + \frac{1}{2}(3c_2 + c_1)[(8R_{23} - 10R_{22} + 4R_{21} - 6R_{12}^2 + 8R_{12}R_{11} - 4R_{11}^2) \\
& + (2R_{23} + 4R_{22} + 2R_{21} - 2R_{12}^2 + 4R_{21}R_{11} - 2R_{11}^2)(c_1 - 1) \\
& + \frac{1}{4}(-R_{12}^2 + 4R_{21}R_{11} - 2R_{12}R_{10} - 4R_{11}^2 + 4R_{11}R_{10} - R_{10}^2)(3c_2 - 2c_1 - 1)] \\
& + (R_{23} - 3R_{22} + 3R_{21} - R_{20})(5c_3 + 2c_2 + c_1) \\
& + \frac{3}{16}(R_{12}^2 - 4R_{23} + 6R_{22} - 4R_{22} + R_{20})(35c_4 + 15c_3 + 9c_2 + 5c_1) \\
& + \frac{1}{16}(-2R_{22} + 4R_{21} - R_{20} + 2R_{12}R_{10} - 4R_{11}R_{10} + R_{10}^2)(9c_4 - 3c_3 - 5c_2 - c_1) \\
& + \frac{1}{4}(-R_{22} + R_{11}^2)(27c_4 + 3c_3 + c_2 + c_1) \\
& \left. + \frac{1}{4}(R_{22} - R_{12}R_{11})(45c_4 + 9c_3 + 7c_2 + 3c_1) \right\} + 0((n_i - 1)^{-3}) \quad (2.10)
\end{aligned}$$

olarak önermiştir.

Buna göre, $T_N(\bar{X}_1, \dots, \bar{X}_k; S_1^2, \dots, S_k^2)$ istatistiği için gözlenen değer $T_N(\bar{x}_1, \dots, \bar{x}_k; s_1^2, \dots, s_k^2)$ olmak üzere

$$T_N(\bar{x}_1, \dots, \bar{x}_k; s_1^2, \dots, s_k^2) > J_\alpha \quad (2.11)$$

ise H_0 hipotezi reddedilir.

2.3. Welch Testi

Welch (1951) tarafından homojen olmayan varyanslar varsayımının sağlanmadığı durumda yığın ortalamalarının eşitliği hipotezini test etmek amacıyla çalışmalar yapılmıştır.

i . örnek hacmi n_i , i . örnek varyansı S_i^2 ve $w_i = n_i/S_i^2$, $i = 1, \dots, k$ olmak üzere Welch (1951), yığın ortalamalarının eşitliği hipotezinin testi için

$$W = \frac{T_N(\bar{X}_1, \dots, \bar{X}_k; S_1^2, \dots, S_k^2)/(k-1)}{1 + \left(\frac{2(k-2)}{k^2} - 1 \right) \sum_{i=1}^k \left(\frac{1}{n_i - 1} \right) (1 - w_i / \sum w_j)^2} \quad (2.12)$$

istatistiğini önermiştir. Burada

$$f = \left[\frac{3}{k^2 - 1} \sum_{i=1}^k \frac{1}{n_i - 1} \left(1 - \frac{w_i}{\sum w_i} \right)^2 \right]^{-1} \quad (2.13)$$

olmak üzere, Welch (1951), H_0 hipotezinin doğruluğu altında, W istatistiğinin yaklaşık olarak $k - 1$ ve f serbestlik dereceli F dağılımına sahip olduğunu göstermiştir.

Buna göre α 'nın verilen bir değeri ve W istatistiğinin gözlenen bir değeri w için

$$P(F_{k-1, f} > w) < \alpha \quad (2.14)$$

ise yığın ortalamalarının eşitliği hipotezi reddedilir.

2.4. Brown-Forsythe Testi

Brown ve Forsythe (1974), klasik F testinin düzenlenmiş (modified) bir biçimi olan Brown-Forsythe testini önermişlerdir. Bu test istatistiği

$$B = \frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2}{\sum_{i=1}^k (1 - n_i/n) S_i^2} \quad (2.15)$$

olarak önerilmiştir. Eş. 1.2'de verilen H_0 hipotezinin doğruluğu altında, B istatistiği $k - 1$ ve v serbestlik dereceli F dağılımına sahiptir. Bu F dağılım için v serbestlik derecesi

$$v = \frac{\left[\sum_{i=1}^k \left(1 - \frac{n_i}{n} \right) S_i^2 \right]^2}{\sum_{i=1}^k \frac{\left(1 - \frac{n_i}{n} \right)^2 S_i^4}{n_i - 1}} \quad (2.16)$$

olarak verilmiştir. Buna göre B istatistiği için hesaplanan değer B_h olmak üzere,

$$P(F_{k-1, v} > B_h) < \alpha \quad (2.17)$$

ise yığın ortalamalarının eşitliği hipotezi reddedilir.

2.5. Alexander ve Govern Testi

Varyansların homojen olmadığı durumda k bağımsız yığın ortalamalarının eşitliği hipotezinin testi için Alexander ve Govern (1994) tarafından bir yaklaşım önerilmiştir. Normal dağılımlı yığınlardan rassal örnek sayısı k , i . örnek hacmi n_i ve i . örnek ortalaması $\bar{X}_i$ ile gösterilsin. i . örnek ortalaması için standart hata

$$S_i = S_{\bar{X}_i} = \left\{ \frac{\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{n_i(n_i-1)} \right\}^{\frac{1}{2}}, i = 1, \dots, k, \quad (2.18)$$

ve ağırlık katsayıları da

$$w_i = \frac{1/S_i^2}{\sum_{i=1}^k 1/S_i^2} \quad (2.19)$$

olarak tanımlansın. Yukarıdaki ağırlık katsayıları için $\sum_{i=1}^k w_i = 1$ olduğu açıktır. Ağırlıklandırılmış genel ortalama X^+ ve her bir örnek için bir örnek t istatistiği t_i ile gösterilsin ve aşağıdaki gibi tanımlansın. Burada

$$X^+ = \sum_{i=1}^k w_i \bar{X}_i$$

ve (2.20)

$$t_i = \frac{\bar{X}_i - X^+}{S_i}$$

olarak tanımlanmıştır. Bu t_i istatistiği $v_i = n_i - 1$ serbestlik dereceli Student t dağılımına sahiptir (Alexander ve Govern, 1994). Bu yaklaşımda her bir örnek için, $a = v_i - 5$, $b = 48a^2$ ve $c = [a \ln(1 + t_i^2 / v_i)]^{1/2}$ olmak üzere, Standart normal dağılıma sahip olan

$$z_i = c + \frac{(c^3 + 3c)}{b} - \frac{(4c^7 + 33c^3 + 240c^3 + 855c)}{(10b^2 + 8bc^4 + 1000b)}, i = 1, \dots, k \quad (2.21)$$

değişkenleri tanımlansın (Alexander ve Govern, 1994). Ayrıca

$$\sum_{i=1}^k z_i^2 = A \quad (2.22)$$

olarak tanımlanan A istatistiği yaklaşık olarak $k - 1$ serbestlik derecesi ile χ^2 dağılır. Buna göre,

$$P(\chi_{k-1}^2 > A) < \alpha \quad (2.23)$$

ise yığın ortalamalarının eşitliği hipotezi reddedilir.

2.6. Düzenlenmiş (Modified) Brown-Forsythe Testi

Mehrotra (1997), Brown-Forsythe testini geliştirerek Düzenlenmiş Brown-Forsythe testini önermiştir. Varyansların heterojenliği altında, yığın ortalamalarının eşitliği hipotezinin testi için

$$B^* = \frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X}_{..})^2}{\sum_{i=1}^k \left(1 - \frac{n_i}{n}\right) S_i^2} \quad (2.24)$$

istatistiği önerilmiştir. Bu istatistik,

$$v_1 = \frac{\left[\sum_{i=1}^k \left(1 - \frac{n_i}{n}\right) S_i^2\right]^2}{\sum_{i=1}^k S_i^4 + \left[\sum_{i=1}^k n_i S_i^4 / n\right]^2 - 2 \sum_{i=1}^k n_i S_i^4 / n} \quad (2.25)$$

ve

$$v = \frac{\left[\sum_{i=1}^k \left(1 - \frac{n_i}{n}\right) S_i^2\right]^2}{\sum_{i=1}^k \frac{\left(1 - \frac{n_i}{n}\right)^2 S_i^4}{n_i - 1}} \quad (2.26)$$

olmak üzere, H_0 hipotezinin doğruluğu altında v_1 ve v serbestlik dereceli F dağılımına sahiptir. Buna göre,

$$P(F_{v_1, v} > B^*) < \alpha \quad (2.27)$$

ise H_0 hipotezi reddedilir.

2.7.Düzenlenmiş (Modified) Welch Testi

Hartung, Argac ve Makambi (2002), Welch testine dayanan Düzeltilmiş Welch testini önermişlerdir. Düzenlenmiş Welch testinde $w_i = n_i/S_i^2$ ağırlık katsayıları yerine Böckenhoff ve Hartung (1998) tarafından geliştirilen ağırlık katsayıları kullanılmıştır. Varyanslar homojen değil iken yığın ortalamalarının eşitliği hipotezinin testi için

$$W^* = \frac{\sum_{i=1}^k w_i^* (\bar{X}_i - \sum_{i=1}^k h_i^* \bar{X}_i)^2}{(k-1) + 2(k-2)(k+1)^{-1} \sum_{i=1}^k (n_i-1)^{-1} (1-h_i^*)^2} \quad (2.28)$$

istatistiğini önermişlerdir (Hartung ve diğerleri, 2002). Bu test istatistiğindeki ağırlık katsayılarını belirlemek için önce δ_1 ve δ_2 reel sayılar olmak üzere $1 \leq \varphi_i \leq [(n_i - 1) / (n_i - 3)]$ koşulunu sağlayan $\varphi_i = (n_i + \delta_1) / (n_i + \delta_2)$ katsayıları bulunur (Hartung ve diğerleri, 2002). Buna göre ağırlık katsayıları da

$$w_i^* = n_i / (\varphi_i S_i^2) \quad (2.29)$$

ve

$$h_i^* = \frac{w_i^*}{\sum_{i=1}^k w_i^*}, i = 1, \dots, k \quad (2.30)$$

olarak önerilmiştir (Hartung ve diğerleri, 2002). W^* istatistiği, H_0 hipotezinin doğruluğu altında $k - 1$ ve

$$v_w^* = \frac{k^2 - 1}{3 \sum_{i=1}^k (n_i - 1)^{-1} (1 - h_i^*)^2} \quad (2.31)$$

serbestlik dereceleri ile F dağılımına sahiptir. Buna göre

$$P(F_{k-1, v_w^*} > W^*) < \alpha \quad (2.32)$$

ise H_0 hipotezi reddedilir.

2.8. Tek Aşamalı Test

S. Chen ve H. Chen (1998), tek yönlü varyans analizinde normal dağılımlı ancak heterojen varyanslı yığınların ortalamaların eşitliği hipotezi için iki aşamalı test yerine tek aşamalı bir test önermişlerdir. Monte Carlo simülasyon çalışması sonucunda tek aşamalı yöntemin ikili aşamalı yöntemle göre hem I. tip hata oranı hem de testin gücü bakımından daha iyi sonuçlar verdiği görülmüştür. Ayrıca tek aşamalı yöntemin iki aşamalı yöntemle göre sağlamış olduğu ilave örnek gerektirmeme avantajı da göz önünde bulundurulduğunda tek aşamalı yöntem tercih edilmelidir. S. Chen ve H. Chen (1998) tarafından önerilen bu tek aşamalı test aşağıda kısaca tanıtılmıştır.

