Journal of Statistical Research *ISSN 0256 - 422 X*
2006, Vol. 40, No. 2, pp. 65-83
Bangladesh

A NEW ALTERNATIVE IN TESTING FOR HOMOGENEITY OF VARIANCES

Mehmet MENDEŞ
Departmanet of Animal Science, University of Canakkale Onsekiz Mart,
Faculty of Agriculture, 17020 Canakkale-Turkey
Email: mmendes@comu.edu.tr

Necla ÖZKAYA TURHAN
Department of Statistics, University of Ankara,
Faculty of Sciences, 06100 Tandogan/Ankara-Turkey
Email: nturhan@science.ankara.edu.tr

Fikret GÜRBÜZ
Departmanet of Animal Science, University of Ankara,
Faculty of Agriculture, 06100 Tandogan/Ankara-Turkey
Email: fikretgurbuz@isbank.net.tr

SUMMARY

A new alternative procedure is proposed for testing the hypothesis that the variances of k independent groups are equal under non-normality. Extensive simulations indicated that the new procedure always gave the experimenter more control over the probability of a Type I error than do the Bartlett χ^2 and Levene's tests. The same statement was true for unequal sample sizes. Overall, the new procedure was more powerful than the other procedures tested, especially when the variance ratios were high.

Keywords and phrases: homogeneity of variance, Levene's test, Bartlett test, power of test, Type I error rate, normal distribution

1 Introduction

Analysis of variance techniques has been used in variety of fields such as agriculture, medicine and sociology to compare independent group means. However, a set of assumptions, such as normal distribution, homogeneity of variances, and independent observations has to be met for the method to work properly. Studies of robustness of the ANOVA F test have shown that the violation of normality has little effect on inferences about the means, especially if there is sufficient number of observations in the groups, and the data are balanced. However, violation of equality of variances can have a serious effect on inferences about the means, especially if sample sizes are unequal (Bryk and Raudenbush, 1987; Wilcox, 1989; Schneider and Penfield 1997; Chen and Chen, 1998; Luh and Guo, 2000; Mendeş, 2003; Çamdeviren and Mendeş, 2005). Therefore, homogeneity of variances is the most influential assumption on sensitivity of an analysis of variance F test. It is vital to correctly test for homogeneity of variances when analysis of variance is to be performed. Preliminary to an ANOVA test, it is useful to check the equality of variances assumption using appropriate tests (Brown and Forsythe, 1974; Wilcox et al., 1986). When this assumption is violated, the results of the analysis may be untrustworthy, namely that the reported p-value from the significance test may be too liberal or too conservative. Therefore, before performing ANOVA, homogeneity of variance assumption must be checked to increase the reliability of the results. There are variety of methods to test for homogeneity of variances for different situations (Conover et al., 1981; Lim and Loh, 1996; Nelson, 2002; Wilcox, 2002). Most common tests for the one-way ANOVA are Bartlett's and Levene tests. Unfortunately, these tests are very sensitive to departure from normality (Cochran and Cox, 1957; Levene, 1960; Weerahandi, 1995; Zar, 1999). Several parametric alternative tests to ANOVA have been proposed for testing the equality of k ($k \geq 2$) population means when population variances are not equal (Wilcox, 1988; Alexander and Govern, 1994). However, these alternatives are affected adversely by non-normality (Hsuing and Olejnik, 1996). As Wilcox (1995) pointed out that none of the procedures directly handle the problem of skewness bias. Oshima and Algina (1992) also argued that failure to consider the impact of the combined violations of variance equality and distribution normality is an important omission of a statistical procedure. Therefore, developing new alternative tests, such as trimming, transforming statistics, bootstrapping (Wei-ming, 1999; Keselman et al., 2002) to deal with unequal variances, and non-normality is worthwhile. Many of these tests do not take place in the statistical packages commonly used by researchers in practice, such as SPSS, STATISTICA, NCSS, SYSTAT, SAS, and MINITAB.

