

Levene, Bartlett, Neyman-Pearson ve Bartlett 2 Testlerinin 1.Tip Hata Olasılıkları Bakımından Karşılaştırılması

Mehmet MENDEŞ¹

Geliş Tarihi: 08.04.2002

Özet: Bu çalışmada, Levene, Bartlett, Neyman-Pearson ve Bartlett 2 testlerinin değişik deneme koşullarında gerçekleşen 1.Tip hata olasılıkları bakımından karşılaştırılmaları yapılmıştır. Yapılan 50000 simülasyon denemesi sonunda ele alınan bütün testlerin dağılım şekli ve örnek hacminden oldukça etkilendikleri görülmüştür.

Anahtar Kelimeler: varyansların homojenliği, varyans analizi, 1.Tip hata olasılığı varyansların homojenlik testleri

The Comparisons of Levene, Bartlett, Neyman-Pearson and Bartlett 2 Tests in Terms of Actual Type I Error Rates

Abstract: In this study, Levene, Bartlett, Neyman-Pearson and Bartlett 2 tests were compared in terms of actual Type I error rates under various experimental conditions. At the end of 50000 simulation experiment it was seen that all of these tests were effected distribution shapes and sample size.

Key Words: homogeneity of variance, variance analysis, Type I error rates, homogeneity of variance tests

Giriş

Normal dağılım gösteren populasyonların ortalamalarını, bunlardan çekildiği kabul edilen örnekler yardımıyla karşılaştırmada kullanılan en yaygın teknik varyans analizi tekniğidir (Sokal ve Rohlf 1995). Ancak bu kadar yaygın bir şekilde kullanım alanı bulan bu teknikten beklenen yararların sağlanabilmesi için, elde edilen verilerde bazı ön şartların yerine getirilmesi gerekmektedir (normal dağılım, varyansların homojenliği, gözlemlerin bağımsızlığı, etkilerin eklenebilir oluşu). Zira ön şartların sağlanmadığı durumlarda varyans analizi tekniğinin kullanılması bir yandan deneme başında kararlaştırılan I.Tip hata olasılığının deneme sonunda korunamamasına neden olurken, bir yandan da testin gücünde olumsuz değişmelerin meydana gelmesine neden olmaktadır (Winer ve ark. 1991, Zar 1999). Bu durum ise, elde edilecek sonuçların yorumlanma aşamasında yanlışlılara düşülmesine neden olmaktadır.

Bu ön şartlardan gözlemlerin bağımsızlığı denemenin bir fonksiyonu olduğu için deneme başında sağlanabilir. Ancak normal dağılım ve varyansların homojenliği ön şartları ise çalışılan populasyonun bir fonksiyonu oldukları için deneme başında sağlanabilmeleri mümkün olamamaktadır (Brown ve Forsythe 1974). Söz konusu ön şartlardan varyans analizi sonuçlarının önemlilik düzeyini ve hassasiyetini en fazla etkileyen ön şart varyansların heterojen olmasıdır (Piepho 1996, Ott 1998, Ghost ve Kim 2001). Zira normallikten sapmaların aşırı derecede olmadıkları sürece bu ön şartın sağlanmaması deneme başında kararlaştırılan I.Tip hata olasılığını pek etkilememektedir (Tomarken ve Serlin 1986, Wludyka ve Nelson 1997, Zimmerman 1998).

Bu çalışmada, değişik deneme koşullarının ele alınan testlerin deneme sonunda gerçekleşen 1.Tip hata olasılıklarını nasıl etkilediği irdelenmiştir.

