A MONTE CARLO SIMULATION STUDY FOR COMPARING PERFORMANCES OF SOME HOMOGENEITY OF VARIANCES TESTS

Hamit MIRTAGIOĞLU
Bitlis Eren University, Faculty of Arts and Sciences, Department of Statistics, Bitlis-Turkey

E-mail: hamit_68@hotmail.com

Soner YIĞIT
Çanakkale Onsekiz Mart University, Agriculture Faculty, Animal Science Department, Biometry and Genetics Unit, Çanakkale-Turkey

E-mail: soneryigit@comu.edu.tr

Emel MENDEŞ
TOKI Anadolu Lisesi, Çanakkale-Turkey

E-mail: emelmendes@yahoo.com

Mehmet MENDEŞ
Çanakkale Onsekiz Mart University, Agriculture Faculty, Animal Science Department, Biometry and Genetics Unit, Çanakkale-Turkey

E-mail: mmendes@comu.edu.tr

Abstract

This simulation study has been carried out to compare empirical Type I error and test power of five tests namely Anom, Bartlett's, Levene's, Brown-Forsythe, and Conover tests to check homogeneity of variances. For this purpose, a comprehensive Monte Carlo Simulation study has been carried out for different number of groups (k = 3, 4, 5, and 10), variance ratios (i.e: 1, 5, 10, 15, and 20), and sample size combinations (equal and unequal sample sizes) under normality assumption. Based on results of 50,000 simulation, it is observed that the best robust tests are the Anom and Bartlett's even if studying with very small sample size (n = 5) and a large number of group cases (k = 10). They were followed by the Levene's and Conover tests in general. But, both the Conover and Levene's tests have been slightly negatively affected from increases in the number of groups when sample sizes were small (n ≤ 20). On the other hand, since the Brown-Forsythe test did not give satisfactory results for any of the experimental conditions, it was concluded that the use of this test should not be preferred for checking homogeneity of variance assumption. As a result, since the Anom and Bartlett's are robust tests for all experimental conditions, it is possible to propose to the researchers to use this test to check homogeneity of variances assumption prior to ANOVA and t-test.

Key words: Homogeneity of variances, Anom test, Bartlett's test, Levene's test, Conover's test, test power, Type I error, simulation

Introduction

Homogeneity of variances assumption is the major assumption of parametric tests like ANOVA and t-test (Armitage and Berry, 1994; Shuqiang, 1998; Zar, 1999; Mendeş, 2003; Hatchavanich, 2014; Wang et al., 2016). Since this assumption has a critical effect on reliability and validity of the results of ANOVA and t-test, it is very critical to check this assumption before perform the ANOVA or t-test. Unreliable results occur, when this assumption is not met or variance of one group is significantly larger than the others. Therefore, when ANOVA performs over the data sets in which this assumption is not met, in that case, the results of ANOVA will not be reliable (Cochran, 1947; Bodhisuwan, 1991; Srisunsanee, 1998; Mendeş, 2003; Vorapongsathorn et al., 2004). A previous simulation studies revealed that ANOVA and t-test are very sensitive to the heterogeneity of the variances, even when the data are normally distributed (Glass et al., 1972; Piepho, 1996, Ott 1998, Ghost and Kim 2001; Mendeş, 2003). Since any degree of heterogeneity may cause not to be retain Type I error rate at the nominal level and cause negative changes in test power estimates, it is very important to determine which degree of the heterogeneity has significant effect on Type I error and test power estimates. That way, it will be possible to obtain more detailed information about the performances of homogeneity of variances tests.

Although many statistical tests and approaches have been proposed in the literature to check the homogeneity of variance assumption, it is noticed that especially Levene and Bartlett tests are most commonly used in practice (Gartside, 1972; Veitch and Roscoe, 1974; Phil, 1999; Zar, 1999; Mendeş, 2003). There are different reasons for why many of these tests are not known / used by the researchers. For example, many of those tests have been revealed to be very sensitive to non-normality, only a few of them are presented in the current statistical textbooks, and majority of them have not been included yet in the commonly used statistical software's such as Minitab, SPSS, SAS etc. Therefore, the main questions addressed in this study are:

(a) What level of heterogeneity in the variances can cause a serious problem in terms of reliability of the ANOVA results, and;

(b) Which of Anom, Bartlett's, Levene, Brown-Forsythe, and Conover tests is more appropriate for checking homogeneity of variances assumption.

