

Normal Olmayan Dağılımlı Populasyonlardan Alınan Örneklerde Hesaplanan Çeşitli Test İstatistiklerinin I. Tip Hata Olasılıkları Bakımından Karşılaştırılması

Mehmet MENDEŞ¹

Ensar BAŞPINAR²

Geliş Tarihi: 11.01.2002

Özet: Bu çalışmada, çeşitli varyanslara sahip Beta (4,14) ve Ki-Kare (5) dağılımı gösteren populasyonlardan alınan 3 veya 4 grulu örneklerden yararlanılarak F, Marascuilo, Welch, Brown-Forsythe ve Alexander-Govern test istatistiklerinin 100 000 simülasyon denemesi sonunda gerçekleşen I. tip hata olasılıkları bakımından karşılaştırılması yapılmıştır. Simülasyon denemeleri sonunda, normal dağılım ön şartının yerine gelmemesinin (dağılım şeklinin), F testini etkilemediği, populasyon varyansları homojenken veya aşırı heterojen değilken örnek genişlikleri dengeli olmak şartıyla, F testinin güvenilir sonuçlar verdiği görülmüştür. Diğer alternatif testlerin, genel olarak dağılım şeklinden ve örnek genişlikleri ile bu örneklerin dengeli olup olmadıklarından oldukça etkilendikleri sonucuna varılmıştır.

Anahtar Kelimeler : I. tip hata, varyansların homojenliği, varyans analizi, normal olmayan dağılımlar

Comparison of Some Test Statistics About Realized Type I Error Rate in the Non-Normal Populations

Abstract: In this study, F, Marascuilo, Welch, Brown-Forsythe and Alexander-Govern tests were compared on realized type I error rates via three or four samples which are taken from Beta (4,14) and Chi-Square (5) populations by simulation technique. At the end of 100 000 simulation trials it was determined that the violation of normality assumption was not as important as homogeneity of variance for F test but it was important for the other tests. It was concluded that other alternative tests were highly affected by the distribution shape, sample size and being balanced or unbalanced of observations in samples.

Key Words: type I error, homogeneity of variance, Non-normal distributions

Giriş

Bilindiği üzere birbirinden bağımsız iki ve daha fazla grup ortalamasını karşılaştırmak amacıyla yaygın olarak varyans analizi tekniği kullanılmaktadır (Tomarken and Serlin 1986, Tabatabaia and Tan 1986, Oshima and Algina 1992). Bu tekniğin söz konusu karşılaştırmaların yapılmasında kullanılabilmesi için; normal dağılım, gözlemlerin bağımsızlığı, etkilerin eklenebilirliği ve varyansların homojenliği gibi bazı varsayımların yerine getirilmesi gerekmektedir (Levy 1978b, Wilcox 1986, 1988, 1989) Ancak uygulamalarda bu varsayımların sağlanamadığı durumlarla çok sık karşılaşmaktadır. Özellikle, normallik ve varyansların homojenliği varsayımlarının sağlanamadığı durumlarda, deneme başında karşılaştırılan I. tip hata olasılığının (α), deneme sonunda korunamamasına ve dolayısıyla da testin gücünün ($1-\beta$) olumsuz yönde etkilenmesi söz konusu olmaktadır (Krutchkoff 1988, Lix ve ark. 1996. Böyle durumlarda genel olarak üç çözüm yolu önerilmektedir. Bunlar;

1. Verileri, transformasyon yoluyla varyans analizi tekniğine uygun hale getirmek,

2. Uygun bir parametrik olmayan (non-parametric) test ile analiz etmek,

3. Varyans analizi tekniği yerine alternatif parametrik testleri kullanmaktır.

Verilerin transformasyona tabi tutulması veya parametrik olmayan testlerin kullanılması birçok durumda iyi bir çözüm yolu olamamaktadır. Çünkü, elde edilen sonuçların yorumlanması orijinal değerler üzerinden yapılamamakta, transformasyon sonucu elde edilen değerler üzerinden yapılmakta ve sonuçların yorumlanması aşamasında bazı yanlışlıklara sebep olabilmektedir. Parametrik olmayan testler ise; varyansların heterojen olmasından olumsuz yönde etkilendikleri ve etkinlikleri de düşük olduğu için pek tavsiye edilmemektedirler. Bu durumda en iyi çözüm yolu, varyans analizine alternatif olabilecek bazı parametrik testlerin kullanılmasıdır.