Normal dağılımlı ancak heterojen varyanslı i . yığından $i = 1, \dots, k$, n_i hacimli örnekler seçilsin. $n_i \geq 3$ olmak üzere, i . örnekteki j . gözlem X_{ij} , $j = 1, \dots, n_i$ ile gösterilsin.

$n_i - 1$ hacimli örnek için örnek ortalaması ve örnek varyansı, sırasıyla,

$$\bar{X}_i = \sum_{j=1}^{n_i-1} \frac{X_{ij}}{n_i - 1}$$

ve

(2.33)

$$S_i^2 = \sum_{j=1}^{n_i-1} \frac{(X_{ij} - \bar{X}_i)^2}{n_i - 2}$$

olsun. Ağırlıklı örnek ortalaması $\tilde{X}_i$ istatistiğini hesaplamak için öncelikle gözlemlerin ağırlıkları bulunmalıdır. Örnek varyansları olan $S_1^2, \dots, S_k^2$ istatistiklerinin en büyüğü (maksimum değeri) $S_{[b]}^2$ ile gösterilsin .

$$U_i = \frac{1}{n_i} + \frac{1}{n_i} \sqrt{\frac{1}{n_i-1} \left[\frac{S_{[b]}^2}{S_i^2} - 1 \right]}$$

ve (2.34)

$$V_i = \frac{1}{n_i} - \frac{1}{n_i} \sqrt{n_i - 1 \left[\frac{S_{[b]}^2}{S_i^2} - 1 \right]}$$

olmak üzere, gözlem ağırlıkları

$$W_{ij} = \begin{cases} U_i & 1 \leq j \leq n_i - 1 \\ V_i & j = n_i \end{cases} \quad (2.35)$$

biçiminde tanımlansın.

$$(n_i - 1)U_i + V_i = 1$$

ve (2.36)

$$(n_i - 1)U_i^2 + V_i^2 = \frac{S_{[b]}^2}{n_i S_i^2}$$

koşullarını sağlamak üzere, i . örnek için ağırlıklı örnek ortalaması

$$\tilde{X}_i = \sum_{j=1}^{n_i} W_{ij} X_{ij} \quad (2.37)$$

olarak tanımlanmıştır (S. Chen ve H. Chen 1998). Ayrıca genel ağırlıklı örnek ortalaması da

$$\tilde{X}_\cdot = \sum_{i=1}^k \frac{\tilde{X}_i}{k} \quad (2.38)$$

olarak hesaplanır. Eş. 2.37'deki $\tilde{X}_i$ istatistiği, ortalaması μ_i ve varyansı $\sum W_{ij}^2 \sigma_i^2$ olan normal dağılıma sahiptir. Buna göre

$$t_i = \frac{\tilde{X}_i - \mu_i}{\sqrt{S_i^2 \sum_{j=1}^{n_i} W_{ij}^2}} \quad (2.39)$$

dönüşüm ile t_i istatistiği türetilir. Eş. 2.36'da verilen koşullar sağlandığında ve örnek varyansları verilmişken,

$$t_i = \frac{\bar{X}_i - \mu_i}{S_{[b]}/\sqrt{n_i}}, i = 1, \dots, k \quad (2.40)$$

olarak tanımlanır ve t_i istatistiklerinin her biri $(n_i - 2)$ serbestlik derecesi ile bağımsız Student-t dağılımlarıdır (S. Chen ve H. Chen, 1998). Yığın ortalamalarının eşitliği hipotezinin testi için S. Chen ve H. Chen (1998) tarafından test istatistiği, bağımsız t_i istatistiklerinin karelerinin toplamı biçiminde,

$$\tilde{F}^1 = \sum_{i=1}^k \left(\frac{\bar{X}_i - \bar{X}}{S_{[b]}/\sqrt{n_i}} \right)^2 \quad (2.41)$$

olarak önerilmiştir. Eş. 2.41'deki test istatistiğinin dağılımını bulmak için bu istatistik

$$\tilde{F}^1 = \sum_{i=1}^k \left(\frac{\bar{X}_i - \mu_i}{S_{[b]}/\sqrt{n_i}} - \frac{\bar{X}_{..} - \mu}{S_{[b]}/\sqrt{n_i}} + \frac{\mu_i - \mu}{S_{[b]}/\sqrt{n_i}} \right)^2 \quad (2.42)$$

şeklinde de yazılabilir. Eş. 2.42'deki ikinci terimin

$$\frac{\bar{X}_{..} - \mu}{S_{[b]}/\sqrt{n_i}} = \frac{(1/k) \sum_{i=1}^k (\bar{X}_i - \mu_i)}{S_{[b]}/\sqrt{n_i}} = \frac{\sqrt{n_i}}{k} \sum_{i=1}^k \frac{t_i}{\sqrt{n_i}} \quad (2.43)$$

olduğu gösterilebilir (S. Chen ve H. Chen, 1998). H_0 hipotezinin doğruluğu altında, gerekli işlemler sonucunda Eş. 2.27'deki $\tilde{F}^1$ istatistiği için

$$\tilde{F}^1 = \sum_{i=1}^k \left(t_i - \frac{\sqrt{n_i}}{k} \sum_{i=1}^k \frac{t_i}{\sqrt{n_i}} \right)^2 \quad (2.44)$$

elde edilir (S. Chen ve H. Chen, 1998). Örnek hacimleri eşit iken yani $n_1 = n_2 = \dots = n_k = n$ olduğu durumda H_0 hipotezinin doğruluğu altında, Eş. 2.42'deki $\tilde{F}^1$ istatistiği, her biri $n - 2$ serbestlik dereceli bağımsız Student-t değişkenlerinden karesel formu olan

$$Q = \sum_{i=1}^k (t_i - \bar{t})^2 \quad (2.45)$$

istatistiğini verir (S. Chen ve H. Chen, 1998). Bu test için kritik değerler Q rassal değişkenin dağılımından çıkarılabilir. Q rassal değişkenin dağılımı $\tilde{F}^1$ istatistiğinin yokluk dağılımı ile aynıdır ve belli bir dağılıma yakınsamadığı için k ve $n - 1$ değerlerine göre m sayıda tekrar yapılarak dağılımın elde edilmesi gerekir. Buna göre Monte Carlo Simülasyon yöntemine göre oluşturulan dağılım $Q_{k,n-1}$ ve $\tilde{F}^1$ istatistiği için gözlenen değer de $\tilde{F}_h^1$ olmak üzere

$$P(Q_{k,n-1} > \tilde{F}_h^1) < \alpha \quad (2.46)$$

ise yokluk hipotezi reddedilir (S. Chen ve H. Chen, 1998).

2.9. Güvenilir (Fiducial) Yaklaşım Testi

Li, Wang ve Liang (2011), normal dağılımlı ancak heterojen varyanslı yığınların ortalamalarının eşitliği hipotezinin testi için Güvenilir (Fiducial) Yaklaşım testini geliştirmişlerdir. Bu test Dawid ve Stone (1982) tarafından geliştirilen güvenilir (fiducial) yaklaşıma dayanmaktadır. Özellikle biyomedikal ve klinik denemeler gibi oluşturulan kararın büyük önem arz ettiği durumlar için bu test önerilmektedir.

i . örnek hacmi n_i , $i = 1, \dots, k$ olmak üzere, i . örnek ortalaması $\bar{X}_i$ ve i . örnek varyansı S_i^2 ile gösterilsin ve bu istatistikler

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$$

ve (2.47)

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

olarak tanımlansın. Bu yaklaşımda test istatistiği, yığın varyansları biliniyorken,

$$T(\bar{X}; \sigma^2) = \sum_{i=1}^k \frac{n_i}{\sigma_i^2} \left\{ (\bar{X}_i - \mu_i) - \frac{\sum_{i=1}^k \frac{n_i(\bar{X}_i - \mu_i)}{\sigma_i^2}}{\sum_{i=1}^k \frac{n_i}{\sigma_i^2}} \right\}^2 \sim \chi_{k-1}^2 \quad (2.48)$$

olarak önerilmiştir. Genelde yığın varyansları bilinmez. Eğer yığın varyansları bilinmiyorsa, σ_i^2 yerine S_i^2 alınarak

$$T(\bar{X}; S^2) = \sum_{i=1}^k \frac{n_i}{S_i^2} \left\{ (\bar{X}_i - \mu_i) - \frac{\sum_{i=1}^k \frac{n_i(\bar{X}_i - \mu_i)}{S_i^2}}{\sum_{i=1}^k \frac{n_i}{S_i^2}} \right\}^2 \quad (2.49)$$

elde edilir. H_0 hipotezinin doğruluğu altında bu istatistik için

$$T(\bar{X}; S^2) = \sum_{i=1}^k \frac{n_i}{S_i^2} \left\{ \bar{X}_i - \frac{\sum_{i=1}^k \frac{n_i \bar{X}_i}{S_i^2}}{\sum_{i=1}^k \frac{n_i}{S_i^2}} \right\}^2 \quad (2.50)$$

elde edilir. Eş. 2.49'da $\bar{X}_i$ ve S_i^2 yerlerine $\bar{x}_i$ ve s_i^2 yazıldığında test istatistiği

$$\sum_{i=1}^k \frac{n_i}{s_i^2} \left\{ (\bar{x}_i - \mu_i) - \frac{\sum_{i=1}^k \frac{n_i(\bar{x}_i - \mu_i)}{s_i^2}}{\sum_{i=1}^k \frac{n_i}{s_i^2}} \right\}^2 \quad (2.51)$$

olarak elde edilir (Li ve diğerleri, 2011).

Bu test istatistiği ile p-değeri hesaplamak amacıyla birbirinden bağımsız $U_{1i} \sim N(0,1)$ ve $U_{2i} \sim \chi_{n_i-1}^2$ değişkenleri tanımlansın. Bu istatistikler ile

$$\bar{X}_i = \mu_i + \frac{\sigma_i}{\sqrt{n_i}} U_{1i}$$

ve (2.52)

$$(n_i - 1)S_i^2 = \sigma_i^2 U_{2i}$$

olduğu gösterilebilir. Ayrıca

$$\bar{X}_i \sim N(\mu_i, \frac{\sigma_i}{\sqrt{n_i}})$$

ve

(2.53)

$$(n_i - 1)S_i^2 \sim \sigma_i^2 \chi_{n_i-1}^2$$

eşitlikleri de gösterilebilir (Li ve diğerleri, 2011). $(\bar{x}_i, s_i^2)$ ve (u_{1i}, u_{2i}) gözlenen değerler olmak üzere

$$\bar{x}_i = \mu_i + \frac{\sigma_i}{\sqrt{n_i}} u_{1i}$$

ve

(2.54)

$$(n_i - 1)s_i^2 = \sigma_i^2 u_{2i}$$

biçimde de yazılabilir. Eş. 2.54 için

$$\mu_i = \bar{x}_i - \frac{u_{1i}}{\sqrt{\frac{u_{2i}}{n_i-1}}} \sqrt{\frac{s_i^2}{n_i}}$$

ve

(2.55)

$$\sigma_i^2 = \frac{(n_i-1)s_i^2}{u_{2i}}$$

olarak bulunur. Buna göre $(\bar{x}_i, s_i^2)$ verildiğinde bu istatistiğin güvenilir (fiducial) dağılımı ile

$$T_{\mu_i} = \bar{x}_i - t_i \sqrt{S_i^2/n_i}, i = 1, \dots, k \quad (2.56)$$

istatistiğinin dağılımı aynıdır (Li ve diğerleri, 2011). $(n_i - 1)$ serbestlik dereceli student t dağılımı $t(n_i - 1)$ olmak üzere bu istatistik $t_i \sim t(n_i - 1)$, $i = 1, \dots, k$ biçimde de ifade

edilir. Ayrıca bu istatistik yokluk hipotezini test etmek için p-değerlerini elde etmek amacıyla kullanılır. Bu amaçla $t_i \sim t(n_i - 1)$, $i = 1, \dots, k$ olmak üzere

$$T_F(t, s^2) = \sum_{i=1}^k t_i^2 - \frac{\left(\sum_{i=1}^k \frac{\sqrt{n_i}}{s_i} t_i \right)^2}{\sum_{i=1}^k \frac{n_i}{s_i^2}} \quad (2.57)$$

ile Eş. 2.48'deki istatistiğinin güvenilir (fiducial) dağılımı türetilir. H_0 hipotezinin doğruluğu altında p-değeri

$$p = P [T_F \geq T(\bar{X}, S^2)] \quad (2.58)$$

şeklinde elde edilir. Buna göre H_0 hipotezinin testi için Eş. 2.58 ile p-değeri bulunur ve

$$p\text{-değeri} < \alpha \quad (2.59)$$

olursa nominal α ile H_0 reddedilir. Eş. 2.59'da verilen p-değerinin elde edilebilmesi için kullanılan algoritma aşağıda verilmiştir.