Materyal ve Yöntem

Bu çalışmada, dağılım şekilleri simetrik olan t (10), Normal (0,1) ve Beta (14,10) populasyonlarının kendileri dahil ikili kombinasyonlarından alınan ve 3,4,10, 20,30 gözlem içeren örneklerde ele alınan testlerin I. Tip hata bakımından karşılaştırmaları simülasyon tekniği ile yapılmıştır. Çalışmanın materyalini, yukarıdaki dağılımlara uygun olarak Microsoft Fortran Developer Stüdyo'nun IMSL kütüphanesinden üretilen tesadüf sayıları oluşturmaktadır (Anonymous 1994). Söz konusu populasyonlar belirtilen parametrelerine göre farklı ortalama ve varyansa sahip olmalarından dolayı bu populasyonlar standardize edilmiştir. Böylece bu populasyonların dağılım şekilleri değiştirilmeden ortalamalarının 0, standart sapmalarının da 1 olması sağlanmıştır. Daha sonra farklı gözlem kombinasyonlarının etkisini irdelemek amacıyla örnek genişlikleri; 3:5, 5:15 ve 15:30 olarak ta alınmıştır. Çalışma için gerekli olan bütün hesaplamalar FORTRAN programları ile yürütülmüştür. Ele alınan testlerin I.Tip hata olasılığı bakımından karşılaştırılmasında, göz önüne alınan her deneme koşulu 50000 defa tekrarlanmıştır. Bu çalışmada göz önüne alınan testler;

Bartlett testi: Bartlett testi, varyansların homojenliğinin test edilmesinde en yaygın olarak kullanılan testlerden birisidir. Özellikle normal dağılım verilerde iyi sonuçlar veren bu test için test istatistiği, $B = \ln S_p^2 \left(\sum v_j \right) - \sum v_j \ln S_j^2$ şeklinde hesaplanmaktadır. Bu şekilde hesaplanan test istatistiği yaklaşık olarak (k-1) serbestlik dereceli χ^2 dağılımı gösterir. Diğer yandan, χ^2 dağılımına yaklaşımı daha da arttırmak amacıyla B istatistiğinde,

¹ Ankara Üniv. Çankırı Orman Fak.-Çankırı

$$C = 1 + \frac{1}{3(k-1)} \left[\sum \frac{1}{v_j} - \frac{1}{\sum v_j} \right]$$

şeklinde bir düzeltme yapılmıştır. Yapılan bu düzeltmeden sonra düzeltilmiş Bartlett istatistiği, $B' = \frac{B}{C}$ olarak bulunur.

B' istatistiği, $(k-1)$ serbestlik dereceli χ^2 dağılımı gösterir.

Bu eşitliklerdeki; S_p^2 ; toplanmış varyansı, S_j^2 j. grubun varyansını, k; grup sayısını ve v_j ise j. grubun serbestlik derecesini göstermektedir. Eğer B' istatistiği $(k-1)$ serbestlik dereceli χ^2 tablo değerinden büyüğe ilgili H_0 hipotezi ret edilir (Zar 1999).

Bartlett 2 testi: Bu test için test istatistiği;

$$B_2 = \frac{WB}{(k-1)(b-B)}$$

şeklinde hesaplanmaktadır (Conover ve ark. 1981).

$$\text{Buradaki; } W = \frac{(k+1)}{(C-1)^2} \text{ ve } b = \frac{W}{(C+2W)}$$

hesaplanmaktadır. B_2 istatistiği, $(k-1)$ ve w serbestlik dereceli merkezi F dağılımı gösterdiği varsayılır.

Neyman-Pearson testi: 1931 yılında Neyman ve Pearson tarafından önerilen bu test varyansların homojenliğini test etmek için kullanılan ilk testlerden birisidir. Normal dağılım gösteren verilerin söz konusu olması durumunda kullanılması önerilen bu test için test

$$\text{istatistiği, } NP = N \ln \left(\frac{N-k}{N} S_p^2 \right) - \sum n_j \ln \left(\frac{n_j-1}{n_j} S_j^2 \right)$$

Buradaki N, denemede toplam gözlem adedini göstermektedir. Bu şekilde hesaplanan test istatistiği, $(k-1)$ serbestlik dereceli χ^2 dağılımı gösterir (Conover ve ark. 1981).

Levene testi: Bu test, ele alınan grupların kendi ortalamasından olan sapmalarının mutlak değerlerine varyans analizinin uygulanması esasına dayanır. Bu test için test istatistiği,

$$L = \frac{\sum n_j (\bar{X}_j - \bar{X}_{..})^2 / (k-1)}{\sum \sum (X_{ji} - \bar{X}_j)^2 / (N-k)}$$

şeklinde hesaplanmaktadır. Burada $X_{ji} = |Y_{ji} - \bar{Y}_j|$ olup,

Y_{ji} , j. gruptaki i. gözlemi göstermektedir. Hesaplanan L istatistiği, $(k-1)$ ve $(N-k)$ serbestlik dereceli F tablo değerinden büyük ise ilgili H_0 hipotezi ret edilir.