Bu çalışmada, varyans analizi tekniğinin en önemli iki ön şartı olan normal dağılım ve varyansların homojenliği ön şartlarının ayrı ayrı veya birlikte sağlanamadığı durumlarda F testi ile birlikte, bu teste alternatif olan Welch, Marascuilo, Brown-Forsythe ve Alexander-Govern's testlerinin değişik deneme koşulları altında gerçekleşen I. tip hata olasılıkları bakımından karşılaştırılması yapılmıştır.

¹ Ankara Üniv. Orman Fak. Orman Hasılatı ve Biyometri Anabilim Dalı-Çankırı

² Ankara Üniv. Ziraat Fak. Biyometri ve Genetik Anabilim Dalı-Ankara

Materyal ve Yöntem

Çalışmanın materyalini, "Microsoft Fortran Power Station Developer Studio" yazılımının IMSL kütüphanesinden (Anonymous 1994) yararlanılarak üretilen tesadüf sayıları oluşturmuştur. Çalışmada değişik varyanslı Beta (4,14) ve χ^2 (5) dağılım gösteren populasyonlardan rasgele alınan 3 ve 4 örneğin (grubun) bulunması durumları ele alınmıştır. Böylece değişik varyans-örnek genişliği-dağılım şekli kombinasyonları için 100 000 simülasyon denemesi sonunda ele alınan testlerin test istatistikleri hesaplanarak, gerçekleşen I. tip hata olasılıkları ampirik olarak elde edilmiş ve bunların deneme başında kararlaştırılan I. tip hata olasılığından ne ölçüde saptıklarını irdeleme imkanı sağlanmıştır. Deneme başında I. tip hata olasılığı $\alpha=0.05$ olarak kararlaştırılmıştır. 100 000 simülasyon denemesi sonunda ampirik olarak %4.5-%5.9 arasında I. tip hata olasılığı gerçekleştiren testlerin iyi sonuçlar verdikleri varsayılmıştır. Bu varsayım, gerçekleşen I. tip hata olasılığının buradakilerden çok daha farklı ve/veya yüksek olduğu durumlarda bile, karşılaştırma bakımından önemli bir sakınca olmadığını vurgulayan araştırmacılara (Hsiung and Olejnik 1996, Ware 1997, Cliff 1997, Gorham 1998) dayanarak bir miktar dar bir aralıkta tutulmuştur. Bu çalışmadaki hesaplamalar için FORTRAN 90 programlama dilinde yazılan programlar kullanılmıştır.

Çalışmada ele alınan populasyonlar belirlenen parametrelerine göre değişik ortalama ve varyanslara sahiptirler. Bu durum göz önüne alınarak her populasyondaki gözlem değerleri olarak kabul edilen tesadüf sayıları (X_{ij});

$$Y_{ij} = \frac{X_{ij} - \mu_i}{\sigma_i}; i=1,2,\dots,k \text{ ve } j=1,2,\dots,n_i$$

şeklinde standardize edilmiştir. Böylece ele alınan dağılımların şekli değiştirilmeden $\mu=0$ ve $\sigma^2=1$ olan merkezi olmayan dağılımlara dönüştürülmüştür. Yani standardizasyondan sonra bütün populasyonlar, ortalamaları ve varyansları bakımından özdeş hale getirilmiştirlerdir. Bu ifade de;

X_{ij} : i. populasyondaki j. gözlemin standardize edilmiş değerini göstermekte olup ortalaması μ_i ve varyansı

da σ_i^2 dir,

Y_{ij} : i. populasyondaki j. Gözlemin standardize edilmiş değeri olup ortalaması ($\mu=0$) sıfır, varyansı ise ($\sigma^2=1$) birdir,

μ_i : i. populasyonun ortalamasını,

σ_i : i. populasyonun standart sapmasını,

k: populasyon sayısını,

n_i : i. populasyondan alınan örneğin genişliğini göstermektedir.

Bu dönüşümden sonra, ele alınan testlerin gerçekleşen I. tip hata olasılıkları, 100 000 simülasyon denemesi sonunda, ret edilen H_0 hipotez sayılarının

sayılıp (ret edilen hipotez sayısı / 100 000) %'ye dönüştürülerek elde edilmiştir. Ele alınan populasyonların varyansları arasında bir heterojenliğin sağlanması için gözlem değeri olarak kabul edilen standardize edilmiş tesadüf sayılarının tamamı, her populasyon için belirlenen sabit sayılarla (1, $\sqrt{2}$, $\sqrt{3}$, $\sqrt{4}$, $\sqrt{5}$, $\sqrt{7}$, $\sqrt{9}$, $\sqrt{10}$) çarpılmışlardır. Böylece, varyans oranlarının 3 populasyon için 1:1:1, 1:2:3 ve 1:5:9 ve 4 populasyon için de 1:1:1:1, 1:2:3:4 ve 1:4:7:10 olması sağlanmıştır.