Algoritma

$\bar{x}$, s^2 ve Eş. 2.50'deki test istatistiği hesaplanır.

Eş. 2.50'de verilen $T(\bar{X}; S^2)$, R_0 olarak adlandırılır.

Döngü $k = 1, \dots, M$ olmak üzere 1'den M 'ye kadar aşağıdaki hesaplamaları içerecek şekilde çalıştırılır.

Döngüde $t_i \sim t(n_i - 1)$, $i = 1, \dots, k$ üretilir.

$$R_j = \sum_{i=1}^k t_i^2 - \left(\sum_{i=1}^k \frac{\sqrt{n_i}}{s_i} t_i \right)^2 / \sum_{i=1}^k \frac{n_i}{s_i^2} \quad (2.60)$$

hesaplanır.

Bu hesaplama sonucunda $R_j > R_0$ olduğu durum için $W_j = 1$ ve aksi durumda $W_j = 0$ ataması yapılır. Döngü tamamlandıktan sonra bu simülasyon ile p-değerinin Monte Carlo tahmini

$$\text{p-değeri} = \frac{1}{m} \sum_{j=1}^m W_j \quad (2.61)$$

olarak hesaplanır.

Sonuç olarak p-değeri $< \alpha$ olduğunda ortalamaların eşitliği hipotezi reddedilir.

2.10. Parametrik Bootstrap Testi

Parametrik Bootstrap (PB) testi, Krishnamoorthy, Lu ve Mathew (2007) tarafından önerilmiştir. Eş. 2.8’de verilen $T_N(\bar{X}_1, \dots, \bar{X}_k; S_1^2, \dots, S_k^2)$ istatistiğine dayalı olarak PB testi aşağıdaki gibi tanımlanır.

Bu test için i . yığından n_i hacimli örnek $X_{i1}, X_{i2}, \dots, X_{in_i}$, $i = 1, 2, \dots, k$ olmak üzere bu örneklerin herbirinden B sayıda tekrar ile yerine koyarak yöntemi ile n_i hacimli Bootstrap örnekleri seçilsin ve bunlar $X_{i1}^l, X_{i2}^l, \dots, X_{in_i}^l$, $l = 1, 2, \dots, B$ olarak gösterilsin. Bu Bootstrap örneklerinin ortalamaları $\bar{X}_{Bi}$ ve varyansları S_{Bi}^2 ile gösterilsin. İşlem kolaylığı olması için ortalama parametresinin değişmezlik özelliğinden yararlanarak ortalama değeri “0” olarak alınmıştır. Ayrıca S_{Bi}^2 istatistiğinin konum parametresinin değişmezlik özelliği de vardır. Buna göre Bootstrap örnek ortalamalarının ve varyanslarının asimptotik dağılımları

$$\bar{X}_{Bi} \sim N(0, S_i^2/n_i)$$

$$\text{ve} \quad (2.62)$$

$$S_{Bi}^2 \sim S_i^2 \chi_{n_i-1}^2 / (n_i - 1), i = 1, \dots, k$$

olur (Krishnamoorthy ve diğerleri, 2007). Eş. 2.8’de verilen T_N istatistiği $\bar{X}_{Bi}$ ve S_{Bi}^2 değişkenlerine dayalı olarak

$$T_N(\bar{X}_1, \dots, \bar{X}_k; S_1^2, \dots, S_k^2) = \sum_{i=1}^k \frac{n_i}{S_{Bi}^2} \bar{X}_{Bi}^2 - \frac{\left[\sum_{i=1}^k n_i \bar{X}_{Bi} / S_{Bi}^2 \right]^2}{\sum_{i=1}^k n_i / S_{Bi}^2} \quad (2.63)$$

biçiminde ifade edilir (Krishnamoorthy ve diğerleri, 2007). Bu istatistik Parametrik Bootstrap (PB) değişkeni olarak adlandırılır. Bu istatistikte standart normal rassal değişken Z_i olmak üzere, $\bar{X}_{Bi} \sim Z_i(S_i/\sqrt{n_i})$ bilgisi ile yukardaki PB değişkeni için

$$T_{NB}(Z_i, \chi_{n_i-1}^2; S_i^2) = \sum_{i=1}^k \frac{Z_i^2(n_i-1)}{\chi_{n_i-1}^2} - \frac{\left[\sum_{i=1}^k (\sqrt{n_i} Z_i(n_i-1) / \{S_i \chi_{n_i-1}^2\}) \right]^2}{\sum_{i=1}^k (n_i(n_i-1) / \{S_i^2 \chi_{n_i-1}^2\})} \quad (2.64)$$

istatistiği tanımlanmıştır (Krishnamoorthy ve diğerleri, 2007). Belirli bir α değeri ve $(S_1^2, \dots, S_k^2)$ istatistiğinin gözlenen $(s_1^2, \dots, s_k^2)$ değerleri için ,

$$P(T_{NB}(Z_i, \chi_{n_i-1}^2; s_i^2) > T_{N0}) < \alpha \quad (2.65)$$

ise PB testi ile H_0 hipotezi reddedilir.

T_{N0} değeri, Eş. 2.63'deki T_N istatistiğinin gözlenen bir değeridir. Sabit bir $(s_1, \dots, s_k)$ için, yukarıdaki olasılık herhangi bir bilinmeyen parametreye bağlı değildir. Bu nedenle, bu olasılık aşağıdaki Algoritma ile verilen Monte Carlo benzetimi kullanılarak tahmin edilir.

Algoritma

$(n_1, n_2, \dots, n_k)$, $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)$ ve $(s_1^2, s_2^2, \dots, s_k^2)$ olarak verilsin.

Eş. 2.63'de verilen T_N, T_{N0} olarak adlandırılır.

Döngüye başlanır.

j , 1'den m 'ye kadar döngü çalıştırılır.

$Z_i \sim N(0,1)$ ve $\chi_{n_i-1}^2$, $i = 1, \dots, k$ üretilir.

Eş. 2.64'deki $T_{NB}(Z_i, \chi_{n_i-1}^2; s_i^2)$ formülde verildiği şekilde hesaplanır.

Eğer $T_{NB}(Z_i, \chi_{n_i-1}^2; s_i^2) > T_{N0}$ olursa $Q_j = 1$ atanır.

$j=m$ olduğunda döngü sonlandırılır. Sonra

$$(1/m) \sum_{j=1}^m Q_j \quad (2.66)$$

hesaplanır.

Bu değer p-değerinin Monte Carlo tahmin değeridir.

Eğer p-değeri $< \alpha$ olduğunda ortalamaların eşitliği hipotezi reddedilir.



3. GERÇEK VERİ UYGULAMASI

Bu bölümde ikinci bölümde tanıtılan testlerin gerçek veri setleri üzerinde uygulamaları verilmiştir. Bu veriden birincisi Anderson (1935) tarafından derlenen iris çiçeğinin üç farklı türü olan setosa, versicolor ve virginica ile ilgili taç yaprak uzunluk değerleri üzerinedir. Bu üç türe ait taç yaprak uzunlukları Çizelge 3.1’de verilmiştir.

Çizelge 3.1. Çiçek türleri için taç yaprak uzunlukları

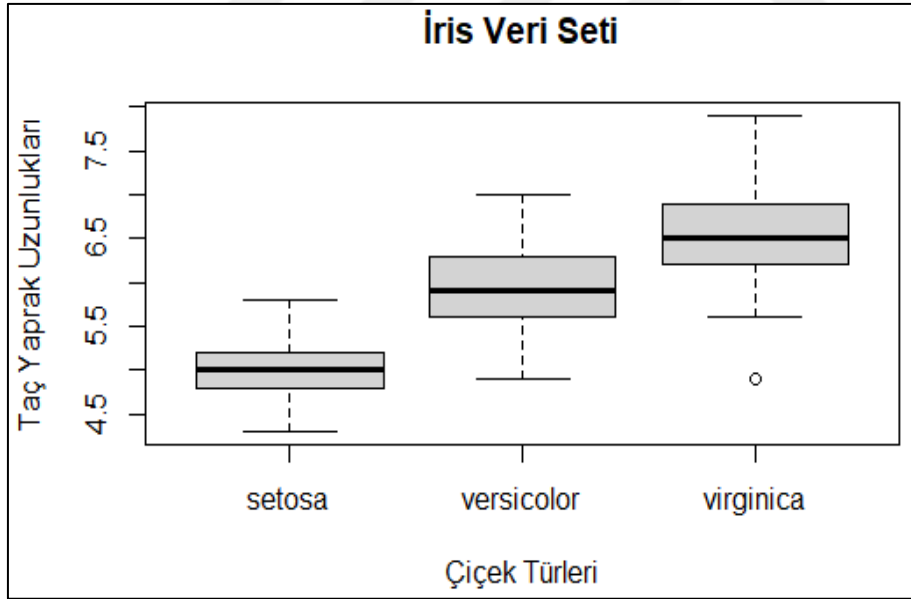
Çiçek Türleri					
Setosa		Versicolor		Virginica	
5.1	5.0	7.0	6.6	6.3	7.2
4.9	5.0	6.4	6.8	5.8	6.2
4.7	5.2	6.9	6.7	7.1	6.1
4.6	5.2	5.5	6.0	6.3	6.4
5.0	4.7	6.5	5.7	6.5	7.2
5.4	4.8	5.7	5.5	7.6	7.4
4.6	5.4	6.3	5.5	4.9	7.9
5.0	5.2	4.9	5.8	7.3	6.4
4.4	5.5	6.6	6.0	6.7	6.3
4.9	4.9	5.2	5.4	7.2	6.1
5.4	5.0	5.0	6.0	6.5	7.7
4.8	5.5	5.9	6.7	6.4	6.3
4.8	4.9	6.0	6.3	6.8	6.4
4.3	4.4	6.1	5.6	5.7	6.0
5.8	5.1	5.6	5.5	5.8	6.9
5.7	5.0	6.7	5.5	6.4	6.7
5.4	4.5	5.6	6.1	6.5	6.9
5.1	4.4	5.8	5.8	7.7	5.8
5.7	5.0	6.2	5.0	7.7	6.8
5.1	5.1	5.6	5.6	6.0	6.7
5.4	4.8	5.9	5.7	6.9	6.7
5.1	5.1	6.1	5.7	5.6	6.3
4.6	4.6	6.3	6.2	7.7	6.5
5.1	5.3	6.1	5.1	6.3	6.2
4.8	5.0	6.4	5.7	6.7	5.9

Çizelge 3.1’deki gerçek veri seti için çiçek türlerinin taç yaprak uzunluklarına ilişkin tanımlayıcı istatistikler Çizelge 3.2’de ve kutu grafiği de Şekil 3.1’de verilmiştir.