A simulation study (Conover et. al, 1981) was carried out to assess the robustness and power of 56 procedures for testing the equality of variances. Distribution of data, number of groups and the number of observations in these groups affect the results of these tests. The fact that all tests are sensitive to departure from normality (Bishop and Dudewicz, 1978; Winer et al., 1991; Piepho, 1997) suggests that a new test, which adapts to variety of distributions, is required. In this study, a new approximation test, which is adaptable to different kinds of distributions ($\chi^2(3)$, $\beta(6, 1.5)$, $\omega(1.5, 1)$, and $t(5)$ distributions) as is proposed.

1.1 Definition of Statistical Tests

The model for one-way analysis of variance is $Y_{ij} = \mu + \alpha_j + e_{ij}$ where Y_{ij} ($i = 1, \ldots, n_j$; $j = 1, \ldots, k$) denotes the measure for the ith subject in the jth group, μ is population mean, α_j is the effect of the jth treatment ($j = 1, \ldots, k$), and e_{ij} random error term. The goal is to test the following hypothesis: $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ and $H_1 : \sigma_i^2 \neq \sigma_k^2$ for at least one pair (i, k).

1.1.1 Bartlett χ^2 Test (Bart)

The most common method employed to test for homogeneity of variance is Bartlett's χ^2 test. In this procedure, the test statistic is;

$$B = \ln S_p^2 \left(\sum v_j\right) - \sum v_j \ln S_j^2$$

where $v_j = n_j - 1$ and n_j is the size of sampled j. S_p^2 is pooled variance and S_j^2 is the variance of sampled j. Distribution of B statistic is approximated by the $\chi^2(k-1)$ distribution, but a more accurate chi-square approximation is obtained by computing a correction factor,

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

with the corrected test statistics being

$$B_C = \frac{B}{C}$$

(Conover et al. 1981; Lim and Loh, 1996; Ott, 1998; Zar, 1999).

1.1.2 Levene 1 Test (Lev1)

A procedure suggested by Levene (1960) defines new variables,

$$X_{ij}^1 = \left|Y_{ij} - \overline{Y}_{i.}\right|$$

where Y_{ij} is the original variable, X_{ij}^1 is the transformed or new variable and $\bar{Y}_{i.}$ is the mean of the ith group. An ANOVA procedure is then carried out using X_{ij}^1 in place of Y_{ij}. Levene 1 test statistic is defined as:

$$\mathbf{Lev1} = \frac{\sum\limits_{i=1}^{k} n_i \left(\bar{X}_{i.} - \bar{X}_{..}\right)^2 / (k-1)}{\sum\limits_{i=1}^{k} \sum\limits_{j=1}^{n_i} \left(X_{ij} - \bar{X}_{i.}\right)^2 / (N-k)} \quad , \quad \left(N = \sum_{i=1}^{k} n_i\right) \tag{1.1}$$

which has the same distribution as the F-statistic with $(k-1)$ and $(N-k)$ degrees of freedom where $\bar{X}_{i.}$ is the mean of the ith group, $\bar{X}_{..}$ is the overall mean, k is the number of group, n_i is the number of observations in the ith group, and N is the number of total observations.

The Levene 1 test rejects the null hypothesis if $\mathbf{Lev1} > F_{(\alpha, k-1,\, N-k)}$. Then researcher will adopt a test instead of ANOVA for mean equality that is robust to heterogeneity of variances, such as the Welch test, James second-order test, Alexander-Govern's test (Conover et al, 1981; Boos and Brownie, 1989).

1.1.3 Levene 2 Test (Lev2)

Levene 2 test is defined as: The original variables Y_{ij}, transforms to a new variable,

$$X_{ij}^2, \text{ as } X_{ij}^2 = (Y_{ij} - \overline{Y}_{i.})^2$$

An ANOVA procedure (1.1) is then carried out using X_{ij}^2 in place of Y_{ij}. The Levene 2 test rejects the null hypothesis if **Lev2** $> F_{(\alpha,k-1,\ N-k)}$.