Çalışmada ele alınan testler ve bunların hesaplama işlemleri aşağıda verilmiştir.

1. F testi:

$$F = \frac{S_1^2}{S_2^2} \text{ şeklinde hesaplanan test istatistiği (k-1) ve } [\Sigma(n_i-1)] \text{ serbestlik dereceli merkezi F dağılımı gösterir (Zar, 1999). Burada;}$$

S_1^2 : Gruplar arası varyansı,

S_2^2 : Gruplar içi varyansı,

k : Grup sayısı,

n_i : i. Gruptaki gözlem sayısını göstermektedir.

2. Marascuilo testi:

$$F_M = \frac{\sum_i W_i (\bar{X}_{.i} - \bar{X}_{..})^2 / (k-1)}{\Lambda}$$

şeklinde tanımlanan test istatistiğinin serbestlik dereceleri (k-1) ve $(1/\Lambda)$ olan merkezi F-dağılımı gösterdiği varsayılır. Λ hata serbestlik derecesinde bir düzeltme yapmak için kullanılır ve

$$\Lambda = \frac{3 \sum_{i=1}^k (1 - W_i / \sum_{i=1}^k W_i)^2 / (n_i - 1)}{(k^2 - 1)}$$

formülü ile hesaplanır (Marascuilo 1971).

3. Welch testi:

$$\text{Bu test istatistiği; } F_w = \frac{\sum_{i=1}^k W_i (\bar{X}_{.i} - \bar{X}_{..})^2 / (k-1)}{\left[1 + \frac{2}{3} (k-2) \Lambda\right]} \text{ şeklinde}$$

bulunur. Bu istatistiğin serbestlik dereceleri (k-1) ve $(1/\Lambda)$ olan merkezi F-dağılımı gösterdiği varsayılır. Burada,

$$W_i = \frac{n_i}{S_i^2}, i=1,2,\dots,k \text{ ve } \bar{X}_{..} = \frac{\sum_{i=1}^k W_i \bar{X}_{.i}}{\sum_{i=1}^k W_i} \text{ şeklinde}$$

hesaplanır (Welch 1951).

4. Brown-Forsythe testi: Brown-Forsythe tarafından bulunan bu test istatistiği;

$$F_{BF} = \frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2}{\sum_{i=1}^k (1 - n_i/N) S_i^2} \text{ şeklinde olup burada } N = \sum_{i=1}^k n_i \text{ dir}$$

(Brown-Forsythe, 1974). Bu istatistiğin, dağılımının yaklaşık olarak (k-1) ve f serbestlik dereceli merkezi F-dağılımı olduğu varsayılır, f,

$$\frac{1}{f} = \sum_{i=1}^k C_i^2 / (n_i - 1) \text{ şeklinde hesaplanan hata serbestlik}$$

derecesidir (Satterthwaite, 1941). C_i ise aşağıda verilen formül yardımıyla bulunur.

$$C_i = (1 - n_i/N) S_i^2 / \left[\sum_{i=1}^k (1 - n_i/N) S_i^2 \right]$$

5. Alexander-Govern's testi:

Test istatistiği:

$$Z_j = c + \frac{(c^3 + 3c)}{b} - \frac{(4c^7 + 33c^5 + 240c^3 + 855)}{(10b^2 + 8bc^4 + 1000b)} \text{ olma}$$

k üzere $AG = \sum_{i=1}^k Z_i^2$ şeklinde hesaplanır ve yaklaşık (k-1)

serbestlik dereceli χ^2 -dağılımı gösterir (Schneider ve Penfield, 1997).

Burada;

$$a = v_i - 0.5, \quad c = \sqrt{a \cdot \ln\left(1 + \frac{t_i^2}{v_i}\right)}, \text{ ve } b = 48a^2 \text{ dir.}$$

$X^+ = \sum_{i=1}^k W_i \bar{X}_i$ genel ortalamanın varyansla tartılı tahmini

$$\text{ve } W_i = \frac{1/S_i^2}{\sum_{i=1}^k 1/S_i^2} \text{ olmak üzere } t_i = \frac{\bar{X}_i - X^+}{S_{\bar{X}_i}} \text{ şeklinde}$$

hesaplanır. $v_i = n_i - 1$ dir (Hill, 1970).