Çizelge 3.2. Çiçek türleri için taç yaprak uzunluklarına ilişkin tanımlayıcı istatistikler

	Çiçek Türleri		
	Setosa	Versicolor	Virginica
Örnek Hacmi	50	50	50
Ortalama	5.01	5.94	6.59
Standart Sapma	0.35	0.52	0.64
Medyan	5	5.9	6.5
En Küçük	4.3	4.9	4.9
En Büyük	5.8	7	7.9
Birinci Çeyreklik	4.8	5.6	6.23
Üçüncü Çeyreklik	5.2	6.3	6.9
Çarpıklık	0.1165	0.1022	0.1144
Basıklık	2.6542	2.4012	2.9121

Çizelge 3.2'deki tanımlayıcı istatistikler incelendiğinde, çiçek türlerinin çarpıklık ve basıklık değerlerinin normal dağılımının çarpıklık ve basıklık değerlerine yakın olduğu söylenebilir. Çarpıklık değerlerinin sıfıra yakın olmasından dolayı çarpıklığın olmadığı söylenebilir.



Şekil 3.1. Çiçek türleri için taç yaprak uzunluklarına ilişkin kutu grafiği

Şekil 3.1'deki kutu grafiğine göre çiçek türlerinin varyansların homojen olmadığı kanısı oluşmaktadır.

Bu örneklerin geldikleri yığınların dağılımlarının normal dağılıma uygun olup olmadığını belirlemek için kullanılan Shapiro-Wilk testi sonuçları Çizelge 3.3'de verilmiştir.

Çizelge 3.3. Normal dağılıma uygunluk için Shapiro-Wilk uyum iyiliği testi sonuçları

Çiçek Türü	Shapiro-Wilk Test İstatistiği Değeri	p-değeri
setosa	0.9777	0.4595
versicolor	0.9778	0.4647
virginica	0.9712	0.2583

Bu çizelgedeki p-değerlerinin hepsi de $\alpha = 0,05$ değerinden daha büyük olduğundan %95 güven düzeyinde bu örneklerin geldikleri yığınların dağılımlarının normal olduğu söylenir. Bu veri ile varyansların homojenliği varsayımının sağlanıp sağlanmadığını belirlemek amacıyla kullanılan Levene testine ilişkin sonuçlar Çizelge 3.4’de verilmiştir.

Çizelge 3.4. Levene testi sonuçları

Levene Test İstatistiği Değeri	Sd1	Sd2	p-değeri
7.381092	2	147	0.0008

Varyansların homojenliği testi için kullanılan Levene testi sonucunda p-değeri 0,0008 olarak hesaplanmıştır. Bu p-değeri $\alpha = 0,05$ değerinden küçük olduğu için yığın varyanslarının homojen olmadığı %95 güven düzeyi ile söylenir. Bu incelemeler sonucunda normallik varsayımının sağlandığı ancak varyansların homojenliği varsayımının sağlanmadığı anlaşılmaktadır.

Çizelge 3.1’deki veri ile ikinci bölümdeki testler için gerekli işlemler yapılmış ve James testi hariç diğer testler için elde edilen sonuçlar Çizelge 3.5’de verilmiştir.

Çizelge 3.5. Test sonuçları

Test	Test İstatistiği Değeri	p-değeri
Welch Testi	138.9083	0.000
Brown-Forsythe Testi	119.2645	0.000
Parametrik Bootstrap Testi	279.8251	0.000
Cochran Testi	279.8251	0.000
Alexander ve Govern Testi	146.3573	0.000
Düzenlenmiş Brown-Forsythe Testi	119.2645	0.000
Düzenlenmiş Welch Testi	133.2386	0.000
Tek Aşamalı Test	153.7579	0.000
Güvenilir Yaklaşım Testi	279.8251	0.000

Çizelge 3.5’deki sonuçlara göre Welch, Brown-Forsythe, Parametrik Bootstrap, Cochran, Alexander ve Govern, Düzenlenmiş Brown-Forsythe, Düzenlenmiş Welch, Tek Aşamalı ve

Güvenilir Yaklaşım testleri için p-değerlerinin 0,000 olduğu görülmektedir. Buna göre $\alpha = 0,05$ iken bu testler için p-değeri $< \alpha$ olduğundan ortalamaların eşitliği hipotezi reddedilir. James testi karar kuralında p-değeri değil kritik değer kullanıldığı için bu teste ilişkin sonuçlar Çizelge 3.5’de verilmemiştir. Bu teste ilişkin istatistik değeri 279,8251 ve $\alpha = 0,05$ için kritik değerde 6,233185 olarak bulunur. Buna göre James testi ile %95 güven düzeyinde ortalamaların eşitliği hipotezi reddedilir.

İkinci veri Weerahandi (1995b) tarafından derlenen homojen toprak koşullarında 22 parselde ekilmiş dört farklı hibrit mısır türünün verimi üzerinedir. Bu dört türe ait verim değerleri Çizelge 3.6’de verilmiştir.

Çizelge 3.6. Tohum türleri için mısır verim değerleri

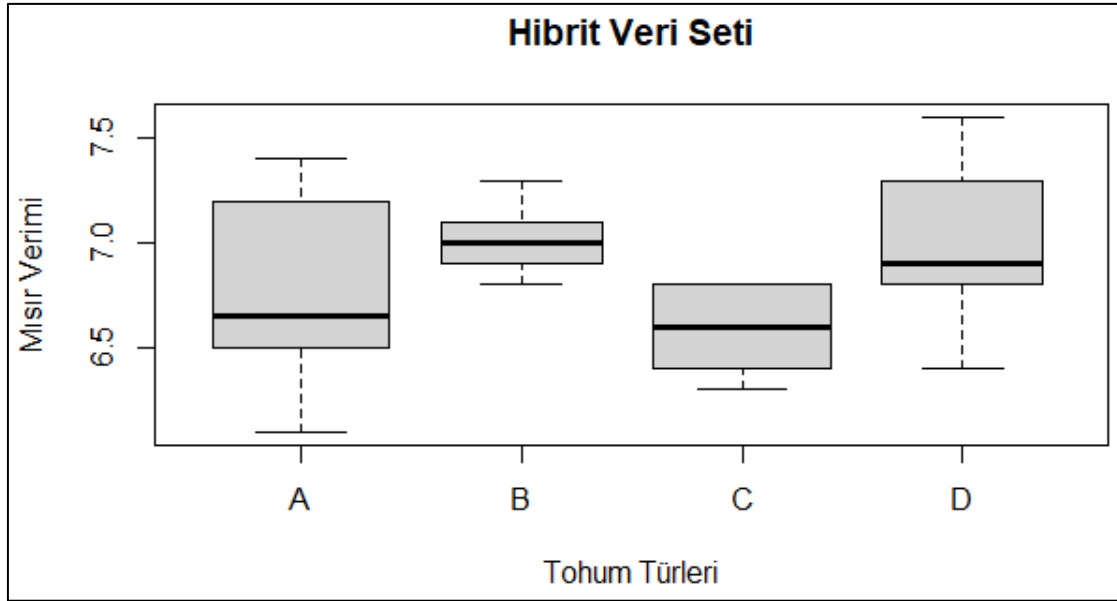
Tohum Türleri			
Hibrit A	Hibrit B	Hibrit C	Hibrit D
7.4	7.1	6.8	6.4
6.6	7.3	6.3	6.9
6.7	6.8	6.4	7.6
6.1	6.9	6.7	6.8
6.5	7.0	6.5	7.3
7.2		6.8	

Çizelge 3.6’deki gerçek veri seti için mısır verimine ilişkin tanımlayıcı istatistikler Çizelge 3.7’de ve türlere göre kutu grafiği de Şekil 3.2’de verilmiştir.

Çizelge 3.7. Tohum türleri için mısır verimine ilişkin tanımlayıcı istatistikler

	Tohum Türleri			
	A	B	C	D
Örnek Hacmi	6	5	6	5
Ortalama	6.75	7.02	6.58	7
Standart Sapma	0.48	0.19	0.21	0.46
Medyan	6.65	7	6.6	6.9
En Küçük	6.1	6.8	6.3	6.4
En Büyük	7.4	7.3	6.8	7.6
Birinci Çeyreklik	6.53	6.9	6.43	6.8
Üçüncü Çeyreklik	7.08	7.1	6.78	7.3
Çarpıklık	0.1500	0.4000	-0.1700	0.0500
Basıklık	1.8700	1.9900	1.4100	1.8200

Çizelge 3.7’deki tanımlayıcı istatistikler incelendiğinde, tohum türlerinin çarpıklık değerlerinin normal dağılımının çarpıklık değerine oldukça yakın olduğu söylenebilir.



Şekil 3.2.Tohum türleri için mısır verimine ilişkin kutu grafiği

Şekil 3.2'deki kutu grafiğine göre tohum türlerine ilişkin herhangi bir yorum yapılamamaktadır.

Her bir örneğin geldiği yığının dağılımının normal dağılıma uygun olup olmadığı belirlemek için kullanılan Shapiro-Wilk test sonuçları Çizelge 3.8'de verilmiştir.

Çizelge 3.8. Normal dağılıma uygunluk için Shapiro-Wilk uyum iyiliği testi sonuçları

Tür	Shapiro-Wilk Test İstatistiği Değeri	p-değeri
A	0.9551273	0.7815109
B	0.9787162	0.9276364
C	0.8924716	0.3313410
D	0.9804393	0.9369553

Çizelge 3.8'deki p-değerlerinin hepsi de $\alpha = 0,05$ değerinden daha büyük olduğundan %95 güven düzeyinde bu örneklerin geldikleri yığınların dağılımlarının normal olduğu söylenir. Varyansların homojenliği varsayımının sağlanıp sağlanmadığını belirlemek amacıyla kullanılan Levene testine ilişkin sonuçlar Çizelge 3.9'de verilmiştir.

Çizelge 3.9. Levene testi sonuçları

Levene Test İstatistiği Değeri	sd1	sd2	p-değeri
2.1628	3	18	0.1279

Varyansların homojenliği için Levene testi sonucunda p-değeri 0,1279 olarak hesaplanmıştır. Levene testi için p-değeri $\alpha = 0,05$ değerinden büyük olduğu için yığın varyanslarının homojen olduğu %95 güven düzeyinde söylenir. Bu testler sonucunda normallik varsayımının ve varyansların homojenliği varsayımının sağlandığı görülmektedir.

Bu veri ile yığın ortalamalarının eşitliği hipotezini test etmek için ikinci bölümde tanıtılan testler için istatistiklerin değerleri ve p-değerleri bulunmuş ve Çizelge 3.10'da verilmiştir. James testi için karar kuralı olarak p-değeri kullanılmamaktadır. Bundan dolayı çizelgede James testi yer almamaktadır.

Çizelge 3.10. Test sonuçları

Test	Test İstatistiği Değeri	p-değeri
Welch Testi	3.980426	0.0446
Brown-Forsythe Testi	1.850501	0.1914
Parametrik Bootstrap Testi	13.63805	0.0506
Cochran Testi	13.63805	0.0034
Alexander ve Govern Testi	7.911063	0.0479
Düzenlenmiş Brown-Forsythe Testi	1.850501	0.1971
Düzenlenmiş Welch Testi	2.179458	0.1579
Tek Aşamalı Test	5.118701	0.4117
Güvenilir Yaklaşım Testi	13.63805	0.0720

Çizelge 3.10'da verilen sonuçlara göre $\alpha = 0,05$ iken Cochran, Welch ve Alexander ve Govern testlerine göre %95 güven düzeyinde ortalamaların eşitliği hipotezi reddedilir iken diğer testlere göre ortalamaların eşitliği hipotezi reddedilemez. James testi karar kuralında p-değeri değil kritik değer kullanıldığı için bu teste ilişkin sonuçlar çizelgede verilmemiştir. Bu teste ilişkin istatistik değer 13,63805 ve $\alpha = 0,05$ için kritik değerde 12,83385 olduğundan dolayı %95 güven düzeyinde ortalamaların eşitliği hipotezi reddedilir.