1.1.4 Levene 3 Test (Lev3)

Levene 3 test is defined as: The original variablesY_{ij}, transforms to a new variable,

$$X_{ij}^3, \text{ as } X_{ij}^3 = \ln(Y_{ij} - \overline{Y}_{i.})^2$$

An ANOVA procedure (1.1) is then carried out using X_{ij}^3 in place of the Y_{ij}. The Levene 3 test rejects the null hypothesis if **Lev3** $> F_{(\alpha,k-1,\ N-k)}$.

1.1.5 Levene 4 Test (Lev4)

Levene 4 test is defined as: The original variablesY_{ij}, transforms to a new variable,

$$X_{ij}^4, \text{ as } X_{ij}^4 = \sqrt{|Y_{ij} - \overline{Y}_{i.}|}$$

An ANOVA procedure (1.1) is then carried out using X_{ij}^4 in place of the Y_{ij}. The Levene 4 test rejects the null hypothesis if **Lev4** $> F_{(\alpha,k-1,\ N-k)}$.

1.1.6 Proposed Test (MP Test) (MP)

Almost all variance homogeneity tests are affected adversely by deviations from normality (Conover et al., 1981; Weerahandi, 1995; Zar, 1999). We adopt to eliminate this handicap by developing a new procedure. Let Z_{1j} be the minimum observation in group n_j and Z_{2j} be the maximum observation in the same group. Then we transform Z_{1j} and Z_{2j} to a new observation Z_{ij} and $Z_{ij} = \dfrac{Z_{2j} - Z_{1j}}{2}$.

Z_{ij} is designed to mitigate the effect of extreme observations in the tails of a distribution. $Z_{ij}$is then added to the group n$_j$. Thus, the group is composed of $n_j - 1$ observations. There are n_j observations in group j. When the observations with maximum and minimum values are removed, the number of observations is $n_j - 2$ in the group j. When the new observation calculated as $Z_{ij} = \dfrac{Z_{2j} - Z_{1j}}{2}$ is added to the group j, the number of observations increases to $n_j - 2 + 1 = n_j - 1$. Mean of the group j is given as: $\tilde{X}_j = \dfrac{\sum X_{ij}^5}{n_j - 1}$.

Where $\tilde{X}_j$ is the mean of the jth group, which is composed of $n_j - 1$ observations. X_{ij}^5 is the ith observation in jth group. Then, the original variable, X_{ij}^5, transforms to a new variable, Y_{ij}, as $Y_{ij} = \left| X^5_{ij} - (\tilde{X}_j)^2 \right|$ gives the new observations for each group, where Y_{ij} is the new variable. An ANOVA procedure is then carried out using Y_{ij} in place of X^5_{ij}. Therefore, **MP** test statistics defined as;

$$\textbf{Lev1} = \frac{\sum\limits_{i=1}^{k} n_i \left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2 / (k-1)}{\sum\limits_{i=1}^{k} \sum\limits_{j=1}^{n_i-1} \left(Y_{ij} - \bar{Y}_{i.}\right)^2 / (N-2k)} \tag{1.2}$$

1.1.7 Sampling Distribution of MP Test

The numerator of **MP** test statistics is the same as the Levene 1. Therefore, the numerator degrees of freedom for the **MP** and the **Lev1** tests are equal and under the null hypothesis, the numerator of **MP** test statistic is distributed asymptotically as a central χ^2-variable with $k-1$ degrees of freedom. That is,

$$\sum_{i=1}^{k} \left(\frac{\bar{Y}_{i.} - \bar{Y}_{..}}{\sigma/\sqrt{n_i}} \right)^2 = \sum_{i=1}^{k} \frac{n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2}{\sigma^2} \approx \chi^2_{(k-1)} \tag{1.3}$$

The **MP** test used $N-2k$ denominator degrees of freedom while Levene's test (1960) used $N-k$ denominator degrees of freedom. The denominator of **MP** test statistic is distributed asymptotically as a central χ^2-variable with $N-2k$ degrees of freedom. Because;

Let Y_{ij} be the jth observation in the ith group $(i = 1, \ldots, k)$, $(j = 1, \ldots, n_i - 1)$ and expected values μ_i and variance σ_i^2. Let $t_i = n_i - 1$. The best linear unbiased estimates of σ_i^2 are

$$S_i^2 = \frac{1}{t_i - 1} \sum_{j=1}^{t_i} (Y_{ij} - \bar{Y}_{i.})^2 \tag{1.4}$$