Although many different simulation studies had been carried out in the past to compare performances of different homogeneity of variances tests, small number of simulation (in general 10,000 or lesser), a few sample size combinations (three or four combination), and a few variance ratios (one or two) had been considered in these studies. And many of the tests considered are not included in the textbooks and statistical packages in general. At the same time, since it is well known that many of the homogeneity of variances tests are very sensitive to departures from the normality, it will be more beneficial to focus on investigating performances of these tests for normality rather than non-normality under many different group sizes, sample sizes, and variance ratios. From this point of view, in this paper, a comprehensive Monte Carlo Simulation study has been carried out to answer above questions under normality.

2. Material and Methods

2.1. Methodology

Random numbers generated from normal distributed populations for the experimental conditions given in Table 1 by using RNNOA library function of Microsoft Fortran Developer Studio's. Two performance criteria namely Type I error rate and test power were used to determine appropriate test(s) and level of heterogeneity in the variances which may cause a serious problem in terms of reliability of the ANOVA or t-tests results. Then the Anom, Bartlett's, Levene's, Brown-Forsythe, and Conover test values have been computed over the random numbers generated for each simulation and these values have been compared with the critical values of these tests. This process has been done 50,000 times and the values that rejected null hypothesis have been counted.

A statistical test has been accepted as a robust, if deviations of the empirical Type I error estimates of that test from the nominal level of significance does not exceed the predetermined alpha level (0.05) and has at least 80% test power. Evaluation of the robustness has been done based on the Cochran criteria as: if empirical Type I error estimates are between 0.04-0.06 at 0.05 significance level then that test has been accepted as a robust in terms of Type I error rates. In *case* of test power, if empirical test power estimates are equal or greater than 80.0% then that test has been accepted as a robust in terms of test power.

Empirical Type I error and test power estimates have been obtained as below:

$$\alpha^* = \frac{\text{The number of Ho rejection when Ho is actually true}}{\text{The number of simulation runs (50,000)}}$$

$$\beta^* = \frac{\text{The number of Ho rejection when Ho is actually true}}{\text{The number of simulation runs (50,000)}}$$

Test Power $=1-\beta^*$

Table 1. Experimental conditions for simulation study

Number of Group	3	4	5	10
Variance ratio	1:1:1 1:1:8 1:1:12 1:1:20	1:1:1:1 1:1:1:8 1:1:1:12 1:1:1:20	1:1:1:1:1 1:1:1:1:8 1:1:1.1:12 1:1:1:1:20	1:1:1:……:1 1:1:1:…….:8 1:1:1:……:12 1:1:1:……:20
Sample Size (equal)	5,10,15,20,25,30 35,40,45,50	5,10,15,20,25,30 35,40,45,50	5,10,15,20,25,30 35,40,45,50	5,10,15,20,25,30 35,40,45,50
Sample Size (unequal)	5:10:15 10:20:30	5:10:15:20 10:20:30:40	5:10:15:20:25 10:20:30:40:50	5:10:15:…….:50 10:20:30:….:100

2.2. Statistical Tests

Since it is well known that majority of the homogeneity of variances tests are extremely sensitive to departures from normality that is why we have focused on the performances of those tests under normality.

2.2.1. Bartlett's Test

This test has been proposed by Bartlett (1937) to test homogeneity of variances. Since the Bartlett's test assumes that all populations are normally distributed, it is not recommended when the normality assumption is not met (Snedecor and Cochran, 1989; Armitage and Berry, 1994; Zar, 1999; Mendeş, 2003).

$$B = \frac{(N-k)\log\left(S_p^2\right) - \sum_{i=1}^{k}(n_i - 1)\log\left(S_i^2\right)}{1 + \frac{1}{3k-1}\left(\sum_{i=1}^{k}\frac{1}{n_i - 1}\right) - \frac{1}{(N-k)}}$$

Where

k is the number of groups,

S_i^2 *is* the variance of the i[th] group

S_p^2 is the pooled variance

If the assumptions are met, the B-test statistic follows the Chi-Squared distribution with *k-1* d.f.

2.2.2. Levene's Test (LV)

This test was proposed by Levene (1960) to test if *k* samples have equal variances. The Levene's test is an alternative to the Bartlett's test. The Levene's test is less sensitive than the Bartlett test to departures from normality. If the authors have strong evidence that their data sets do in fact come from a normal, or nearly normal distribution, then Bartlett's test has better performance.