4. SİMÜLASYON ÇALIŞMASI

İkinci bölümde tanıtılan testleri I. tip hata oranı ve güç bakımından karşılaştırmak amacıyla bu bölümde bir simülasyon çalışması yapılmıştır. Bu simülasyon çalışmasında yığın sayıları, örnek hacimleri, varyans değerleri ve ortalama değerleri gibi tasarım parametreleri ile farklı simülasyon senaryoları kullanılmıştır.

I. tip hata oranı karşılaştırmaları için simülasyon senaryolarında ortalamaların eşit ve 0 olduğu durumda $k = 3, 4, 5, 6$ ve $n = 5, 7, 9, 10, 11, 15, 20, 25$ ile çeşitli σ^2 parametrelili normal dağılımlardan 10 000 tekrar ile üretilen veriler kullanılmıştır. Diğer yandan güç bakımından karşılaştırmalar içinde $k = 3, 4, 5$ ve $n = 5, 7, 9, 10, 11, 15$ ile çeşitli μ ve σ^2 parametrelili normal dağılımlardan 10 000 tekrar ile üretilen verilerle simülasyon yapılmıştır. Sonra her bir simülasyon senaryosunda tüm testler için H_0 hipotezini ret oranları hesaplanmıştır. Bu ret oranları yokluk hipotezini ret etme sayısının 10 000 ile bölümüyle elde edilmiştir. Ayrıca Parametrik Bootstrap, Tek Aşamalı ve Güvenilir (Fiducial) Yaklaşım testlerinin test istatistiğinin dağılımı elde etmek amacıyla kullanılan algoritmalar için tekrar değeri 10 000 alınmıştır.

Simülasyon çalışması tasarlanırken, varyansların çeşitli heterojenlik seviyeleri göz önünde bulundurulmuştur. Örnek hacimleri eşit olmadığı durumlarda heterojenliğin farklı seviyeleri ile ilişkileri de incelenmiştir. Kullanılan testler aşağıdaki kısaltmalar ile yer almaktadır.

J	:James Testi
W	:Welch Testi
BF	:Brown-Forsythe Testi
PB	:Parametrik Bootstrap Testi
CF	:Cochran Testi
AG	:Alexander ve Govern Testi
MBF	:Düzenlenmiş (modified) Brown-Forsythe Testi
MW	:Düzenlenmiş (modified) Welch Testi
OS	:Tek Aşamalı Test
FA	:Güvenilir (Fiducial) Yaklaşım Testi

Yığın ortalamaları eşit iken bu ret oranları testlere ilişkin deneysel I. tip hata oranlarıdır. Bu deneysel I. tip hata oranları Çizelge 4.1-4.12’de verilmiştir. Daha sonra varyansların eşit ve farklı olduğu durumlarda normal dağılımlı yığınların ortalamaları farklı olarak alınmış ve bu testlere ilişkin ret oranları hesaplanmıştır. Bu ret oranlarının deneysel güç değerleri olduğu açıktır. Testlere ilişkin deneysel güç değerleri Çizelge 4.13-4.24’de verilmiştir.

Çizelge 4.1. $k=3, n_1=n_2=n_3=5$, nominal $\alpha=0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

$\sigma_1^2, \sigma_2^2, \sigma_3^2$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
4, 4, 4	0.048	0.044	0.039	0.044	0.123	0.043	0.036	0.007	0.012	0.024
2, 2, 2	0.054	0.049	0.045	0.048	0.127	0.049	0.042	0.009	0.014	0.026
1, 1, 1	0.049	0.044	0.039	0.043	0.121	0.045	0.038	0.007	0.013	0.022
2, 2, 1	0.050	0.044	0.044	0.043	0.125	0.043	0.041	0.007	0.014	0.023
2, 1.5, 1.5	0.055	0.051	0.046	0.050	0.134	0.050	0.043	0.009	0.015	0.028
2, 1.5, 1	0.051	0.046	0.043	0.046	0.128	0.045	0.040	0.007	0.014	0.022
2, 1, 1	0.050	0.045	0.044	0.044	0.120	0.043	0.040	0.008	0.014	0.024
2, 1, 0.2	0.059	0.057	0.054	0.055	0.147	0.055	0.049	0.011	0.015	0.035
2, 0.5, 0.5	0.050	0.047	0.052	0.046	0.125	0.045	0.048	0.009	0.017	0.026
2, 0.5, 0.2	0.055	0.051	0.060	0.049	0.139	0.047	0.055	0.012	0.017	0.030
2, 0.2, 0.2	0.052	0.049	0.064	0.047	0.137	0.047	0.057	0.012	0.018	0.030
2, 0.2, 0.02	0.054	0.054	0.061	0.050	0.151	0.051	0.054	0.013	0.020	0.040

Çizelge 4.1’deki simülasyon sonuçları incelendiğinde J, W, BF, PB, AG ve MBF testlerinin deneysel I. tip hata oranlarının nominal I. tip hata oranına yakın oldukları söylenebilir. Sonuçlara göre, bu testler içinde I. tip hata oranı bakımından J testinin en iyi test ve W ile PB testlerinin de sonraki iyi testler olduğu görülmektedir. Diğer yandan CF, MW, OS ve FA testlerinin deneysel I. tip hata oranları nominal I. tip hata oranına yakın değildir. Bu testler içerisinde I. tip hata oranı bakımından en kötü sonuçları CF testinin sonraki kötü sonuçları da MW testinin verdiği görülmektedir.

Çizelge 4.2. $k=3, n_1=n_2=n_3=10$, nominal $\alpha=0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

$\sigma_1^2, \sigma_2^2, \sigma_3^2$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
2,2,2	0.051	0.050	0.050	0.050	0.086	0.049	0.047	0.025	0.025	0.039
2,2,1	0.050	0.049	0.049	0.049	0.083	0.048	0.046	0.026	0.025	0.039
2,1.5,1.5	0.050	0.048	0.047	0.049	0.082	0.047	0.044	0.026	0.026	0.037
2,1.5,1	0.049	0.047	0.046	0.047	0.080	0.045	0.044	0.023	0.025	0.036
2,1,1	0.053	0.051	0.053	0.051	0.089	0.049	0.049	0.025	0.025	0.040
2,1,0.2	0.052	0.052	0.054	0.052	0.089	0.051	0.047	0.025	0.025	0.043

Çizelge 4.2. (devam) $k = 3$, $n_1 = n_2 = n_3 = 10$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

$\sigma_1^2, \sigma_2^2, \sigma_3^2$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
2,0.5,0.5	0.053	0.052	0.058	0.052	0.086	0.050	0.051	0.025	0.025	0.040
2,0.5,0.2	0.046	0.046	0.059	0.045	0.087	0.045	0.050	0.023	0.025	0.038
2,0.2,0.2	0.050	0.049	0.066	0.049	0.083	0.047	0.052	0.025	0.024	0.041
2,0.2,0.02	0.048	0.049	0.068	0.048	0.097	0.048	0.053	0.026	0.027	0.044

Çizelge 4.2'deki simülasyon sonuçları incelendiğinde J, W, BF, PB, AG ve MBF testlerinin nominal I. tip hata oranına yakın deneysel I. tip hata oranları verdiği görülmektedir. Bu testler içinde I. tip hata oranı bakımından J testinin en iyi test ve W testinin sonraki en iyi test olduğu sonuçların incelenmesiyle görülür. Ayrıca varyanslar homojen değil iken BF testinin deneysel I. hata oranları nominal I. tip hata oranından uzaklaşmaktadır. Diğer yandan CF, MW, OS ve FA testlerinin deneysel I. tip hata oranları nominal I. tip hata oranına yakın değildir. Bu testler içerisinde I. tip hata oranı bakımından en kötü sonuçları CF testinin sonraki kötü sonuçları da MW ve OS testlerinin verdiği anlaşılmaktadır.

Çizelge 4.3. $k = 3$, $n_1 = n_2 = n_3 = 20$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

$\sigma_1^2, \sigma_2^2, \sigma_3^2$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
4, 4, 4	0.048	0.048	0.049	0.047	0.063	0.047	0.048	0.033	0.036	0.042
2, 2, 2	0.049	0.048	0.049	0.049	0.063	0.047	0.048	0.036	0.035	0.043
1, 1, 1	0.052	0.052	0.050	0.051	0.064	0.051	0.048	0.036	0.037	0.045
2, 1.5, 1	0.048	0.047	0.048	0.047	0.063	0.046	0.045	0.035	0.037	0.043
2, 1, 0.5	0.053	0.052	0.056	0.053	0.068	0.052	0.050	0.040	0.036	0.048
1, 0.8, 0.6	0.051	0.051	0.051	0.051	0.066	0.050	0.049	0.036	0.038	0.046
2, 0.08, 0.06	0.053	0.053	0.073	0.053	0.071	0.052	0.051	0.038	0.039	0.048
0.8, 0.6, 0.4	0.054	0.053	0.053	0.054	0.070	0.052	0.050	0.037	0.039	0.046

Çizelge 4.3'deki deneysel I. tip hata oranları incelendiğinde J, W, BF, PB, AG, MBF ve FA testlerinin deneysel I. tip hata oranlarının nominal I. tip hata oranına yakın oldukları anlaşılmaktadır. Bu çizelgede I. tip hata oranı bakımından testlerin çoğunun iyi sonuçlar verdiği görülmektedir. Sadece BF ve FA testleri diğer testlere göre I. tip hata oranı bakımından biraz kötü sonuçlar vermiştir. Diğer yandan CF, MW ve OS testlerinin deneysel I. tip hata oranları nominal I. tip hata oranına yakın değildir. Bu testlerin deneysel I. tip hata oranı bakımından birbirlerine yakın sonuçlar verdiği görülmektedir.

Çizelge 4.4. $k = 3$, $n_1 = 7$, $n_2 = 9$, $n_3 = 11$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

$\sigma_1^2, \sigma_2^2, \sigma_3^2$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
4,4,4	0.050	0.049	0.047	0.048	0.085	0.049	0.044	0.022	0.026	0.037
2,2,2	0.050	0.049	0.047	0.048	0.087	0.047	0.045	0.021	0.025	0.035
1,1,1	0.051	0.050	0.047	0.050	0.089	0.049	0.045	0.021	0.029	0.038
2,1,5,1	0.053	0.052	0.051	0.051	0.095	0.053	0.048	0.022	0.024	0.039
1,1,5,2	0.047	0.045	0.044	0.046	0.081	0.044	0.041	0.018	0.031	0.033
2,1,0,5	0.051	0.051	0.051	0.051	0.094	0.051	0.047	0.023	0.023	0.040
0,5,1,2	0.051	0.049	0.053	0.049	0.084	0.047	0.046	0.021	0.032	0.034
1,0,8,0,6	0.050	0.049	0.048	0.049	0.088	0.047	0.045	0.020	0.023	0.036
0,6,0,8,1	0.053	0.052	0.049	0.052	0.090	0.051	0.044	0.019	0.029	0.036
2,0,0,8,0,0,6	0.054	0.054	0.071	0.052	0.105	0.052	0.058	0.025	0.025	0.044
0,0,6,0,0,8,2	0.048	0.048	0.067	0.048	0.087	0.045	0.051	0.022	0.033	0.035
0,8,0,6,0,4	0.052	0.051	0.050	0.052	0.094	0.051	0.048	0.023	0.025	0.039
0,4,0,6,0,8	0.052	0.051	0.051	0.051	0.086	0.049	0.046	0.022	0.030	0.037

Çizelge 4.4'deki deneysel I. tip hata oranlarına göre J, W, BF, PB, AG ve MBF testlerinin deneysel I. tip hata oranlarının nominal I. tip hata oranına yakın oldukları gözlenmiştir. Bu testler içerisinde I. tip hata oranı bakımından J, W ve PB testlerinin en iyi testler ve AG testinin sonraki iyi test olduğu sonuçların incelenmesiyle görülür. Diğer yandan CF, MW, OS ve FA testlerinin deneysel I. tip hata oranlarının nominal I. tip hata oranına yakın olmadıkları görülür. Bu testler içerisinde en kötü sonuçları CF testinin sonraki kötü sonuçların da MW testinin verdiği anlaşılmaktadır.