Sum of squares for error is

$$SS_E = \sum_{i=1}^{k} \sum_{j=1}^{t_i} (Y_{ij} - \bar{Y}_{i.})^2 = \sum_{i=1}^{k} \left[\sum_{j=1}^{t_i} (Y_{ij} - \bar{Y}_{i.})^2 \right]$$

Then we can calculate sample variance as

$$S_i^2 = \frac{1}{t_i - 1} \sum_{j=1}^{t_i} (Y_{ij} - \bar{Y}_{i.})^2$$

Then $(t_i - 1)S_i^2 = \sum\limits_{j=1}^{t_i} (Y_{ij} - \bar{Y}_{i.})^2$ and we get $\dfrac{(t_i-1)S_i^2}{\sigma^2} = \dfrac{1}{\sigma^2} \sum\limits_{j=1}^{t_i} (Y_{ij} - \bar{Y}_{i.})^2$. Using $n_i - 1$ in place of t_i we can get following equality;

$$\frac{(n_i - 1 - 1)S_i^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{j=1}^{n_i-1} (Y_{ij} - \bar{Y}_{i.})^2 \tag{1.5}$$

$$\frac{(n_i-2)S_i^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{j=1}^{n_i-1}(Y_{ij}-\bar{Y}_{i.})^2 \tag{1.6}$$

and then we get

$$\frac{(n_i-2)S_i^2}{\sigma^2} \approx \chi^2_{(n_i-2)} \tag{1.7}$$

In equation 14, the sum on the righthand-side gives,

$$\sum_{i=1}^{k}(n_i-2)S_i^2 \tag{1.8}$$

Pooled variance is

$$S_p^2 = \frac{(n_1-2)S_1^2+(n_2-2)S_2^2+...+(n_k-2)S_k^2}{(n_1-2)+(n_2-2)+...+(n_k-2)} = \frac{\sum\limits_{i=1}^{k}\left[\sum\limits_{j=1}^{t_i}(Y_{ij}-Y_{i.})^2\right]}{\sum\limits_{i=1}^{k}(n_i-2)} \tag{1.9}$$

where $\sum\limits_{i=1}^{k}(n_i-2) = \sum\limits_{i=1}^{k}n_i-2k$, because $\sum\limits_{i=1}^{k}n_i = N$. Then, pooled variance can be shown as

$$S_p^2 = \frac{1}{N-2k}\sum_{i=1}^{k}(n_i-2)S_i^2 \tag{1.10}$$

Therefore we get $\frac{(N-2k)S_p^2}{\sigma^2} \approx \chi^2_{(N-2k)}$. Therefore, the denominator of **MP** test statistic is distributed asymptotically as a central χ^2-variable with $N-2k$ degrees of freedom. Because the ratio of two chi square distributions has the F distribution, the **MP** test has an F distribution with $(k-1, N-2k)$ degrees of freedom. That is, distribution of the proposed **MP** test is thus,

$$\frac{\dfrac{\sum\limits_{i=1}^{k}n_i(Y_{i.}-Y_{..})^2}{\sigma^2/(k-1)}}{\dfrac{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{t_i}(Y_{ij}-Y_{i.})^2}{\sigma^2/(N-2k)}} \sim F_{(k-1,\,N-2k)} \tag{1.11}$$

The **MP** test rejects the null hypothesis if $\mathbf{MP} > F_{(\alpha,k-1,\,N-2k)}$. The researcher then needs to adopt a test that is robust to heterogeneity of variances, such as the Welch test, James second-order test, Alexander-Govern's test. If the null hypothesis can not be rejected, ANOVA F test can be used to compare the group means.