Bulgular ve Tartışma

Değişik varyans-örnek genişliği-dağılım şekli kombinasyonları için 100 000 simülasyon denemesi sonunda F testi, Marascuilo testi (M), Welch testi (W), Brown-Forsythe testi (BF) ve Alexander-Govern testi (AG) için elde edilen ampirik I tip hata olasılıkları Çizelge 1-Çizelge 4'te verilmiştir.

Çizelge 1'de populasyon varyansları homojen olduğunda (1:1:1) deneme başında karşılaştırılan %5'lik I. tip hata olasılığını deneme sonunda F testinin koruduğu ampirik olarak görülmektedir.

Populasyon varyanslarının 1:2:3 olacak şekilde heterojenleştirilmesi, F testinin, örneklerde eşit sayıda gözlemin (dengeli) olması halinde, bu heterojenlikten etkilenmediği, örneklerdeki gözlem sayılarının farklılaşması (dengesiz) halinde ise, oldukça olumsuz

Çizelge 1. Beta (4,14) dağılımı gösteren 3 populasyondan alınan örneklerde 100000 simülasyon denemesi sonunda gerçekleşen I. tip hata olasılıkları (%)

Varyans oranları	$n_1 : n_2 : n_3$	F	M	W	BF	AG
1:1:1	3:3:3	5.00	5.20	3.92	2.97	4.00
	4:4:4	4.92	5.44	4.34	3.58	4.40
	5:5:5	4.85	5.41	4.47	3.92	4.54
	10:10:10	4.94	5.49	4.99	4.66	4.89
	15:15:15	5.04	5.37	5.04	4.92	4.95
	20:20:20	4.92	5.27	5.01	4.86	4.93
	100:100:100	5.00	5.15	5.08	5.00	5.06
	3:4:5	4.95	5.85	4.73	4.04	5.26
	3:6:9	5.01	7.41	6.39	6.33	9.45
	5:10:15	5.03	6.54	5.85	6.56	8.52
1:2:3	10:20:30	4.91	5.62	5.29	6.52	7.51
	5:30:50	4.92	7.24	6.61	15.52	21.28
	3:3:3	5.85	5.90	4.44	3.47	4.56
	4:4:4	5.75	6.04	4.83	4.27	4.81
	5:5:5	5.77	6.09	5.13	4.65	5.06
	10:10:10	5.58	5.76	5.23	5.21	5.14
	15:15:15	5.43	5.53	5.19	5.17	5.08
	20:20:20	5.35	5.38	5.13	5.18	5.05
	100:100:100	5.46	5.23	5.17	5.44	5.15
	3:4:5	4.38	5.66	4.53	4.16	4.72
1:5:9	3:6:9	2.91	6.03	5.20	5.50	6.86
	5:10:15	2.82	5.65	5.06	6.23	6.73
	10:20:30	2.62	5.30	4.99	6.45	6.73
	5:30:50	1.79	6.73	6.21	12.36	19.20
	3:3:3	7.35	7.35	5.75	4.31	5.75
	4:4:4	6.89	7.30	5.91	4.80	5.88
	5:5:5	6.78	7.16	5.90	5.13	5.92
	10:10:10	6.33	6.29	5.75	5.56	5.62
	15:15:15	6.28	5.88	5.48	5.78	5.41
	20:20:20	5.97	5.54	5.28	5.62	5.24
1:5:9	100:100:100	5.91	5.16	5.10	5.83	5.07
	3:4:5	4.21	6.27	5.08	4.64	4.87
	3:6:9	2.47	5.66	4.77	6.07	5.62
	5:10:15	2.35	5.54	5.00	6.66	5.82
	10:20:30	2.24	5.33	5.03	6.87	5.94
	5:30:50	1.33	5.87	5.49	9.75	13.64

olasılıklarının %5'ten düşük olduğu gözlenmektedir. Bu deneme koşullarında (örneklerin dengesiz olması), genel olarak en iyi sonuçları Marascuilo ve Welch testlerinin verdiği görülmektedir. Diğer yandan Brown-Forsythe ve Alexander-Govern testleri özellikle dengesiz gözlem kombinasyonlarından olumsuz yönde etkilenmektedir.

Populasyon varyansları 1:5:9 olduğunda; ele alınan bütün gözlem kombinasyonlarında çok güvenilir sonuçlar gerçekleştiren Welch testi ile örneklerde 5:30:50 gibi oldukça dengesiz sayılabilecek gözlem kombinasyonunu dışında, diğer gözlem kombinasyonlarında çok güvenilir sonuçlar veren Alexander-Govern testlerinin, en güvenilir testler oldukları söylenebilir. Dolayısıyla bu deneme koşullarında, bu iki testin diğer testlere tercih edilebileceği ileri sürülebilir.