Bulgular ve Tartışma

50000 simülasyon denemesi sonunda elde edilen sonuçlar Çizelge 1 ve Çizelge 2'de topluca verilmiştir. Çizelge 1 incelendiğinde her iki ön şartın sağlandığı durumda 1.Tip hatayı %5 seviyesinde koruma bakımından en güvenilir sonuçları Bartlett testinin verdiği görülür. Dikkat edileceği üzere nisbeten Levene testi hariç, diğer testler özellikle küçük gözlem kombinasyonlarından olumsuz yönde etkilenmektedir. Ele alınan populasyonların t (10) olması durumunda, %5.5-8.0 arasında 1.Tip hata olasılıkları gerçekleştiren Levene testinin en güvenilir sonuçları verdiği görülmektedir. Buna karşın bu dağılımlar bakımından Bartlett testinin olumsuz yönde etkilediği dikkati çekmektedir. Neyman-Parson ve Bartlett2 testleri ise bu deneme koşullarında en sapmalı sonuçları gerçekleştiren testlerdir. Her iki populasyonun dağılımının Beta (14,10) olması durumunda ise; Bartlett testinin genel olarak %4.0 civarında 1.Tip hata olasılıkları gerçekleştirdiği ve bu hata olasılıklarının da Bradly ya da Cochran kriterlerine göre bir değerlendirme yapıldığı zaman kabul edilebilir hata olasılıkları oldukları söylenebilir (Cochran 1954, Bradly 1978). Levene testi bakımından gerçekleşen 1.Tip hata olasılıklarının genel olarak söz konusu populasyonların dağılımlarının t (10) olması durumu ile çok benzer oldukları dikkati çekmektedir. Dolayısıyla populasyon dağılımlarının t (10) ve Beta (14,10) olmasının, kararlaştırılan 1.Tip hata olasılığının korunması bakımından Levene testini pek etkilemediği ileri sürülebilir. Diğer yandan bu deneme koşullarından en olumlu yönde etkilenen testin dağılımların Normal (0,1) ve t (10) olması durumunda en sapmalı sonuçları veren testlerden birisi olan Bartlett2 testinin olduğu dikkati çekmektedir. Bu deneme koşullarında en sapmalı sonuçları gerçekleştiren test ise Neyman-Pearson testidir. Ele alınan örneklerden birinci örneğin alınmış olduğu populasyonun dağılımın t(10), diğerinin ise Beta (14,10) olması durumunda, genel olarak en güvenilir sonuçları Levene ve Bartlett testlerinin gerçekleştirdikleri görülmektedir. Ancak her iki test bakımından gerçekleşen 1.Tip hata olasılıklarının ele alınan bir çok koşulda %6.0 ve daha fazla olduğu dikkati çekmektedir. Ancak bu hata olasılıklarının büyük çoğunluğu Bardly ya da Cochran kriterlerine uygun oldukları gözlenmektedir. Bu koşullarda Neyman-Pearson ve Bartlett 2 testleri en sapmalı sonuçları gerçekleştiren testlerdir. Ele alınan dağılımlardan birincisinin Normal (0,1) yani Z-dağılımı, diğerinin ise t (10) olması durumunda genel olarak en güvenilir sonuçları başta Bartlett testi olmak üzere Levene testinin verdiği söylenebilir. Söz konusu dağılımlardan ilkinin Normal (0,1), diğerinin ise Beta (14,10) olması durumundan genel olarak bütün testlerin olumlu yönde etkilendikleri, ancak en güvenilir sonuçları Bartlett testinin verdiği görülmektedir.

Sonuç

Yapılan 50000 simülasyon denemesi sonunda;

1-Örneklerin alınmış oldukları populasyonların her ikisinin de Normal (0,1) olması durumunda en güvenilir sonuçları Bartlett testinin gerçekleştirdiği görülmüştür.