Çizelge 4.5. $k = 4$, $n_1 = n_2 = n_3 = n_4 = 10$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

$\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
4,4,4,4	0.052	0.050	0.049	0.049	0.094	0.048	0.044	0.023	0.017	0.039
2,2,2,2	0.054	0.053	0.050	0.052	0.097	0.052	0.045	0.023	0.019	0.042
1,1,1,1	0.054	0.053	0.050	0.052	0.099	0.053	0.046	0.026	0.019	0.042
2,1,5,1,0,5	0.052	0.052	0.052	0.050	0.097	0.050	0.045	0.025	0.019	0.040
2,1,5,1,0,2	0.054	0.054	0.056	0.052	0.104	0.053	0.045	0.026	0.015	0.045
2,1,0,8,0,6	0.049	0.048	0.055	0.048	0.092	0.045	0.048	0.022	0.016	0.038
2,0,5,0,2,0,0,2	0.054	0.056	0.073	0.053	0.110	0.053	0.053	0.027	0.020	0.049
2,0,2,0,0,2,0,0,2	0.052	0.053	0.085	0.050	0.102	0.050	0.056	0.025	0.017	0.043

Yukarıdaki deneysel I. tip hata oranları incelendiğinde J, W, BF, PB, AG ve MBF testlerinin deneysel I. tip hata oranlarının nominal I. tip hata oranına yakın olduğu görülür. Bu çizelgede deneysel I. tip hata oranı bakımında en iyi testlerin J, PB ve AG testleri olduğu

anlaşılmaktadır. Diğer yandan CF, MW, OS ve FA testlerinin deneysel I. tip hata oranları nominal I. tip hata oranına yakındır. Bu testler içerisinde deneysel I. tip hata oranı bakımından en kötü sonuçları CF testi sonraki kötü sonuçları OS testi vermiştir.

Çizelge 4.6. $k = 4$, $n_1 = n_2 = n_3 = n_4 = 20$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

$\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
4, 4, 4, 4	0.053	0.052	0.052	0.053	0.074	0.052	0.049	0.037	0.034	0.048
2, 2, 2, 2	0.049	0.048	0.050	0.049	0.070	0.048	0.048	0.034	0.035	0.043
1, 1, 1, 1	0.050	0.050	0.048	0.050	0.071	0.049	0.046	0.034	0.033	0.045
2, 1.5, 1, 0.5	0.049	0.049	0.050	0.048	0.069	0.047	0.043	0.030	0.030	0.043
2, 1, 0.5, 0.05	0.049	0.049	0.063	0.049	0.075	0.048	0.047	0.034	0.031	0.046
1, 0.8, 0.6, 0.4	0.048	0.048	0.053	0.048	0.069	0.048	0.046	0.032	0.032	0.044
2, 0.08, 0.06, 0.04	0.052	0.052	0.093	0.051	0.074	0.051	0.053	0.036	0.029	0.047

Çizelge 4.6'deki simülasyon sonuçlarından J, W, BF, PB, AG, MBF ve FA testlerinin nominal I. tip hata oranına yakın deneysel I. tip hata oranları verdiği anlaşılmaktadır. Bu testler içinde I. tip hata oranı bakımından J, W, PB ve AG testlerinin birbirlerine yakın sonuçlar verdiği söylenebilir. Diğer yandan CF, MW ve OS testlerinin deneysel I. tip hata oranları nominal I. tip hata oranına yakın değildir. Bu testler içerisinde I. tip hata oranı bakımından en kötü sonuçları CF testinin sonraki kötü sonuçları da MW ve OS testlerinin verdiği anlaşılmaktadır.

Çizelge 4.7. $k = 4$, $n_1 = n_2 = n_3 = n_4 = 25$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

$\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
4, 4, 4, 4	0.050	0.050	0.048	0.050	0.066	0.050	0.047	0.037	0.039	0.047
2, 2, 2, 2	0.051	0.051	0.051	0.051	0.067	0.050	0.048	0.038	0.038	0.047
1, 1, 1, 1	0.047	0.047	0.047	0.047	0.064	0.046	0.045	0.035	0.035	0.044
2, 1.5, 1, 0.5	0.050	0.050	0.053	0.050	0.066	0.049	0.046	0.039	0.036	0.047
2, 1, 0.5, 0.05	0.050	0.050	0.069	0.050	0.070	0.050	0.050	0.038	0.037	0.050
1, 0.8, 0.6, 0.4	0.051	0.051	0.056	0.051	0.068	0.050	0.052	0.038	0.038	0.047
2, 0.08, 0.06, 0.04	0.048	0.048	0.087	0.048	0.066	0.047	0.049	0.034	0.033	0.044

Çizelge 4.7'deki deneysel I. tip hata oranlarına göre J, W, BF, PB, AG, MBF ve FA testlerinin nominal I. tip hata oranına yakın deneysel I. tip hata oranları verdiği görülür. Diğer yandan CF, MW ve OS testlerinin deneysel I. tip hata oranları nominal I. tip hata

oranına yakın değildir. Bu testler içerisinde I. tip hata oranı bakımından en kötü sonuçları CF testinin sonraki kötü sonuçları da MW ve OS testlerinin verdiği anlaşılmaktadır.

Çizelge 4.8. $k = 4$, $n_1 = 5$, $n_2 = 10$, $n_3 = 15$, $n_4 = 20$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

$\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
4, 4, 4, 4	0.054	0.056	0.051	0.053	0.114	0.056	0.045	0.019	0.044	0.039
2, 2, 2, 2	0.053	0.057	0.047	0.053	0.110	0.055	0.042	0.019	0.047	0.038
1, 1, 1, 1	0.053	0.054	0.048	0.053	0.111	0.054	0.043	0.019	0.046	0.040
2, 1.5, 1, 0.5	0.057	0.062	0.058	0.056	0.124	0.059	0.054	0.024	0.034	0.046
0.5, 1, 1.5, 2	0.051	0.051	0.055	0.050	0.097	0.050	0.044	0.018	0.081	0.034
2, 1.5, 1, 0.2	0.058	0.064	0.057	0.057	0.130	0.057	0.050	0.022	0.033	0.049
2, 1, 1, 1	0.053	0.056	0.055	0.053	0.111	0.053	0.051	0.021	0.039	0.042
2, 1, 0.8, 0.6	0.057	0.062	0.060	0.058	0.121	0.058	0.055	0.023	0.034	0.045
2, 0.5, 0.5, 0.5	0.056	0.062	0.071	0.056	0.124	0.055	0.066	0.021	0.033	0.044
2, 0.5, 0.2, 0.02	0.052	0.056	0.070	0.049	0.130	0.050	0.057	0.021	0.029	0.046
0.02, 0.2, 0.5, 2	0.054	0.053	0.075	0.052	0.090	0.050	0.050	0.027	0.121	0.040
2, 0.2, 0.02, 0.02	0.059	0.065	0.078	0.055	0.132	0.052	0.065	0.025	0.030	0.051

Yukarıdaki çizelgedeki deneysel I. tip hata oranlarına göre J, W, BF, PB, AG ve MBF testlerinin nominal I. tip hata oranına yakın deneysel I. tip hata oranları verdiği görülmektedir. Bu testler içinde I. tip hata oranı bakımından PB testinin en iyi test ve AG testinin sonraki iyi test olduğu sonuçların incelenmesiyle görülür. Diğer yandan CF, MW, OS ve FA testlerinin deneysel I. tip hata oranları nominal I. tip hata oranına yakın değildir. Bu testler içerisinde I. tip hata oranı bakımından en kötü sonuçları CF testinin sonraki kötü sonuçları da MW testinin verdiği anlaşılmaktadır.

Çizelge 4.9. $k = 4$, $n_1 = 10$, $n_2 = 15$, $n_3 = 20$, $n_4 = 25$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

$\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
4, 4, 4, 4	0.050	0.050	0.049	0.050	0.080	0.049	0.047	0.028	0.033	0.042
2, 2, 2, 2	0.055	0.055	0.052	0.054	0.086	0.053	0.049	0.033	0.037	0.048
1, 1, 1, 1	0.050	0.050	0.050	0.050	0.081	0.050	0.047	0.031	0.034	0.043
2, 1.5, 1, 0.5	0.048	0.049	0.058	0.049	0.085	0.049	0.052	0.030	0.019	0.046
0.5, 1, 1.5, 2	0.047	0.046	0.054	0.046	0.072	0.044	0.045	0.029	0.084	0.040
2, 1.5, 1, 0.2	0.051	0.054	0.066	0.050	0.095	0.051	0.051	0.032	0.015	0.050
2, 1, 0.8, 0.6	0.049	0.049	0.066	0.049	0.075	0.048	0.048	0.030	0.110	0.044
2, 0.5, 0.2, 0.02	0.050	0.051	0.052	0.050	0.084	0.051	0.048	0.031	0.022	0.046
0.02, 0.2, 0.5, 2	0.051	0.051	0.054	0.050	0.077	0.050	0.048	0.032	0.071	0.044
2, 0.2, 0.02, 0.02	0.051	0.052	0.080	0.051	0.088	0.049	0.053	0.030	0.013	0.047

Bu çizelgedeki deneysel I. tip hata oranlarına göre J, W, BF, PB, AG ve MBF testlerinin nominal I. tip hata oranına yakın deneysel I. tip hata oranları verdiği görülmektedir. Bu testler içinde I. tip hata oranı bakımından J, W, BF, AG ve MBF testlerinin oldukça iyi sonuçları verdiği anlaşılmaktadır. Diğer taraftan CF, MW, OS ve FA testlerinin deneysel I. tip hata oranları nominal I. tip hata oranına yakın değildir. Bu testler içerisinde I. tip hata oranı bakımından en kötü sonuçları CF testinin sonraki kötü sonuçları da MW testinin verdiği görülmektedir.

Çizelge 4.10. $k = 5$, $n_1 = n_2 = n_3 = n_4 = n_5 = 10$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

σ_i^2	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
V1	0.052	0.051	0.048	0.049	0.107	0.050	0.043	0.022	0.018	0.039
V2	0.051	0.050	0.045	0.048	0.106	0.050	0.040	0.023	0.014	0.038
V3	0.050	0.049	0.045	0.047	0.103	0.048	0.040	0.021	0.016	0.038
V4	0.052	0.056	0.055	0.051	0.115	0.051	0.041	0.025	0.011	0.045
V5	0.055	0.059	0.077	0.053	0.120	0.053	0.050	0.028	0.016	0.049
V6	0.053	0.057	0.058	0.052	0.120	0.053	0.044	0.026	0.015	0.047
V7	0.054	0.054	0.092	0.052	0.110	0.050	0.054	0.023	0.016	0.043
V8	0.050	0.052	0.060	0.048	0.117	0.047	0.045	0.026	0.016	0.044

V1.(4, 4, 4, 4, 4), V2.(2, 2, 2, 2, 2), V3.(1, 1, 1, 1, 1), V4.(2, 1.5, 1, 0.5, 0.05), V5.(2, 1, 0.5, 0.02, 0.01), V6.(1, 0.8, 0.6, 0.4, 0.02), V7.(2, 0.08, 0.06, 0.04, 0.02), V8.(0.8, 0.6, 0.4, 0.2, 0.02)

Çizelge 4.10'daki deneysel I. tip hata oranları gözden geçirilirse J, W, BF, PB, AG, MBF ve FA testlerinin nominal I. tip hata oranına yakın deneysel I. tip hata oranları verdiği görülür. Bu testler içinde I. tip hata oranı bakımından J, PB ve AG testlerinin diğer testlere göre daha iyi sonuçlar verdiği söylenebilir. Diğer yandan CF, MW ve OS testlerinin deneysel I. tip hata oranları nominal I. tip hata oranına yakın değildir. Bu testler içerisinde I. tip hata oranı bakımından en kötü sonuçları CF testinin sonraki kötü sonuçları da MW ve OS testlerinin verdiği anlaşılmaktadır.