Distributions	Mean	Variance	Skewness	Kurtosis
t (5)	0.00	1.67	0.00	6.00
$\chi^2(3)$	3.00	6.00	1.63	4.00
$\beta(6, 1.5)$	0.80	1.36	-1.00	0.60
$\omega(1.5, 1)$	0.90	0.37	1.07	1.40

Table 1: The characteristics of the distributions. t (5): t distribution with 5 d.f, $\chi^2(3)$: chi-square distribution with 3 d.f $\beta(6, 1.5)$: Beta distribution (6,1.5), $\omega(1.5, 1)$: Weibull distribution $(1.5, 1)$

2 Materials and Methods

Monte Carlo techniques were used to study the power and Type I error rates of the above tests across a variety of situations. The following variables are manipulated in this simulation study:

a) Type of population distribution,
b) Variance homogeneity and heterogeneity,
c) Degree of variance heterogeneity,
d) Group sizes,
e) The relationship between group size and population variances (direct and inverse pairing).

With respect to the effects of distributional shape on Type I error and test power, we chose to investigate non-normal distributions in which the data were obtained from a variety of skewed distributions ($\chi^2(3)$, $\beta(6.5, 1)$, $\omega(1.5, 1)$, and $t(5)$ distributions). These particular types of non-normal distributions were selected since educational and psychological research data typically have skewed distributions, and those distributions are predominantly used in literature to study deviations from normality (Tiku and Balakrishnan, 1984; Tan and Tabatabai, 1986; Loh, 1987; Sharma, 1991; Wilcox, 1994; Lim and Loh, 1996; Wludyka and Nelson, 1999; Keselman et al., 2002). Sawilowsky and Blair (1992) investigated the effects of eight non-normal distributions, which were identified by Micceri (1989) on the robustness of Student's t test, and they found that only the distributions with the most extreme degree of skewness (e.g., skewness=1.64) affected Type I error control of the independent sample t statistics. In this study, maximum degree of skewness used was 1.63. If the standard deviation ratio was $R = (\sigma_{(1)}/\sigma_{(k)})$(where $\sigma_{(1)} \leq \cdots \leq \sigma_{(k)}$represent the ordered standard deviations), Fenstad (1983) argued that having R as large as 4 was not extreme and a survey of studies reported by Wilcox et al. (1986) supported his view. Brown and Forsythe (1974a) considered $R \leq 3$, while Box (1954) limited his numerical results to $R \leq \sqrt{3}$. Many papers have shown that for larger R-values, the ANOVA F test may not provide satisfactory control over the probability of a Type I error (Rogan and Keselman, 1977; Tomarken and Serlin, 1986; Wilcox, 1988).

In this simulation study, maximum heterogeneity of variances used was 1:4 and 1:1:4 because using small variance differences of groups in a simulation enables the researcher to surface the differences between the tests. Unequal variances and unequal group sizes were both positively (direct pairing) and negatively (inverse pairing) paired. For positive pairings, the group having the least number of observations was associated with the population having the smallest variance, while the group having the greatest number of observations was associated with the population having the largest variance. For negative pairings, the group having the least number of observations was associated with the population having the largest variance, while the group having the greatest number of observations was associated with the population having the smallest variance. Data from the $\chi^2(3)$, $\beta(6.5, 1)$, $\omega(1.5, 1)$, and $t(5)$ distributions were generated using random number generators from IMSL (1994). First, using IMSL (1994) subroutine RNCHI, RNBET, RNWIB, and RNSTT, n_j $\chi^2(3)$ distribution, $\beta(6.5, 1)$ distribution, $\omega(1.5, 1)$ distribution, and $t(5)$ distribution were generated for group j. The variance ratios were chosen as 1:1, 1:2 and 1:4 for $k = 2$ and 1:1:1, 1:2:3, and 1:1:4 for $k = 3$. In order to form heterogeneity among the population variances, random numbers in the samples were multiplied by specific constant numbers ($\sigma = 1, \sqrt{2}, \sqrt{3}, \sqrt{4}$). The populations were standardized since they have different means and variances. Shape of distributions did not change while the means changed to 0 and the standard deviations changed to 1. Characteristics of the samples were as follows:

For $k = 2$, 300,000 data sets were generated from $\chi^2(3)$, $\beta(6.5, 1)$, $\omega(1.5, 1)$, $t(5)$ populations employing sample sizes of 4:4, 6:6, 8:8, 10:10, 12:12, 14:14, 16:16, 18:18, 20:20, 4:8, 10:15, 20:30, 8:4, 15:10, and 30:20. Variances of the populations from which these samples were drawn were set to 1:1, 1:2 and 1:4. For $k = 3$, 300,000 data sets were generated from $\chi^2(3)$, $\beta(6.5, 1)$, $\omega(1.5, 1)$, $t(5)$ populations employing sample sizes of 4:4:4, 6:6:6, 8:8:8, 10:10:10, 12:12:12, 14:14:14, 16:16:16, 18:18:18, 20:20:20, 4:8:10, 10:15:25, 20:40:60, 10:8:4, 25:15:10, and 60:40:20. Variances of the populations from which these samples were drawn were set to 1:1:1, 1:2:3 and 1:1:4. Each given set of parameter values and frequencies of samples for the rejection regions were counted for Bartlett's test, Levene 1 test, Levene 2 test, Levene 3 test, Levene 4 test, and **MP** test. For each pair of samples (for $k = 2$ and $k = 3$), Bartlett's test, Levene 1 test, Levene 2 test, Levene 3 test, Levene 4 test, and **MP** test statistics were calculated. We computed Bartlett's statistics and counted the frequency satisfying $\mathbf{Bart} > \chi^2(k-1)$ d.f. for α=0.05. For the Levene 1, Levene 2, Levene 3, and Levene 4 test, we computed $\mathbf{Lev1}, \mathbf{Lev2}, \mathbf{Lev3}$, and $\mathbf{Lev4}$ statistics and counted the frequency satisfying $\mathbf{Lev}_i > F(k-1, N-k), (i = 1, 2, 3, 4)$ d.f. for $\alpha = 0.05$. For **MP** test, we computed MP statistics and counted the frequency satisfying $\mathbf{MP} > F(k-1, N-2k)$ d.f. for $\alpha = 0.05$. For each test, we checked to see if the hypothesis, which was false, was rejected at $\alpha = 0.05$. The proportion of observations falling in the critical regions was recorded for different variance pattern, n, and distributions. This proportion estimation is the test power if the variances from the populations differ. By power, we mean the ability of the test to detect unequal variances when the variances are, in fact, unequal. Therefore, in these kind of studies the effect size is the variance ratios such as 1:1, 1:2 and 1:4 for $k = 2$ and 1:1:1, 1:2:3, and 1:1:4 for $k = 3$. This proportion estimation is the Type I error if the variances from the population do not differ ($\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$). Ferron and Sentovich (2002) estimated statistical power for three randomization tests using multiple-baseline designs. They stated that they used > .80 as the sufficient power level for comparing the tests. Therefore, .80 was assumed to be the sufficient power level in this study.

3 Results

Type I error rates for different tests, variance-ratios and sample sizes are given in Table 2. In all distributions, the proposed method managed to keep the type I error rates closer to the predetermined alpha level than did the other methods, indicating that the new method may be preferred over methods such as Levene's and Bartlett test. Also, the proposed test was superior for unequal sample size conditions, especially when distributions were $\beta(6, 1.5)$, $t(5)$, and $\omega(1.5, 1)$. When number of groups (k) increased, type I error rate deviated out of the predetermined alpha level in all methods except in the proposed method. For instance when $k = 2$, distribution was $\chi^2(3)$ and $n_i = 16$, type I error rates were 25.71 %, 12.25 %, 6.09 %, 12.17 % and 13.93 % for contemporary tests and 5.12 % for the proposed method. When k was 3 and all other conditions were the same, type I error rates were 27.81 %, 15.25 %, 6.46 %, 14.94 % and 17.95 % for contemporary tests and 4.81 % for the proposed method. Bartlett's test deviated from the predetermined alpha level when the distribution was $t(5)$ and the number of groups compared (k) increased. Levene tests generally deviated from the predetermined alpha level when the distribution was $\chi^2(3)$ and the number of groups compared (k) increased. The proposed method remained stable both on the $t(5)$ and the $\chi^2(3)$ distribution. As the number of observations increased, Type I error rate approached to the predetermined alpha level for the proposed method, while it deviated from the alpha level for Bartlett's test. All Levene's tests approached to the nominal alpha level as the number of observations increased.