Çizelge 1. Her iki populasyonun Normal (0,1), t (10) ve Beta (14,10) olması durumunda gerçekleşen 1.Tip hata olasılıkları (α)

n_1, n_2	Normal (0,1):Normal (0,1)				t (10): t (10)				Beta (14,10): Beta (14,10)			
	Lev.	Bar.	NP.	Bar2	Lev.	Bar.	NP.	Bar2	Lev.	Bar.	NP.	Bar2
3:3	0,049	0,048	0,149	0,127	0,055	0,055	0,165	0,140	0,047	0,046	0,146	0,122
4:4	0,081	0,051	0,117	0,099	0,079	0,058	0,133	0,113	0,085	0,046	0,123	0,093
5:5	0,074	0,050	0,098	0,084	0,075	0,065	0,123	0,107	0,076	0,048	0,114	0,080
6:6	0,067	0,050	0,088	0,076	0,068	0,067	0,114	0,100	0,067	0,043	0,108	0,067
7:7	0,064	0,050	0,080	0,071	0,065	0,070	0,109	0,097	0,065	0,044	0,094	0,064
8:8	0,063	0,051	0,079	0,070	0,064	0,076	0,110	0,099	0,061	0,043	0,083	0,059
9:9	0,066	0,054	0,079	0,071	0,060	0,075	0,103	0,094	0,062	0,043	0,073	0,059
10:10	0,061	0,050	0,071	0,064	0,060	0,075	0,101	0,093	0,059	0,044	0,069	0,056
20:20	0,056	0,052	0,061	0,058	0,055	0,089	0,102	0,098	0,055	0,040	0,062	0,045
30:30	0,054	0,051	0,057	0,055	0,055	0,096	0,105	0,102	0,053	0,041	0,058	0,044
3:5	0,074	0,048	0,135	0,108	0,080	0,057	0,156	0,125	0,077	0,047	0,137	0,106
5:15	0,058	0,048	0,091	0,071	0,063	0,073	0,125	0,103	0,059	0,042	0,082	0,062
15:30	0,057	0,051	0,063	0,058	0,055	0,091	0,106	0,100	0,056	0,043	0,066	0,049

Çizelge 2. Normal (0,1), t (10) ve Beta(14,10) populasyonlarının ikili kombinasyonlarının bulunması durumunda gerçekleşen 1.Tip hata olasılıkları (α)

n_1, n_2	t (10): Beta (14,10)				Normal (0,1):t (10)				Normal (0,1): Beta (14,10)			
	Lev.	Bar.	NP.	Bar2	Lev.	Bar.	NP.	Bar2	Lev.	Bar.	NP.	Bar2
3:3	0,055	0,051	0,162	0,138	0,058	0,049	0,158	0,142	0,045	0,045	0,144	0,120
4:4	0,083	0,053	0,126	0,107	0,077	0,057	0,130	0,110	0,082	0,047	0,112	0,093
5:5	0,073	0,054	0,108	0,091	0,072	0,058	0,110	0,096	0,075	0,051	0,098	0,085
6:6	0,068	0,060	0,101	0,089	0,065	0,062	0,105	0,085	0,070	0,049	0,086	0,075
7:7	0,068	0,063	0,100	0,089	0,064	0,062	0,102	0,082	0,064	0,045	0,075	0,066
8:8	0,065	0,063	0,094	0,085	0,064	0,065	0,098	0,080	0,062	0,045	0,069	0,061
9:9	0,063	0,065	0,090	0,082	0,062	0,065	0,093	0,078	0,063	0,046	0,069	0,062
10:10	0,066	0,067	0,091	0,083	0,060	0,067	0,091	0,075	0,060	0,047	0,066	0,059
20:20	0,060	0,083	0,096	0,091	0,060	0,076	0,100	0,094	0,055	0,045	0,053	0,051
30:30	0,063	0,100	0,108	0,106	0,063	0,103	0,108	0,109	0,056	0,045	0,051	0,049
3:5	0,081	0,051	0,135	0,113	0,081	0,055	0,140	0,118	0,075	0,047	0,132	0,105
5:15	0,079	0,069	0,104	0,095	0,079	0,072	0,109	0,092	0,060	0,046	0,091	0,070
15:30	0,073	0,094	0,100	0,102	0,073	0,089	0,099	0,099	0,056	0,046	0,057	0,052