Çizelge 4.11. $k = 5$, $n_1 = n_2 = n_3 = n_4 = n_5 = 20$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

σ_i^2	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
V1	0.047	0.047	0.048	0.047	0.072	0.046	0.045	0.031	0.030	0.042
V2	0.049	0.049	0.049	0.048	0.071	0.047	0.045	0.033	0.031	0.045
V3	0.052	0.052	0.051	0.052	0.078	0.051	0.049	0.036	0.030	0.047
V4	0.052	0.053	0.063	0.052	0.084	0.052	0.046	0.034	0.028	0.050
V5	0.047	0.047	0.058	0.047	0.073	0.046	0.048	0.030	0.029	0.042
V6	0.049	0.050	0.076	0.049	0.075	0.048	0.048	0.032	0.028	0.045

V1.(4, 4, 4, 4, 4), V2.(2, 2, 2, 2, 2), V3.(1, 1, 1, 1, 1), V4.(2, 1.5, 1, 0.5, 0.05), V5.(1, 0.8, 0.6, 0.4, 0.2), V6.(0.8, 0.4, 0.2, 0.04, 0.02)

Çizelge 4.11'deki deneysel I. tip hata oranlarına göre, J, W, BF, PB, AG, MBF ve FA testlerinin nominal I. tip hata oranına yakın deneysel I. tip hata oranları verdiği açıktır. Ayrıca BF testinin varyansların homojen olmadığı durumda deneysel I. tip hata oranlarının nominal I. tip hata oranından uzaklaştığı anlaşılmıştır. Diğer yandan CF, MW ve OS testlerinin I. tip hata oranı bakımından iyi sonuçlar vermediği görülmüştür. I. tip hata oranı bakımından en kötü sonuçları CF ve OS testleri vermektedir.

Çizelge 4.12. $k = 6$, $n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = 10$, nominal $\alpha = 0,05$ ve çeşitli varyans değerleri için testlerin deneysel I. tip hata oranları

σ_i^2	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
V1	0.051	0.051	0.049	0.048	0.116	0.049	0.043	0.019	0.012	0.040
V2	0.058	0.057	0.049	0.054	0.123	0.055	0.042	0.022	0.012	0.042
V3	0.055	0.054	0.046	0.049	0.115	0.051	0.041	0.021	0.013	0.040
V4	0.051	0.054	0.069	0.048	0.123	0.049	0.045	0.023	0.010	0.043
V5	0.054	0.056	0.076	0.052	0.124	0.050	0.044	0.024	0.013	0.045
V6	0.052	0.057	0.065	0.050	0.128	0.048	0.045	0.024	0.011	0.045
V7	0.054	0.056	0.099	0.052	0.130	0.050	0.056	0.023	0.013	0.047

V1.(4, 4, 4, 4, 4, 4), V2.(2, 2, 2, 2, 2, 2), V3.(1, 1, 1, 1, 1, 1), V4.(2, 1.5, 1, 0.5, 0.05, 0.02), V5.(2, 1.5, 1, 0.02, 0.01, 0.01), V6.(1, 0.8, 0.6, 0.4, 0.02, 0.02), V7.(2, 0.08, 0.06, 0.04, 0.02, 0.002)

Yukarıdaki çizelgede yer alan deneysel I. tip hata oranları incelendiğinde J, W, BF, PB, AG ve MBF testlerinin nominal I. tip hata oranına yakın deneysel I. tip hata oranları verdiği görülmektedir. Bu testler içinde I. tip hata oranı bakımından PB ve AG testinin en iyi testler olduğu sonuçların incelenmesiyle görülür. Diğer yandan CF, MW, OS ve FA testlerinin deneysel I. tip hata oranları nominal I. tip hata oranına yakın değildir. Bu testler içerisinde I. tip hata oranı bakımından en kötü sonuçları CF testinin sonraki kötü sonuçları da MW ve OS testlerinin verdiği anlaşılmaktadır.

Yukarıdaki çizelgedeki sonuçlara göre, CF, BF, J, W, PB, AG ve MBF testlerinin güç bakımından diğer testlerden daha iyi sonuçlar verdiği anlaşılmaktadır. Bu çizelgedeki en iyi sonuçları CF testinin sonraki iyi sonuçları da J testinin verdiği görülmektedir. Diğer yandan, güç bakımından en kötü sonuçları OS testinin sonraki kötü sonuçları da MW testinin verdiği görülür.

Çizelge 4.23. Nominal $\alpha = 0,05$, $k = 5$, $n_1 = n_2 = n_3 = n_4 = n_5 = 5$ ve $\sigma_1^2 = 1$, $\sigma_2^2 = 1.5$, $\sigma_3^2 = 2$, $\sigma_4^2 = 2.5$, $\sigma_5^2 = 3$ iken testlerin güç değerleri

$\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
0,0,0,0,0.2	0.056	0.054	0.048	0.044	0.193	0.052	0.040	0.006	0.010	0.023
0,0,0,0.4,0.4	0.069	0.067	0.061	0.056	0.226	0.063	0.053	0.010	0.011	0.032
0,0.2,0.4,0.6,0.8	0.098	0.096	0.079	0.079	0.282	0.088	0.069	0.016	0.011	0.046
0,0,0,0.8,0.8	0.127	0.123	0.124	0.105	0.337	0.115	0.105	0.023	0.011	0.064
0,0,0,1,1	0.173	0.171	0.182	0.143	0.404	0.157	0.160	0.033	0.012	0.089
0,0.4,0.8,1.2,1.6	0.247	0.243	0.223	0.210	0.521	0.234	0.195	0.051	0.015	0.134
0,0.3,0.6,0.9,1.8	0.250	0.245	0.269	0.214	0.528	0.227	0.238	0.059	0.018	0.141
0,0.5,1,1.5,2	0.360	0.354	0.337	0.314	0.656	0.343	0.299	0.091	0.022	0.215
0,0,0,2,2	0.560	0.552	0.626	0.506	0.822	0.545	0.584	0.191	0.039	0.376
0,0.8,1.6,2.4,3.2	0.780	0.773	0.780	0.730	0.950	0.765	0.742	0.361	0.057	0.598
0,1,2,3,4	0.937	0.933	0.939	0.912	0.992	0.931	0.918	0.619	0.129	0.832
0,2,4,6,8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.787	1.000
0,3,6,9,12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.988	1.000
0,4,8,12,16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000

Çizelge 4.23’de yer alan sonuçlar incelenirse CF, BF, J, W, PB, AG ve MBF testlerinin güç bakımından diğer testlerden daha iyi sonuçlar verdiği görülmüştür. Sonuçlara göre, CF testinin en iyi test olduğu sonraki iyi sonuçları da J testinin verdiği anlaşılmaktadır. Diğer taraftan sonuçlar, güç bakımından OS testinin en kötü test ve MW testinin ise sonraki kötü test olduğunu göstermiştir.

Çizelge 4.24. Nominal $\alpha = 0,05$, $k = 5$, $n_1 = n_2 = n_3 = n_4 = n_5 = 5$ ve $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_3^2 = 3$, $\sigma_4^2 = 4$, $\sigma_5^2 = 5$ iken testlerin güç değerleri

$\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
0,0,0,0,0.2	0.057	0.057	0.050	0.046	0.192	0.053	0.041	0.007	0.010	0.026
0,0,0,0.4,0.4	0.066	0.064	0.057	0.053	0.218	0.060	0.046	0.010	0.010	0.033
0,0.2,0.4,0.6,0.8	0.087	0.087	0.072	0.068	0.259	0.080	0.061	0.013	0.010	0.041
0,0,0,0.8,0.8	0.103	0.102	0.096	0.083	0.287	0.092	0.080	0.019	0.012	0.052
0,0,0,1,1	0.133	0.131	0.134	0.108	0.342	0.118	0.115	0.026	0.012	0.068
0,0.3,0.6,0.9,1.8	0.181	0.180	0.176	0.152	0.438	0.166	0.152	0.039	0.014	0.100
0,0.4,0.8,1.2,1.6	0.185	0.184	0.146	0.156	0.432	0.176	0.123	0.036	0.012	0.098

Çizelge 4.24. (devam) Nominal $\alpha = 0,05$, $k = 5$, $n_1 = n_2 = n_3 = n_4 = n_5 = 5$ ve $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_3^2 = 3$, $\sigma_4^2 = 4$, $\sigma_5^2 = 5$ iken testlerin güç değerleri

$\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$	J	W	BF	PB	CF	AG	MBF	MW	OS	FA
0,0.5,1,1.5,2	0.280	0.277	0.227	0.237	0.561	0.269	0.194	0.063	0.016	0.159
0,0,0,2,2	0.381	0.377	0.422	0.334	0.676	0.363	0.376	0.108	0.024	0.239
0,0.8,1.6,2.4,3.2	0.626	0.623	0.568	0.575	0.873	0.617	0.515	0.240	0.036	0.450
0,1,2,3,4	0.828	0.824	0.776	0.787	0.968	0.826	0.733	0.429	0.063	0.679
0,2,4,6,8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.539	1.000
0,3,6,9,12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.915	1.000
0,4,8,12,16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.991	1.000

Yukardaki çizelgede CF, BF, J, W, PB, AG ve MBF testlerinin güç bakımından diğer testlerden daha iyi sonuçlar verdiği görülmüştür. Bu testler içerisinde en iyi sonuçları CF testinin sonraki iyi sonuçları da J testinin verdiği görülür. Diğer yandan, güç bakımından en kötü sonuçları OS testinin sonraki kötü sonuçları da MW testinin verdiği anlaşılmaktadır.



5. SONUÇ

Varyansların homojen olmadığı durumda yığın ortalamaların eşitliği hipotezinin testi için seçilen testlerin karşılaştırmaları yapılmıştır. Testlerin karşılaştırmaları için Monte Carlo Simülasyonu kullanılmıştır ve sonuçları dördüncü bölümde verilmiştir.

Bu bölümde dördüncü bölümdeki simülasyon sonuçlarının değerlendirilmesi yapılmıştır. Yığın sayısı 3 iken tüm örnek hacimleri için J, W ve PB testlerinin I. tip hata oranı bakımından diğer testlerden daha iyi sonuçlar verdiği görülmüştür. Diğer yandan CF, MW ve OS testlerinin de I. tip hata oranı bakımından oldukça kötü sonuçlar verdiği anlaşılmaktadır. Yığın sayısı 4 iken tüm örnek hacimlerinde I. tip hata oranı bakımından J, W, PB ve AG testlerinin diğer testlere göre daha iyi sonuçlar verdiği gözlenmiştir. Ayrıca MBF testinin de bazı durumlarda bu testler kadar iyi sonuçlar verdiği anlaşılmıştır. Bu yığın sayısı için I. tip hata oranı bakımından kötü sonuç veren testlerin CF, OS ve MW testleri olduğu belirlenmiştir. Yığın sayısı 5 ve 6 olduğunda da öncekilere benzer sonuçlarla karşılaşmıştır. Bu yığın sayıları için W, J, PB, BF, AG ve MBF testleri deneysel I. tip hata oranları bakımından oldukça iyi sonuçlar verirken CF, MW ve OS testleri de nominal I. tip hata oranına yakın sonuçlar vermemiştir.