Çizelge 2'de populasyon varyansları 1:1:1 olduğunda en güvenilir sonuçların F testi ile alınabileceği görülmektedir. 1:2:3 varyans kombinasyonunda, özellikle örneklerdeki farklı gözlem kombinasyonlarının F testini olumsuz yönde etkilediği, ancak F testinin bu varyans kombinasyonunda, örneklerde eşit sayıda gözlem bulunması halinde güvenilir sonuçlar verdiği ve karşılaştırılan I. tip

Çizelge 2. χ^2 (5) dağılımı gösteren 3 popülasyondan alınan örneklerde 100000 simülasyon denemesi sonunda gerçekleşen I. tip hata olasılıkları (%)

Varyans oranları	$n_1 : n_2 : n_3$	F	M	W	BF	AG
1:1:1	3:3:3	4.63	4.75	3.56	2.56	3.67
	4:4:4	4.54	4.79	3.76	2.98	3.99
	5:5:5	4.53	4.94	4.00	3.30	4.25
	10:10:10	4.79	5.38	4.84	4.32	4.88
	15:15:15	4.75	5.34	4.97	4.53	4.99
	20:20:20	4.83	5.33	5.06	4.69	5.04
	100:100:100	5.01	5.12	5.07	5.00	5.05
	3:4:5	4.66	5.42	4.35	3.57	5.00
	3:6:9	4.88	7.28	6.26	5.60	9.20
	5:10:15	4.77	6.60	5.94	6.00	8.55
	10:20:30	4.82	6.19	5.79	6.35	8.07
	5:30:50	5.03	8.69	8.07	15.98	22.60
1:2:3	3:3:3	5.52	5.64	4.29	3.12	4.41
	4:4:4	5.44	5.85	4.66	3.66	4.89
	5:5:5	5.43	5.98	4.96	4.09	5.13
	10:10:10	5.55	6.12	5.56	4.94	5.53
	15:15:15	5.51	5.87	5.49	5.20	5.46
	20:20:20	5.46	5.70	5.46	5.26	5.41
	100:100:100	5.37	5.19	5.12	5.34	5.11
	3:4:5	4.20	5.17	4.09	3.54	4.48
	3:6:9	3.06	5.37	4.57	4.79	6.20
	5:10:15	2.87	5.36	4.72	5.48	6.41
	10:20:30	2.79	5.34	5.02	6.22	6.62
	5:30:50	1.99	7.56	7.10	11.63	18.84
1:5:9	3:3:3	7.33	8.00	6.20	4.12	6.34
	4:4:4	7.07	8.46	7.01	4.74	7.00
	5:5:5	6.85	8.27	7.13	5.05	7.08
	10:10:10	6.53	7.46	6.89	5.70	6.80
	15:15:15	6.30	6.84	6.45	5.77	6.42
	20:20:20	6.33	6.59	6.28	5.94	6.23
	100:100:100	5.82	5.41	5.35	5.75	5.34
	3:4:5	4.50	6.87	5.60	4.36	5.64
	3:6:9	2.64	5.61	4.92	5.55	6.09
	5:10:15	2.57	5.67	5.16	6.41	6.19
	10:20:30	2.32	5.49	5.21	7.01	6.28
	5:30:50	1.33	6.03	5.65	9.88	13.18

hata olasılığını %5 seviyesinde koruduğu dikkati çekmektedir. Bu deneme koşullarında genel olarak en iyi sonuçları; Marascuilo ve Welch testlerinin vermektedir. Popülasyon varyanslarının 1:5:9 olması durumunda ise; genel olarak en güvenilir sonuçları Brown-Forsythe testinin verdiği görülmektedir. Diğer yandan F testinin varyansların heterojenlik derecelerine paralel olarak giderek daha sapmalı sonuçlar gerçekleştirdiği dikkati çekmektedir.

Çizelge 3'te, örneklerin alınmış oldukları popülasyonların varyansları homojenken (1:1:1) en güvenilir sonuçları F testinin verdiği, bunu sırasıyla Welch, Alexander-Govern ve Brown-Forsythe testlerinin izlediği görülmektedir. Dikkat edileceği üzere Marascuilo testinin, grup sayısının 3'ten 4'e çıkartılmasından olumsuz yönde etkilendiği gözlenmektedir.