Dolayısıyla dağılımların Normal (0,1) olması durumunda sadece Bartlett testinin uygulanması gerektiği rahatlıkla ileri sürülebilir. Bu testi Levene testinin izlediği görülmektedir. Neyman-Pearson ve Bartlett 2 testlerinin ise özellikle küçük örnek hacimlerinden olumsuz yönde etkilendikleri görülmekle birlikte, örnek hacmi arttıkça gerçekleşen 1.Tip hata olasılıklarının %5'e yaklaşma eğilimine girdiği görülmüştür.

2-Populasyon dağılımlarının t (10):t (10) olması durumunda Bartlett testinin az da olsa %5'den sapan 1.Tip hata olasılıkları gerçekleştirdiği, buna karşın Levene testinin söz konusu dağılımların Normal (0,1) olması durumuyla hemen hemen aynı sonuçları gerçekleştirdiği dikkatli çekmektedir. Diğer yandan bu deneme koşullarından en olumsuz yönde etkilenen testlerin, Neyman-Pearson ve Bartlett 2 oldukları görülmüştür.

3-Dağılımların Beta (14,10) olması durumunda Bartlett testinin genel olarak %4,0-5,0 arasında 1.Tip hata olasılıkları gerçekleştirdiği görülmüştür. Levene testinin ise söz konusu dağılımların Normal (0,1) ya da t (10) olması durumunda gerçekleşen 1.Tip hata olasılıklarıyla hemen hemen aynı olasılıkları gerçekleştirdiği dikkatli çekmektedir. Dolayısıyla ele alınan her iki örneğin de dağılımları aynı olduğu sürece Levene testi bakımından gerçekleşen 1.Tip hata olasılıkları üzerine dağılım şeklinin bir etkisinin olmadığı söylenebilir. Bu deneme

koşullarından en olumlu yönde etkilenen testlerin özellikle 6-30 arasında olması durumunda %5'e yaklaşan 1.Tip hata olasılıkları gerçekleştiren Neyman-Pearson ve Bartlett 2 testlerinin gerçekleştirdiği söylenebilir.

4-Farklı gözlem kombinasyonlarının ele alınan testlerin gerçekleşen 1.Tip hata olasılıkları üzerine etkilerine bakıldığında, genel olarak dağılım şekli ne olursa olsun en olumsuz yönde etkilenen testlerin Neyman-Pearson ve Bartlett 2 testlerinin oldukları söylenebilir. Dolayısıyla bu deneme koşullarında Neyman-Pearson ve Bartlett 2 testlerinden herhangi birisinin kullanılmasının pek uygun olmadığı ileri sürülebilir.

5-Ele alınan populasyonlardan birincisinin t (10), diğerinin ise Beta (14,10) olması halinde, genel olarak en güvenilir sonuçları Levene ve Bartlett testlerinin gerçekleştirdiği görülmüştür. Zira Levene testinin özellikle örnek hacminin 6-30 arasında olması durumunda %6,0-6,8 arasında 1.Tip hata olasılıkları gerçekleştirdiği, Bartlett testinin ise örnek hacminin 3-10 arasında olması durumunda %5,1-6,7 arasında 1.Tip hata olasılıkları gerçekleştirdiği ve her iki test bakımından gerçekleşen 1.Tip hata olasılıklarının özellikle Cochran kriterine göre bir değerlendirme yapıldığında kabul edilebilir 1.Tip hata olasılıkları oldukları söylenebilir. Diğer yandan Bartlett testi bakımından gerçekleşen 1.Tip hata olasılıklarının örnek hacminin artmasına paralel olarak %5,0'den az da olsa

sapma eğilimine girdiği dikkati çekmektedir. Neyman-Pearson ve Bartlett 2 testlerinin ise yine en sapmalı sonuçlar veren testler oldukları görülmüştür. Dolayısıyla bu koşullarda Levene ya da Bartlett testlerinden her hangi birisinin kullanılması, ele edilecek sonuçların güvenilirliği bakımından oldukça önemli olduğu söylenebilir.