Test statistics for the Levene's test is computed as:

$$LV = \frac{(N-k)\sum_{i=1}^{k} n_i(\bar{Z}_{i.} - \bar{Z}_{..})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - \bar{Z}_{i.})^2}$$

Where

$Z_{ij} = \left|Y_{ij} - \bar{Y}_{i.}\right|$

Where $\bar{Y}_{i.}$ is the mean of the i^{th} group.

2.2.3. Brown-Forsythe Test (BF Test)

This test was proposed by Brown and Forsythe (1974). This test that follows the idea of the Levene's test but it uses the group median instead of the group mean in the calculation of the absolute residual values. It is expected to be more robust than the Levene's test when the population distribution is skewed.

2.2.4. ANOMV Test

This test is proposed by Wludyka and Nelson (1997) to test homogeneity of variances. This test is based on upper (UDL) and lower (LDL) decision lines. This test requires equal number of observations in each group. The UDL and LDL are computed as below:

$UDS = U.k.S_p^2$ and $LDL = L.k.S_p^2$

Where:

A and *L* denote ANOMV upper and lower table values which were determined by α, *k*, and *v*

S_p^2 denotes the pooled variance

2.2.5. Conover Test

This test has been proposed by Conover (1999) as a nonparametric test of homogeneity of variances based on ranks. Since this test does not assume that all populations are normally distributed, it is recommended especially when normality assumption is not met. The test assumes that the data are obtained by taking a simple random sample from each of the *k* populations.

Conover test statistic is computed as below:

$$T = \frac{1}{D^2}\left[\sum_{i=1}^{k}\frac{S_i^2}{n_i} - N\bar{S}^2\right]$$

Where

$$S_i = \sum_{i=1}^{n_i} R_{ij}^2$$

$$R_{ij} = Rank\ (Z_{ij})$$

$$Z_{ij} = \left|Y_{ij} - \bar{Y}_j\right|$$

$\bar{S} = \frac{1}{N}\sum_{i=1}^{k} S_i$ and

$$D^2 = \frac{1}{\mathrm{N}-1}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} R_{ij}^4 - N\bar{S}^2\right)$$

If the assumptions are met, the distribution of this test statistic follows approximately the Chi-Squared distribution with *k-1* d.f.

3. Results

3.1. Empirical Type I Error Estimates

Figure 1-4 have been established to present Type I error estimates of Bartlett's, Levene, Brown-Forsythe, Conover, and Anom tests, respectively.

Simulation results showed that the Type I error estimates of all tests, except the Anom and Bartlett's tests, have generally been affected by the sample size and the number of groups (Figure 1-4). As it is seen from Table 1-4, the Anom and Bartlett's are the best tests in terms of retaining the Type I error rate at the nominal alpha level (0.05) regardless of number of groups, sample size combinations, and variation ratios. The Type I error estimates of the Anom test are very close to 0.05 while the Type I error estimates of the Bartlett's test varied from 4.6 to 5.6% even with unequal sample sizes. Levene and Conover tests generally retained Type I error at the nominal alpha level when the number of observations in each group were greater than 30. It is noticed that both Levene and Conover tests are negatively affected from increases in the number of group when sample size was small ($n \leq 20$). For example, in case of the sample sizes of $n \leq 20$, the Type I error estimates of Levene and Conover tests have been estimated between 5.6-7.9% and 6.2-7.6% for k = 3, 5.8-8.3% and 6.1-7.3 % for k = 4, 6.1-9.2 % and 6.1-7.3 % for k = 5 and, 7.1-14.1% and 6.8-7.6 % for k = 10. However, except sample sizes of 5 and 10, the differences between Type I error estimates of these tests are not obvious. Generally, most deviated estimations in terms of protecting Type I error rates at the 0.05 alpha level have been made when Brown-

Forsythe test is used for checking homogeneity of variances assumption. This test could not have kept the Type I error rate at 0.05 level for none of experimental condition. The Type I error of this test were estimated around 2.5-6.5 % in general. The Type I error estimates of this test for k = 3, 4, 5, and 10 changed between 3.1-6.4 %, 2.7-6.2 %, 2.5-6.4 %, and 2.5-6.2 % respectively. As it can be seen from the Type I error estimates, the Levene's test the one which is relatively the most affected test from the increasing in the number of groups when studying with small sample sizes (n < 20). As sample size is increased, then the Type I error estimates of all tests, except for Brown-Forsythe, have begun to approach to 0.05 level and thus, the effect of sample size and number of groups being compared on Type I error estimates decreased to ignorable level.