Popülasyon varyanslarının 1:2:3:4 olması durumunda, deneme başında %5 olarak kararlaştırılan I. tip hatayı koruma bakımından sapmasız sonuçları, ele alınan bütün gözlem kombinasyonlarında oldukça güvenilir sonuçlar veren Welch testinin verdiği görülmektedir. Diğer yandan F testi ve Alexander-Govern testlerinin özellikle örneklerde eşit sayıda gözlemin bulunması durumunda, bu varyans kombinasyonundan etkilenmediği, örneklerde

Çizelge 3. Beta (4,14) dağılımı gösteren 4 popülasyondan alınan örneklerde 100000 simülasyon denemesi sonunda gerçekleşen I. tip hata olasılıkları (%)

Varyans oranları	$n_1 : n_2 : n_3 : n_4$	F	M	W	BF	AG
1:1:1:1	3:3:3:3	4.94	8.00	4.97	2.67	4.67
	4:4:4:4	5.03	7.41	5.06	3.50	4.96
	5:5:5:5	4.97	7.36	5.32	3.94	5.28
	10:10:10:10	4.87	6.24	5.17	4.60	5.12
	15:15:15:15	4.89	5.84	5.13	4.77	5.05
	20:20:20:20	5.00	5.71	5.14	4.92	5.09
	100:100:100:100	4.92	5.06	4.96	4.92	4.93
	3:5:7:9	5.08	8.95	6.80	5.38	9.38
	4:8:12:16	4.92	7.66	5.94	5.04	7.49
	10:14:18:22	5.05	6.21	5.47	5.41	6.22
	5:10:15:20	4.95	7.30	6.15	6.50	9.66
	10:20:30:40	4.88	6.08	5.43	6.63	8.26
1:2:3:4	3:3:3:3	6.04	8.84	5.70	3.33	5.29
	4:4:4:4	5.86	8.21	5.73	4.06	5.52
	5:5:5:5	5.92	7.83	5.75	4.61	5.59
	10:10:10:10	5.81	6.60	5.54	5.25	5.41
	15:15:15:15	5.75	6.19	5.41	5.45	5.33
	20:20:20:20	5.66	5.81	5.26	5.44	5.17
	100:100:100:100	5.80	5.12	5.01	5.77	5.01
	3:5:7:9	2.97	7.47	5.53	5.24	6.78
	4:8:12:16	3.36	6.88	5.28	5.47	6.02
	10:14:18:22	3.35	5.83	5.10	5.82	5.58
	5:10:15:20	2.54	6.59	5.53	6.51	7.82
	10:20:30:40	2.54	5.74	5.22	6.70	7.35
1:4:7:10	3:3:3:3	6.93	10.22	6.78	3.67	6.18
	4:4:4:4	6.76	9.41	6.65	4.52	6.20
	5:5:5:5	6.79	8.80	6.52	4.95	6.15
	10:10:10:10	6.49	8.86	5.78	5.75	5.61
	15:15:15:15	6.37	6.37	5.63	5.95	5.50
	20:20:20:20	6.24	5.92	5.36	5.91	5.27
	100:100:100:100	6.22	5.18	5.07	6.16	5.06
	3:5:7:9	2.85	7.06	5.25	5.74	5.71
	4:8:12:16	3.27	7.08	5.47	6.07	5.71
	10:14:18:22	3.32	5.94	5.27	6.44	5.47
	5:10:15:20	2.19	6.00	5.02	6.80	6.44
	10:20:30:40	2.21	5.55	5.08	7.12	6.52

farklı gözlem kombinasyonlarının bulunmasından olumsuz yönde etkilendikleri ve F testinin genel olarak %5'ten düşük, Alexander-Govern testinin ise %5'ten yüksek I. tip hata olasılıkları gerçekleştirdikleri dikkati çekmektedir.

1:4:7:10 varyans kombinasyonunda, yine en güvenilir sonuçları, örneklerde 3'er, 4'er ve 5'er gözlemin bulunması dışında Welch testinin verdiği görülmektedir. Diğer testlerin genel olarak, 1:2:3:4 varyans kombinasyonuna göre biraz daha sapmalı sonuçlar verdikleri dikkati çekmektedir.