When the distribution was $\chi^2(3)$, all tests tended to deviate from nominal alpha level. The proposed method was not affected from the distribution as much as the other tests (Table 2) and was closest to the alpha level. Bartlett's test had closer alpha values in $\beta(6, 1.5)$ distribution while all Levene tests had closer alpha values in $t(5)$ distribution except for small sample sizes. As the ratio of variances (δ_i) increased, power of all tests increased (Table 3 and 4). The increase in power was clearer as the sample size increased. The new method was more powerful overall for different sample sizes and distributions.

When the variance ratio was 1:2, Bartlett's test was more powerful than all others (Table 3). For t distribution, Bartlett's test was more powerful regardless of the variance ratio. When the variance ratio was 1:4, the proposed method was more powerful than all the other tests for $\chi^2(3)$, $\beta(6.5, 1)$ and $\omega(1.5, 1)$ distributions. This indicates that the proposed method is the most useful one when the heterogeneity of variances is high.

In this study, sufficient power level was accepted as 70 %. For variance ratio of 1:2, none of the tests was able to reach the sufficient power level. When the variance ratio was 1:4, Bartlett's test and the proposed method reached the given power level in large sample sizes. When the distribution was $\chi^2(3)$, the two tests reached 70 % for sample size 20 while Levene tests were below the sufficient power levels for all sample sizes. When the distribution was $\beta(6, 1.5)$, Bartlett's test and the proposed method reached sufficient power levels for sample sizes 16 and more. Levene 1 reached that level of power only when the sample size was 20. When the distribution was $\omega(1.5, 1)$, Bartlett's test and the proposed method reached sufficient power levels for sample sizes 18 and more. Levene 1 reached that level of power only when the sample size was 20. For $t(5)$ distribution, the Bartlett's test reached the given power level for sample sizes 18 and 20, while all other tests remained below that level. For variance ratio of 1:2:3, none of the tests was able to reach the sufficient power level (Table 4). Bartlett's test was powerful than the other tests for this ratio of variances.

When the variance ratio was 1:1:4, Bartlett's test and the proposed method reached the sufficient power level in large sample sizes. When the distribution was $\chi^2(3)$, Bartlett's test reached the sufficient power level for sample sizes 14 and more. The proposed method reached the level for sample sizes 18 and 20. Levene 1 and Levene 4 could reach the sufficient power level only for sample size 20. When the distribution was $\beta(6, 1.5)$, the proposed method was the only test that reached the sufficient power levels (sample sizes 14 and more). When the distribution was $\omega(1.5, 1)$, Bartlett's test reached the power level for sample sizes 14 and more, while the proposed method reached sufficient levels for sample sizes 16 and more. Levene 1 and Levene 4 reached the given power level for large sample sizes. When the distribution was $t(5)$, only the Bartlett's test reached the sufficient power level for sample sizes 14 and more. Overall, Bartlett's test was more powerful than Levene tests, while Levene tests were closer to the predetermined alpha level than Bartlett's test. The proposed method provided an optimum for Type I error rate and power level, giving a performance close to Bartlett's in power, and Levene's in Type I error rate.

When inverse pairing was applied, highest power was achieved by the **MP** test for 1:4 and for Beta and Weibull distribution when sample sizes were relatively high (15:10, 30:20). In other situations, the **MP** test achieved power levels similar to the Barttlet's test (Table 5). When the sample sizes were 60:40:20 and 25:15:10 and variances were 1:1:4, the **MP** test achieved the highest power level in all distributions except in $t(5)$ distribution (Table 6).

4 Discussion

Lim and Loh (1996) ran simulations on samples with $t(5)$ distribution. They reported that the Type I error rates were 21.80 % when Barttlet's test was run on samples with 10 observations and 24.90 % on samples with 20 observations. These values are similar to those in this study; small differences could be attributed to the differences in number of simulations and number of groups. Keyes and Levy (1997) reported that the Type I error rates of Levene 1, Levene 2, and Levene 3 tests were affected adversely from inequality of sample sizes. Pardo et al (1997) reported that Bartlett's test was affected from deviations from normality. Sharma (1991) supported this view, especially when distributions were $\chi^2(3)$ and $t(5)$. Boos and Brownie (1989) reported that when R=1:4, the power of the Bartlett's tests was higher than the Levene 1 test. Our simulations results are consistent with these results.