In case of unequal sample size, he best estimates have been made when Bartlett's test is used. The Type I error estimates of this test are always around 5.0% level regardless imbalance in the sample sizes. The Type I error estimates of this test varied between 4.6 and 5.6% in general. Anom test is the one that the most affected from the increases in the number of groups when sample sizes are not equal. This test could not keep Type I error rate at the nominal level under unequal sample size combinations. The Type I error estimates of this test estimated between 11.4% and 16.5% in general. The other tests gave similar Type I error estimates to the equal sample sizes conditions in general.

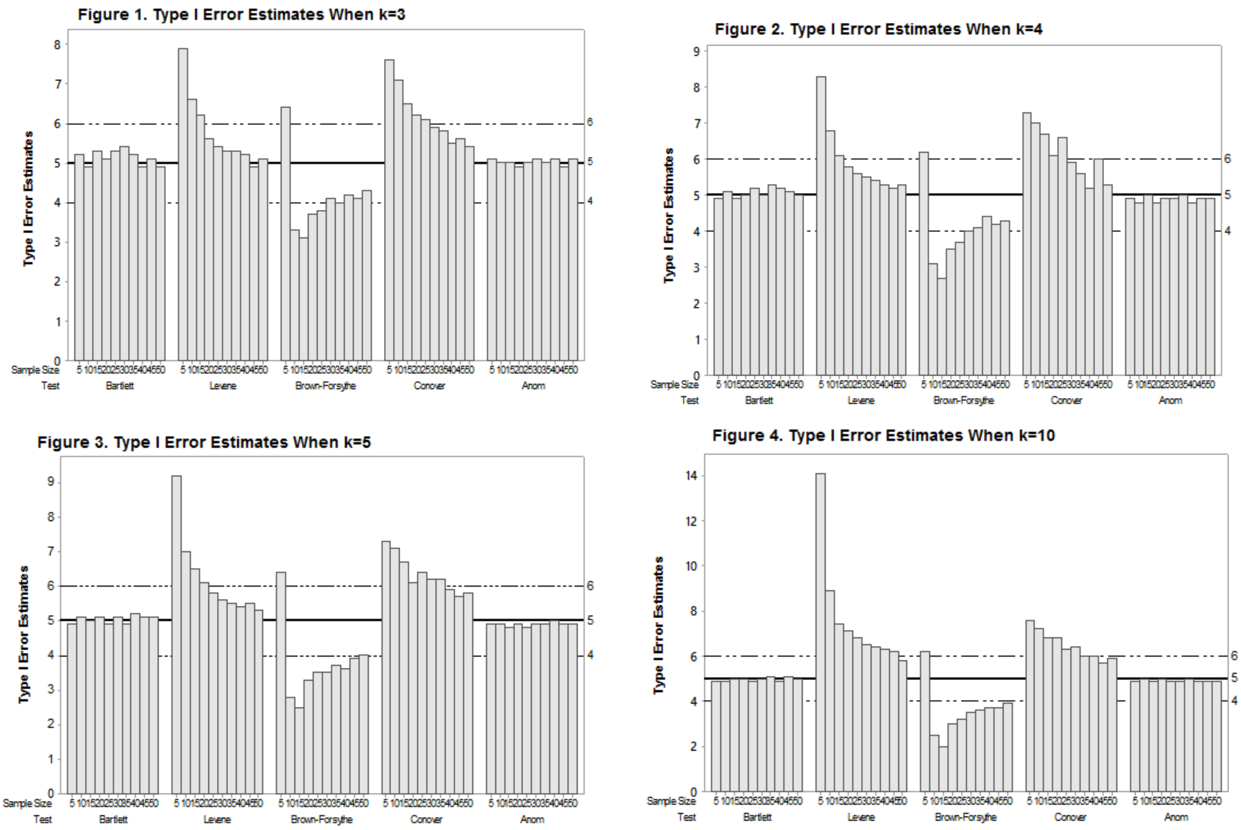


Figure 1-4. Type I Error Estimates

3.2. Empirical Test Power Estimates

Figure 5-21 have been established to present test power estimates of Bartlett's, Levene, Brown-Forsythe, Conover, and Anom tests, respectively. Simulation results showed that the most powerful test is the Anom and it has been followed by the Bartlett and Levene tests in general.