Çizelge 4'te, 1:1:1:1 varyans kombinasyonunda, F testinin dağılım şekline etkilenmediği ve ele alınan bütün gözlem kombinasyonlarında kararlaştırılan I. tip hatayı koruma bakımından oldukça güvenilir sonuçlar verdiği ve bunu Welch testinin izlediği görülmektedir. Bu deneme koşullarında genel olarak en sapmalı sonuçları, grup sayısının artırılmasından olumsuz yönde etkilenen Marascuilo testinin verdiği dikkati çekmektedir. Varyanslarının 1:2:3:4 kombinasyonunda, en güvenilir sonuçları, bütün gözlem kombinasyonlarında kararlaştırılan I. tip hatayı %5 seviyesinde koruyabilen Welch testi vermekte, bunu Brown-Forsythe testi izlemektedir. F testi, örneklerde eşit sayıda gözlemin

Çizelge 4. χ^2 (5) dağılımı gösteren 4 populasyondan alınan örneklerde 100000 simülasyon denemesi sonunda gerçekleşen 1.tip hata olasılıkları (%)

Varyans oranları	$n_1 : n_2 : n_3 : n_4$	F	M	W	BF	AG
1:1:1:1	3:3:3:3	4.66	7.47	4.78	2.42	4.51
	4:4:4:4	4.58	6.96	4.65	2.89	4.85
	5:5:5:5	4.49	6.77	4.75	3.15	5.12
	10:10:10:10	4.71	6.39	5.33	4.24	5.49
	15:15:15:15	4.79	6.08	5.37	4.54	5.44
	20:20:20:20	4.85	5.95	5.41	4.69	5.43
	100:100:100:100	4.88	5.20	5.09	4.88	5.09
	3:5:7:9	4.69	8.66	6.58	4.56	9.37
	4:8:12:16	4.59	7.73	5.96	4.30	7.81
	10:14:18:22	4.77	6.41	5.63	4.96	6.58
	5:10:15:20	4.84	8.18	6.98	6.01	10.72
	10:20:30:40	4.86	6.94	6.28	6.47	9.28
1:2:3:4	3:3:3:3	5.79	8.77	5.59	3.02	5.40
	4:4:4:4	5.56	8.22	5.63	3.54	5.73
	5:5:5:5	5.67	8.12	5.85	4.07	6.12
	10:10:10:10	5.55	7.09	5.89	4.84	6.00
	15:15:15:15	5.73	6.79	5.90	5.34	6.03
	20:20:20:20	5.64	6.32	5.76	5.38	5.73
	100:100:100:100	5.73	5.41	5.30	5.69	5.29
	3:5:7:9	3.15	6.86	4.94	4.59	6.41
	4:8:12:16	3.32	6.76	5.14	4.78	6.08
	10:14:18:22	3.56	6.20	5.45	5.77	6.00
	5:10:15:20	2.59	6.46	5.36	6.04	7.76
	10:20:30:40	2.56	5.99	5.43	6.61	7.61
1:4:7:10	3:3:3:3	6.76	11.1 7	7.44	3.55	6.93
	4:4:4:4	6.62	10.6 2	7.65	4.12	7.39
	5:5:5:5	6.36	10.0 9	7.69	4.44	7.52
	10:10:10:10	6.48	8.65	7.43	5.54	7.29
	15:15:15:15	6.37	7.61	6.82	5.79	6.73
	20:20:20:20	6.29	7.04	6.45	5.89	6.41
	100:100:100:100	6.20	5.50	5.38	6.14	5.37
	3:5:7:9	2.84	7.46	5.56	5.20	6.31
	4:8:12:16	3.16	7.55	5.86	5.38	6.32
	10:14:18:22	3.39	6.88	5.94	6.29	6.21
	5:10:15:20	2.39	6.48	5.44	6.77	6.99
	10:20:30:40	2.40	5.85	5.35	7.18	6.92

bulunması durumunda bu varyans kombinasyonundan etkilenmemekte, farklı gözlem kombinasyonlarının bulunmasından olumsuz yönde etkilenmektedir. En sapmalı sonuçları Marascuilo ve Alexander-Govern testlerinin verdiği göze çarpmaktadır.

Populasyon varyanslarının 1:4:7:10 olarak daha da heterojenleştirilmesi; öncelikle Welch testi olmak üzere bütün testleri olumsuz yönde etkilemektedir.

Sonuç

Çalışmada göz önünde tutulan deneme koşullarında yapılan 100 000 simülasyon denemesi sonunda;

F testinin:

1. Dağılım şekline etkilenmediği, yani normallik ön şartından pek etkilenmediği,
2. Eşit örnek genişliklerinde, homojenlikten olan küçük sapmalardan etkilenmediği,

3. Gruplardaki gözlem kombinasyonlarının dengeli olup olmamasından etkilendiği,

4. Varyanslar homojense en sağlam (robust) test olduğu,

5. Homojenlikten sapma derecesine paralel olarak giderek sapmalı sonuçlar verdiği, yani varyansların homojenliği ön şartının sağlanıp sağlanmamasından oldukça etkilendiği,