Çizelge 4.13-24'deki testlerin deneysel güç değerlerinden aşağıdaki bulgular elde edilmiştir. Yığın sayısı 3 iken hem eşit örnek hacimleri hem de farklı örnek hacimleri için CF ve J testlerinin en güçlü testler olduğu, OS ve MW testlerinin de en düşük güç değerlerinin verdiği anlaşılmaktadır. Bunların yanında W, BF, PB, AG, MBF ve FA testlerinin de genel olarak güç bakımından iyi sonuçlar verdikleri belirlenmiştir.

Yığın sayısı 4 iken testlerin deneysel güç değerleri yığın sayısı 3 iken ki sonuçlara oldukça benzerdir. Bu yığın sayısı için en güçlü testlerin CF ve J testleri olduğu en düşük güce sahip testlerin de OS ve MW testleri olduğu görülmüştür. Diğer yandan W, BF, PB, AG, MBF ve FA testlerin de güç bakımından genel olarak iyi sonuçlar verdikleri gözlenmiştir.

Genel değerlendirme yapıldığında, J testi hem I. tip hata oranı bakımından hem de testin gücü bakımından iyi sonuçlar vermektedir. Deneysel I. tip hata oranı bakımından özellikle örnek hacmi küçük olduğu durumlarda PB testinin diğer testlerden daha iyi sonuçlar verdiği

söylenbilir. Deneysel güç değerleri bakımından en iyi sonuçları CF testinin vermiş olduğu görülmektedir. Ancak deneysel I. tip hata oranı bakımından en kötü sonuçları da CF testinin verdiği anlaşılmaktadır. Ayrıca hem I. tip hata oranı bakımında hem de deneysel güç değerleri bakımından MBF testinin BF testinden daha iyi sonuçlar verdiği anlaşılmıştır. W testinin MW testinden hem I. tip hata oranı bakımında hem de güç değerleri bakımından daha iyi sonuçlar verdiği açıktır. Sonuç olarak özet hali Çizelge 5.1’de verilmiştir.

Çizelge 5.1. Simülasyon sonuçlarına göre iyi sonuç veren testler

	Örnek Hacmi	
	Küçük	Büyük
I. Tip Hata	J, W, BF, PB, AG, MBF	J, W, BF, PB, AG, MBF, FA
Testin Gücü	CF, J, W, BF, PB, AG, MBF	CF, J, W, BF, PB, AG, MBF

KAYNAKLAR

- Anderson, E. (1935). The irises of the Gaspe Peninsula. *Bull American Iris Society*, 59, 2-5.
- Alexander, R. A. and Govern, D. M. (1994). A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational Statistics*, 19(2), 91-101.
- Aspin, A. A. (1948). An examination and further development of a formula arising in the problem of comparing two mean values. *Biometrika*, 35(1/2), 88-96.
- Bartlett, M.S. and Kendal, D.G. (1946), The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation, *Supplement to the Journal of the Royal Statistical Society*, Series B (1), 128-38.
- Bishop, T. A. and Dudewicz, E. J. (1978). Exact analysis of variance with unequal variances: test procedures and tables. *Technometrics*, 20(4), 419-430.
- Bishop, T.A. (1976). *Heteroscedastic ANOVA, MANOVA and multiple comparasion*. Unpublished Ph.D. Thesis, The Ohio State University, 1-50.
- Böckenhoff, A. and Hartung, J. (1998). Some corrections of the significance level in meta-analysis. *Journal of Mathematical Methods in Biosciences*, 40(8), 937-947.
- Boneau, C. A. (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, 57(1), 49-64.
- Box, G. E. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The annals of mathematical statistics*, 25(2), 290-302.
- Brown, M. B. and Forsythe, A. B. (1974a). Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346), 364-367.
- Brown, M. B. and Forsythe, A. B. (1974b). The small sample behavior of some statistics which test the equality of several means. *Technometrics*, 16(1), 129-132.
- Budescu, D. V. and Appelbaum, M. I. (1981). Variance stabilizing transformations and the power of the F test. *Journal of Educational Statistics*, 6(1), 55-74.
- Chen, S. Y. and Chen, H. J. (1998). Single-stage analysis of variance under heteroscedasticity. *Communications in Statistics-Simulation and Computation*, 27(3), 641-666.
- Clinch, J. J. and Keselman, H. J. (1982). Parametric alternatives to the analysis of variance. *Journal of Educational Statistics*, 7(3), 207-214.
- Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Supplement to the Journal of the Royal Statistical Society*, 4(1), 102-118.

- Dajani, A.N. (2002). *Contributions to statistical inference for some fixed and random models*. Doctoral Dissertation, Department of Mathematics and Statistics, University of Maryland, Baltimore County, USA.
- David, F. N. and Johnson, N. L. (1951). The effect of non-normality on the power function of the F-test in the analysis of variance. *Biometrika*, 38(1/2), 43-57.
- Dawid, A. P. and Stone (1982). The functional-model basis of fiducial inference. *The Annals of Statistics*, 1054-1067.
- Dudewicz, E. J. and Dalal, S. R. (1975). Allocation of observations in ranking and selection with unequal variances. *Sankhyā The Indian Journal of Statistics*, Series B, 28-78.
- Gamage, J. and Weerahandi, S. (1998). Size performance of some tests in one-way ANOVA. *Communications in Statistics-Simulation and Computation*, 27(3), 625-640.
- Gamage, J., Mathew, T. and Weerahandi, S. (2004). Generalized p-values and generalized confidence regions for the multivariate Behrens–Fisher problem and MANOVA. *Journal of Multivariate Analysis*, 88(1), 177-189.
- Gamgam, H. ve Altunkaynak, B. (2012). *Parametrik olmayan yöntemler: SPSS uygulamalı*. Ankara:Seçkin Yayıncılık, 35-72.
- Gronow, D. G. C. (1951). Test for the significance of the difference between means in two normal populations having unequal variances. *Biometrika*, 38(1/2), 252-256.
- Hartung, J., Argac, D. and Makambi, K. H. (2002). Small sample properties of tests on homogeneity in one-way Anova and Meta-analysis. *Statistical Papers*, 43(2), 197-235.
- Harwell, M. R., Rubinstein, E. N., Hayes, W. S. and Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one-and two-factor fixed effects ANOVA cases. *Journal of Educational Statistics*, 17(4), 315-339.
- Hays, W. L. (1973) *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart and Winston.
- Hsu, P. L. (1938). Contribution to the theory of Student's t-test as applied to the problem of two samples. *Statistical Research Memoirs*, 2, 1-24.
- James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, 38(3/4), 324-329.
- Keselman, H. J., Rogan, J. C. and Feir-Walsh, B. J. (1977). An evaluation of some non-parametric and parametric tests for location equality. *British Journal of Mathematical and Statistical Psychology*, 30(2), 213-221.
- Krishnamoorthy, K., Lu, F. and Mathew, T. (2007). A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models. *Computational Statistics Data Analysis*, 51(12), 5731-5742.
- Layard, M. W. J. (1973), Robust Large-Sample Tests for Homogeneity of Variance, *Journal American Statistic Association*, 68, 195-198

- Levene, H. (1960), Robust Tests for Equality of Variances, *Contributions to Probability and Statistics*, Palo Alto, California: Stanford University Press, 278-92.
- Li, X., Wang, J. and Liang, H. (2011). Comparison of several means: A fiducial based approach. *Computational Statistics Data Analysis*, 55(5), 1993-2002.
- Mehrotra, D. V. (1997). Improving the Brown-Forsythe solution to the generalized Behrens-Fisher problem. *Communications in Statistics-Simulation and Computation*, 26(3), 1139-1145.
- Moser, B. K., Stevens, G. R. and Watts, C. L. (1989). The two-sample t test versus Satterthwaite's approximate F test. *Communications in Statistics-Theory and Methods*, 18(11), 3963-3975.
- Rasch, D., Kubinger, K. D. and Moder, K. (2011). The two-sample t test: pre-testing its assumptions does not pay off. *Statistical Papers*, 52(1), 219-231.
- Reed, J.F. and Stark, D.B. (1995). Robust analysis of variances: a simulation study. *Journal of Applied Statistics*, 22,87-104.
- Rice, W. R. and Gaines, S. D. (1989). One-way analysis of variance with unequal variances. *Proceedings of the National Academy of Sciences*, 86(21), 8183-8184.
- Rogan, J. C. and Keselman, H. J. (1977). Is the ANOVA F-test robust to variance heterogeneity when sample sizes are equal?. An investigation via a coefficient of variation. *American Educational Research Journal*, 14(4), 493-498.
- Scheffe, H. (1959). *The analysis of variance*. New York: Wiley, 30-80.
- Srivastava, A. B. L. (1959). Effect of non-normality on the power of the analysis of variance test. *Biometrika*, 46(1/2), 114-122.
- Tiku, M. L. (1971). Power function of the F-test under non-normal situations. *Journal of the American Statistical Association*, 66(336), 913-916.
- Tsui, K. W. and Weerahandi, S. (1989). Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association*, 84(406), 602-607.
- Weerahandi, S. (1995a). ANOVA under unequal error variances. *Biometrics*, 589-599.
- Weerahandi, S. (1995b). Generalized confidence intervals. *In Exact Statistical Methods for Data Analysis*. New York: Springer, 143-168.
- Weerahandi, S. and Krishnamoorthy, K. (2017). A note reconciling ANOVA tests under unequal error variances. *Communications in Statistics-Theory and Methods*, 1-5.
- Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29(3/4), 350-362.
- Welch, B. L. (1947). The generalization of student's problem when several different population variances are involved. *Biometrika*, 34(1/2), 28-35.

- Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, 38(3-4), 330-336.
- Wilcox, R. R. (1988). A new alternative to the ANOVA F and new results on James's second-order method. *British Journal of Mathematical and Statistical Psychology*, 41(1), 109-117.
- Wilcox, R. R. (1989). Adjusting for unequal variances when comparing means in one-way and two-way fixed effects ANOVA models. *Journal of Educational Statistics*, 14(3), 269-278.
- Xu, L. and Wang, S. (2008). A new generalized p-value and its upper bound for ANOVA under unequal error variances. *Communications in Statistics-Theory and Methods*, 37(7), 1002-1010.



ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı :YAVUZ, Ahmet Metin
 Uyruğu :T.C.
 Doğum tarihi ve yeri :08.10.1986, Ankara
 Medeni hali :Bekâr
 Telefon :0539 810 60 41
 e-mail :metyav@gmail.com



Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek lisans	Gazi Üniversitesi/ İstatistik	Devam ediyor
Lisans	Eskişehir Osmangazi Üniversitesi/İstatistik	2009
Lise	Kanuni Lisesi	2004

İş Deneyimi

Yıl	Yer	Görev
2011-Devam ediyor	Ticaret Bakanlığı	Memur
2010-2010	SGK	Büro Personeli

Yabancı Dil

İngilizce

Yayınlar

Yavuz, A.M., Gamgam, H. ve Altunkaynak, B. (2018/17-20 Ekim). *Yığın varyansları heterojen iken ortalamaların eşitliği hipotezi için bazı testler ve karşılaştırmaları*. Tam Metin Bildiri Kitabı, Marmara Üniversitesi, 19. Uluslararası Ekonometri, Yöneylem Araştırması ve İstatistik Sempozyumu, İstanbul,1464-1474.

Hobiler

Kitap Okumak, Yüzmek, Yürümek.



GAZİ GELECEKTİR..