Figure 5. Test Power Estimates, k=3, Variance Ratios are 1:1:8

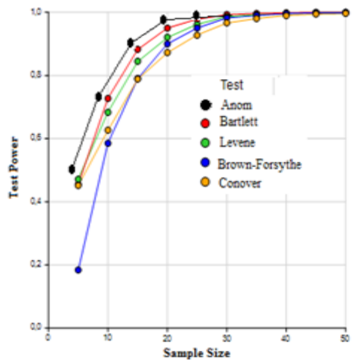


Figure 6. Test Power Estimates, k=3, Variance Ratios are 1:1:12

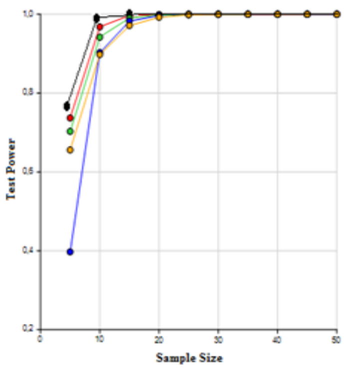


Figure 7. Test Power Estimates, k=4, Variance ratios are 1:1:20

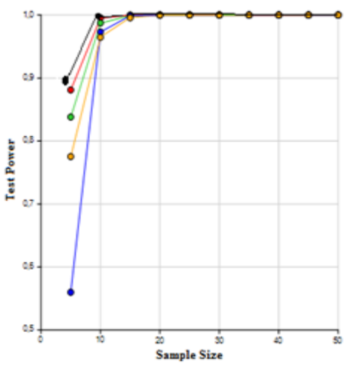


Figure 8. Test Power Estimates, k=4, Variance Ratios are 1:1:1:8

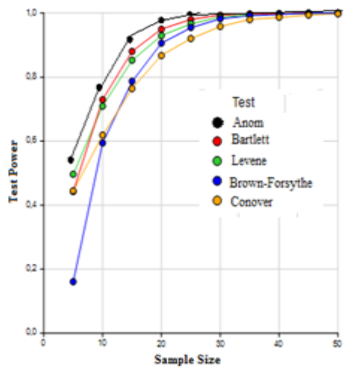


Figure 9. Test Power Estimates, k=4, Variance Ratios are 1:1:1:12

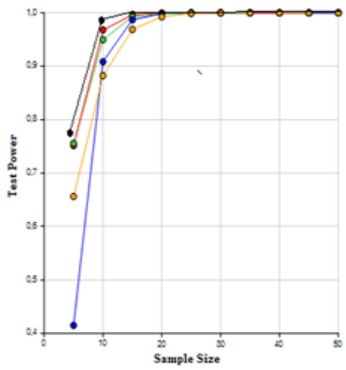


Figure 10. Test Power Estimates, k=4, Variance Ratios are 1:1:1:20

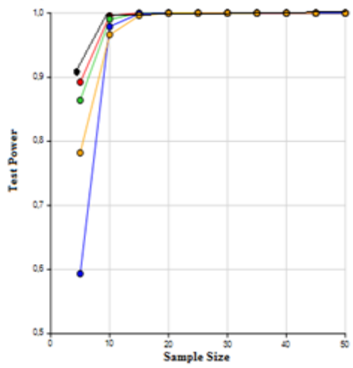


Figure 11. Test Power Estimates, k=5, Variance Ratios are 1:1:1:1:8

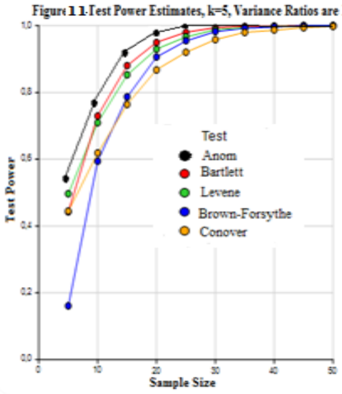


Figure 12. Test Power Estimates, k=5, Variance Ratios are 1:1:1:1:12

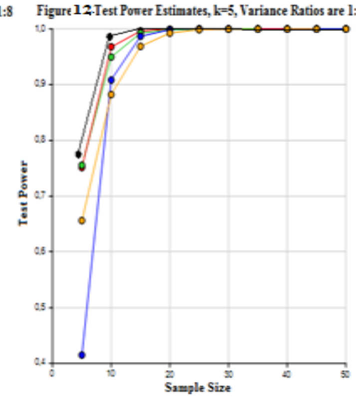
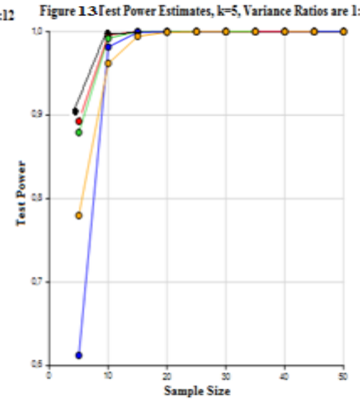