6- Dağılımlardan birisinin Normal (0,1), diğerinin ise t (10) olması durumunda, ele alınan bütün testlerin söz konusu dağılımların t (10):Beta (14,10) olması durumu ile hemen hemen aynı sonuçları gerçekleştirdikleri dikkati çekmektedir.

7- Söz konusu dağılımlardan birincisinin Normal (0,1), diğerinin ise Beta (14,10) olması halinde, en güvenilir sonuçları Bartlett testinin gerçekleştirdiği görülmüştür. Diğer yandan bu deneme koşullarında gerçekleşen 1.Tip hata olasılıklarının, aynı koşullarda dağılımların ikisinin de Normal (0,1) olması durumu ile oldukça yakın sonuçlar oldukları görülmüştür. Dolayısıyla dağılımların her ikisinin de Normal (0,1) olması ya da birinin Normal (0,1), diğerinin ise Beta (14,10) olmasının, Bartlett testini etkilemediği ileri sürülebilir. Gerçekleşen 1.Tip hata olasılıklarının güvenilirliği bakımından bu testi Levene testinin izlediği söylenebilir. Neyman-Pearson ve Bartlett 2 testlerinin ise özellikle küçük örnek hacimlerinden olumsuz yöne etkilendikleri, buna karşın örnek hacimlerinin artmasına paralel olarak %5.0'e yakın 1.Tip hata olasılıkları gerçekleştirme eğilimine girdikleri görülmüştür.

Kaynaklar

- Anonymous, 1994. FORTRAN Subroutines for Mathematical Applications. IMSL MATH/LIBRARY. Vol.1-2. Visual Numerics, Inc., Houston, USA.
- Bradley, J. C. 1978. Robustness. British Journal of Mathematical and Statistical Psychology, 31, 144-152.
- Brown, M. B. and A. B. Forsythe, 1974. The small sample behaviour of some statistics which test the equality of several means. Technometrics, 16, 129-132.
- Canover, W. J., M. E. Johnson and M. M. Johnson, 1981. A comparative study of tests for homogeneity of variances, with applications to outer continental shelf bidding data. Technometrics, 23, 351-361.
- Cochran, W. G. 1954. Some methods for strengthening the common χ^2 -tests. Biometrics, 10, 417-451.
- Ghost, M. and Y. H. Kim, 2001. The Behrens-Fisher problem revisited: A Bayes-frequentist synthesis. The Canadian Journal of Statistics, 29, No:1
- Krutchhoff, R. G. 1988. One-way fixed effects analysis of variance when the error variances may be unequal. J. Statist. Comput. Simul., 30, 259-71
- Ott, L. 1998. An Introduction to Statistical Methods and Data Analysis. Third Edition. PWS-Kent Publishing Company. USA, 835 s.
- Piepho., H. P. 1996. A Monte Carlo test for variance homogeneity in linear models. Biometrical Journal, 38,461-473.
- Sokal, R. R. and F. J. Rohlf, 1995. Biometry. The Principles and Practice of Statistics in Biological Research. Third Ed. W.H. Freeman and Co. New York, 887 s.
- Tomarken, A. J. and R. C. Serlin, 1986. Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. Psychological Bulletin, 99, 90-99
- Winer, B. J., D. R. Brown and K. M. Michels, 1991. Statistical Principles in Experimental Design. Third Ed., McGraw-Hill, Inc. USA, 1057 s.
- Wludyka, P. S. and P. R. Nelson, 1997. An analysis of means type test for variances from normal populations. Technometrics, 39, 274-286.
- Zar, J. H. 1999. Biostatistical Analysis. Fourth Ed., Prentice-Hall, Inc. USA, 683 s.
- Zimmerman, D. W. 1998. Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. J. of Experimental Education, 67, 55-69.

İletişim adresi:

Mehmet MENDEŞ

Ankara Üniv. Çankırı Orman Fakültesi-Çankırı

Tel: 0 542 476 00 27

E-mail: mehmetmendes@hotmail.com