6. Grup sayısından etkilenmediği

Diğer testlerin:

1. Dağılım şekline oldukça etkilendikleri,

2. Gruplardaki gözlem sayılarının dengeli olup olmamasından etkilendikleri,

3. Genel olarak bütün koşullar için ele alınan testlerden hiçbirinin F testinin yerine kullanılamayacağı ve bunların belirli deneme koşullarında F testine alternatif olabilecekleri,

4. Özellikle Marascuilo testinin, grup sayısının artırılmasından olumsuz yönde etkilendiği,

5. Dağılım şekli ve gruplardaki gözlem sayılarına bağlı olarak, bu testlerin de aşırı heterojenlikten oldukça fazla etkilendikleri sonucuna varılmıştır.

Kaynaklar

- Anonymous, 1994. FORTRAN Subroutines for Mathematical Applications. IMSL MATH/LIBRARY. Vol. 1-2. Visual Numerics, Inc., Houston, USA.
- Brown, M. B. and A. B. Forsythe, 1974. the small sample behavior of some statistics which test the equality of several means. *technometrics*, 16, 129-132.
- Cliff, N. 1997. Robustness and Power of D Statistics in Comparison to t For Paired Data. DAI-B 57/07, s. 4774.
- Gorham, J. L. 1998. The Effects on Type I Error Rate and Power of Selected Competitors to the ANOVA F Test in Randomized Block Design Under Non-Normality and Variance Heterogeneity. DAI-B 58/12, s. 6650.
- Hill, G. W. 1970. Algorithm 395. Student's t-Distribution. *Communications of the ACM*, 13, 617-619.
- Hsiung, T. H. and S. Olejnik, 1996. Type I error rates and statistical power for the James 2nd-Order test and the univariate F-test in 2-way fixed effects anova models under heteroscedasticity and/or nonnormality. *J. of Experimental Education*, 65 (1) 57-71.
- Krutchkoff, R. G. 1988. One-way fixed effects analysis of variance when the error variances may be unequal. *J. of Statistical Computation and Simulation*, 30, 259-271.
- Levy, K. J. 1978b. some empirical power results associated with Welch's robust analysis of variance technique. *J. of Statistical Computation and Simulation*, 8, 43-48.
- Lix, L. M., J. C. Keselman, and H. J. Keselman, 1996. Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research*, 66, 579-619.

- Oshima, T. C. and J. Algina, 1992. Type I error rates for James's second-order test and Wilcoxon's H_m test under heteroscedasticity and non-normality. *British Journal of Mathematical and Statistical Psychology*, 45, 255-263.
- Satterhwaite, F. E., 1941. Synthesis of variance. *Psychometrika*, 6, 309-316
- Schneider, P. J. and D. A. Penfield, 1997. Alexander and Govern's Approximation: Providing an alternative to ANOVA under variance heterogeneity. *The J. of Experimental Education*, 65, 271-286.
- Tabatabaia, M. A. and W. Y. Tan, 1986. Some Monte Carlo studies on the comparison of several means under heteroscedasticity and robustness with respect to departure from normality. *J. of Biometry*, 7, 801-814.
- Tomarken, A. J. and R. C. Serlin, 1986. Comparison of ANOVA alternatives under variance heterogeneity and specific non-centrality structures. *Psychological Bulletin*, 99, 90-99.
- Ware, W. B. 1997. Detecting Departures From Normality: A Monte Carlo Simulation of a New Omnibus Test Based on Moments. *DAI-A 58/04*, s. 1196.
- Welch, B. L. 1951. On the comparison of several mean values: An alternative approach. *Biometrika*, 38,330-336.
- Wilcox, R. R. 1986. New Monte Carlo results on the robustness of the ANOVA F , W and F^* statistics. *J. of Statistical Computation and Simulation*, 15, 933-943.
- Wilcox, R. R. 1988. A new alternative to the ANOVA F test and new results on James's second-order method. *J. of Mathematical and Statistical Psychology*, 41, 109-117.
- Wilcox, R. R. 1989. Adjusting for unequal variances when comparing means in one-way and two-way effects ANOVA models. *J. of Educational Statistics*, 14, 269-278.
- Zar, J. H. 1999. *Biostatistical Analysis*. Fourth Edition. Simon & Schuster/A Viacom Co., New Jersey, USA.
-

İletişim adresi:
Ensar BAŞPINAR
Ankara Üniv. Ziraat Fakültesi,
Biyometri ve Genetik Anabilim Dalı-Ankara
Tel : 0 312 317 05 50/1251
E-mail: ensarb@isbank.net.tr