Figure 13. Test Power Estimates, k=5, Variance Ratios are 1:1:1:1:20



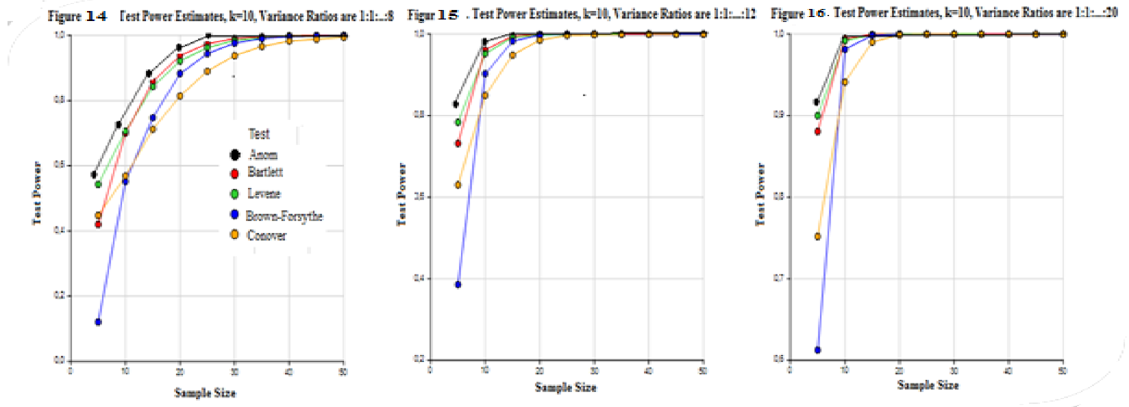


Figure 5-16. Test power curves of Anom, Bartlett, Levene, Conover, and Brown-Forsythe tests

As it can be noticed from the test power graphs (Figure 5-16), the sample size, especially total sample size, is the major factor that affects test power estimates. The second factor influencing the test power estimates is the variance ratio. As it is expected, as the variance ratio is increased, the test power estimates are increased as well. All test power estimates have reached to 100% for large sample sizes. In case of unequal sample sizes, test power estimates have been affected by the total number of observations rather than number of observations in each group. For example, the test power estimates of the Anom, Bartlett's, Levene, Brown-Forsythe, and Conover tests have been estimated as 73%, 75%, 65%, 58%, and 92% respectively for the sample sizes of 5:10:15 (total 30 observations) when $k = 3$. The test power values of these tests have been estimated as 86%, 90%, 83%, 92%, and 94% respectively for the sample size combination of 10:20:30 (total 60 observations). As it can be noticed, there is a clear increase in the test power estimates of all tests when the total number of observations is high.

As it is seen from Figures 5-16, as the sample size and variance ratios are increased, the test power estimates of all tests are increased as well. Test power estimates of all tests 100% when sample size and variance ratios are large. Simulation results showed that the most powerful test is the Anom and it has been followed by the Bartlett, Levene, and Conover tests in general. The Brown-Forsythe, on the other hand, is the weakest test in terms of test power. This test is also one that affected negatively by increases in the number of groups being compared.

4. Discussion

Since homogeneity of variances assumption is major assumption of parametric tests, determining performances of commonly used tests in practice to check homogeneity of variances assumption under different experimental conditions is an important issue (Bartlett, 1937; Levene, 1960; Brown and Forsythe, 1974; Wludyka and Nelson, 1997; Conover, 1999; Mendeş, 2003; Vorapongsathorn et al., 2004; Hatchavanich, 2014; Wang et al., 2016). It is because that way, it will be possible both to determine more suitable test(s) and to reveal the effect of level of heterogeneity on performance of these tests.

Several simulation studies have been performed to compare performances of different homogeneity of variances tests under different experimental conditions (Gartside, 1972; Vietch and Roscoe, 1974; Conover et al., 1981; Lim and Loh, 1996; Shuqiang, 1998; Arsham and Lovric, 2001; Mendeş, 2003; Boos and Brownie, 2004; Vorapongsathorn et al., 2004; Bhandary, and Dai, 2009; Katz et al., 2009; Lee et al., 2010; Parra-Frutos, 2012; Sharma and Kibria, 2013; Nguyen et al., 2014; Hatchavanich, 2014; Wang et al., 2016). Previous simulation studies have shown that small deviations from the assumption of equal variances do not seriously affect the results in the ANOVA. Therefore, the ANOVA is robust to small deviations from the homogeneity of variances assumption. That is why we concerned about large deviations from the homogeneity of variances assumption in this study. At the end of this simulation study it is observed that the Anom and Bartlett's tests are the best choices for checking homogeneity of variances assumption. Although there are some differences between the results of this study and the previous studies due to differences in the experimental conditions, findings of this study are generally consistent with the findings of previous studies. For example, Arsham and Lovric (2001) reported that the Bartlett's test is known to be powerful if the underlying populations are normal. According to some recent results based on simulation, Bartlett's test and Box's test are the best overall methods to test the homogeneity of variances. Mendeş (2003) in his simulation study he compared performances of four different tests namely Bartlett's, Levene, F-Max and Cochran test under different experimental conditions. He reported that Bartlett's test and Levene tests were superior tests compared to the others. Vorapongsathorn et al. (2004) compared Bartlett's, Levene, and Cochran test in terms of Type I error rates and test power. It was reported that although Bartlett's test was sensitive to deviation from the normality in terms of keeping Type I error rate at the nominal alpha level, it had the highest test power in all experimental conditions. Cochran's and Levene's tests were not. It was also reported that the Levene's test had very good performance for both equal and small sample sizes conditions. Cochran's test was the best one especially variance ratio was large. Hatchavanich (2014) compared Bartlett's, Levene's and O'Brien's tests with respected to Type I error and test power. It was reported that the Levene's test was not the best option. Since the Bartlett's test is not affected by sample size, this is one of the best tests to check homogeneity of variances assumption. Likewise, Wang et al. (2016) performed a simulation study to evaluate the performance of 14 tests for the homogeneity of variance assumption in one-way ANOVA models in terms of Type I error control and statistical power. They reported that the Ramsey conditional, O'Brien, Brown–Forsythe, Bootstrap Brown–Forsythe, and Levene with squared deviations tests maintained adequate Type I error control, performing better than the others across all the conditions.

5. Conclusion

Anom and Bartlett's tests are the uniformly most robust tests the homogeneity of variances problem under the assumption that each treatment population is normally distributed regardless of sample size and number of group to be compared. As a conclusion, it is strongly suggested to the authors and researchers to use either the Anom or Bartlett's tests for checking homogeneity of variances assumption.

References

1. Armitage, P. and Berry, G.D. **Statistical Methods in Medical Research.** Third Edition. Wiley-Blackwell, Lincoln, UK, 1994
2. Bartlett, M. S. **Properties of sufficiency and statistical tests.** Journal of the Royal Statistical Society, Series A, vol. 160, 1937, pp. 268-282.
3. Bhandary, M. and Dai, H. **An alternative test for the equality of variances for several populations when the underlying distributions are normal.** Communications in Statistics: Simulation and Computation, vol. 38, 2009, pp. 109-117.
4. Bodhisuwan, W. **A comparison of the test statistics for homogeneity of variances.** Master Thesis. Statistics. Bangkok: Faculty of Graduate Studies, Chulalongkorn University, 1991
5. Boos, D. D. and Brownie, C. **Comparing variances and other measures of dispersion.** Statistical Science, vol. 19, 2004, pp. 571-578.
6. Brown, M. B. and Forsythe, A. B. **Robust tests for the equality of variances.** Journal of tile American Statistical Association, vol. 69, 1974, pp. 364-367.
7. Cochran, W.G. **Some consequences when the assumptions for the analysis of variance are not satisfied.** Biometics, vol. 3, 1947, pp. 22-38.
8. Conover, W. J., Johnson, M. E. and Johnson, M. M. **A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data.** Technometrics, vol. 23, 1981, pp. 351-361.
9. Conover, W. J. **Practical Nonparametric Statistics,** Third Edition, New York: John Wiley & Sons, 1999
10. Gartside, P.S. **A study of methods for comparing several variances.** JASA, vol. 67, no. 338, 1982, pp. 342-346.
11. Ghost, M. and Kim, Y. H. **The Behrens-Fisher problem revisited: A Bayes-frequentist synthesis.** The Canadian Journal of Statistics, vol. 29, 2001, p. 1
12. Hatchavanich, D. **A Comparison of Type I Error and Power of Bartlett's Test, Levene's Test and O'Brien's Test for Homogeneity of Variance Tests.** Southeast-Asian J. of Sciences, vol. 3, no. 2, 2014, pp. 181-194.
13. Katz, G. S., Restori, A. F. and Lee, H. B. **A Monte Carlo study comparing the Levene test to other homogeneity of variance tests.** North American Journal of Psychology, vol. 11, 2009, pp. 511-522.
14. Lee, H. B., Katz, G. S. and Restori, A. F. **A Monte Carlo study of seven homogeneity of variance tests.** Journal of Mathematics and Statistics, vol .6, 2010, pp. 359-366.
15. Levene, H. **Robust tests for equality of variances to Probability and Statistics,** in Olkin, I. (ed.), "Contributions Palo Alto", CA: Stanford University Press, 1960
16. Lim, T. S. and Loh, W. Y. **A comparison of tests of equality of variances.** Computational Statistics and Data Analysis, vol. 22, 1996, pp. 287-301.
17. Mendeş, M. **The Comparisons of Levene, Bartlett, Neyman-Pearson and Bartlett 2 Tests in Terms of Actual Type I Error Rates.** Tarım Bilimleri Dergisi, vol. 9, no. 2, 2003, pp. 143-146.
18. Nguyen, D. T., Pham, T. V., Rodriguez de Gil, P., Hicks, T., Wang, Y., Li, I., and Kromrey, J.D. **ANOVA_HOV: A SAS macro for testing homogeneity of variance in one factor ANOVA models.** In "South East SAS Users Group 2014 Proceedings". Cary, NC: SAS Institute, 2014

19. Ott, L. **An Introduction to Statistical Methods and Data Analysis.** Third Edition. PWS-Kent Publishing Company. USA, 1998
20. Parra-Frutos, I. **Testing homogeneity of variances with unequal sample sizes.** Computational Statistics, vol. 28, 2012, pp. 1269-1297.
21. Phil, E. **Checking assumptions. Education 230 B/C Linear Statistical Models.** [Online]. Graduate School of Education & Information Studies; 1999. Available from: http://www. gseis.ucla.edu/courses/ed230bc1/cnotes1
22. Piepho, H. P. **A Monte Carlo test for variance homogeneity in linear models.** Biometrical Journal, vol. 38, 1996, pp. 461-473.
23. Sharma, D. and Kibria, B. M. G. **On some test statistics for testing homogeneity of variances: A comparative study.** Journal of Statistical Computation and Simulation, vol. 83, 2013, pp. 1944-1963.
24. Shuqiang, Z. **Fourteen Homogeneity of Variance Tests: When and How To Use Them.** Annual Meeting of the American Educational Research Association, San Diego, California, April 13-17, 1998
25. Srisunsanee, J. **Effects of failure to meet assumptions of homogeneity of variance and normal on Type I error in one-way analysis of variance.** Master Thesis in statistics]. Bangkok: Faculty of Graduate Studies, Kasetsart University, 1998
26. Snedecor, G. W. and Cochran, W. G. **Statistical methods** (8th ed.). Ames: Iowa State University Press, 1989
27. Wang, Y., de Gil, P.R., Chen, Y-H.C., Kromrey, J.D., Kim, E.S., Pham, T., Nguyen, D. and Romano, J.L. **Comparing the Performance of Approaches for Testing the Homogeneity of Variance Assumption in One-Factor ANOVA Models.** Educational and Psychological Measurement, 2016, pp. 1-25.
28. Wludyka, P. S. and Nelson, P. R. **An analysis-of-means-type test for variance from normal population.** Technometrics, vol. 39, 1997, pp. 274–285.
29. Vietch, W.R. and Roscoe, J.T. **Homogeneity of Variance: An Empirical Comparison of 4 Statistical Tests.** The Journal of Experimental Education, vol. 43, no. 2, 1974, pp. 73-78.
30. Vorapongsathorn, T., Taejaroenkul, S. and Viwatwongkasem, C. **A comparison of type I error and power of Bartlett's test, Levene's test and Cochran's test under violation of assumptions.** J.Sci.Tech., vol. 26, no. 4, 2004, pp. 537-547.
31. Zar, J. H. **Biostatistical analysis** (Fourth Edition). Upper Saddle River, N.J.: Prentice Hall, 1999