References

[1] Alexander, R.A. and Govern, D.M. (1994): A new and simpler approximation for ANOVA under variance heterogeneity, *Journal of Educational Statistics*, **19**, 91-101.

[2] Anonymous, (1994): FORTRAN Subroutines for mathematical applications. IMSL MATH/LIBRARY. Vol.1-2, Houston: Visual Numerics, Inc.

[3] Bishop, T.A. and Dudewicz, E.J. (1978): Exact analysis of variance with unequal variances: Test procedures and tables, *Technometrics*, **20**, 419-430.

[4] Boos, D.D. and Brownie, C. (1989): Bootstrap methods for testing homogeneity of variances, *Technometrics*, **31(1)**, 69-82.

[5] Box, G.E.P. (1954): Some theorems on quadratic forms applied in the study of analysis variance problems, I. Effect of inequatiley of variance in the one-way model, *The Annals of Mathematical Statistics*, **25,** 290-302.

[6] Brown, M. B. and Forsythe, A.B. (1974): Robust tests for the equality of variances, *Journal of the American Statistical Association*, **69**, 364–367.

[7] Brown, M.B. and Forsythe, A.B. (1974a): The small sample behavior of some statistics which test the equality of several means, *Technometrics*, **16**, 129-132.

[8] Bryk, A.S. and Raudenbush, S.J. (1987): Heterogeneity of variance in experimental studies: A challange to conventional interpretations, *Psychological Bulletin*, **104**, 396-404.

[9] Çamdeviren, H. and Mendeş, M. (2005): A simulation study for type III error rates of some variance homogeneity tests, *Pakistan Journal of Statistics*, **21 (2)**, 223-234.

[10] Chen, S.-Y. and Chen, H.J. (1998): Single-stage analysis of variance under heteroscedasticity, *Communications in Statistics-Simulation and Computation*, **27,** 641-666.

[11] Cochran, W.G. and Cox, G.M. (1957): *Experimental design.* New York: John Willey and Sons Inc.

[12] Conover, W. J., Johnson, M.E. and Johnson, M.M. (1981): A comparative study of tests for homogeneity of variances with applications to the outer continental shelf bidding data, *Technometrics*, **23**, 351–361.

[13] Fenstad, G.U. (1983): A comparison between U and V tests in the behrens-fisher problem, *Biometrika,* **70**, 300-302.

[14] Ferron J. and Sentovich, C. (2002): Statistical power of randomization tests used with multiple- baseline designs, *Journal of Experimental Education*, **70(2)**, 165-178.

[15] Hsiung, T.C. and Olejnik, S. (1996): Type I error rates and statistical power for James second-order test and the univariate F test in two-way ANOVA models under heteroscedasticity and/or non-normality, *The Journal of Experimental Education*, **65**, 57-71.

[16] Keselman, H.J., Wilcox., R.R., Othman, A.R. and Fradette, K. (2002): Trimming, transforming statistics, and bootstrapping: Circumventing the biasing effects of heteroscedasticity and non-normality, *Journal of Modern Applied Statistical Methods*, **1(2)**, 288-309.

[17] Keyes, T.K. and Levy, M.S. (1997): Analysis of Levene's test under design imbalance, *Journal of Educational Behavior Statistics*, **22(2)**, 227-236.

[18] Levene, H. (1960): *Robust tests for equality of variances.* Contributions to Probability and Statistics, edited by I. Olkin. Stanford University Press. Palo Alto. CA., 278–292.

[19] Lim, T.S. and Loh, W.Y. (1996): A comparison of tests of equality of variances, *Computational Statistics and Data Analysis*, **22**, 287-301.

[20] Loh, W. -Y. (1987): Some modifications of Levene's test of variance homogeneity, *Journal of Statistical Computation and Simulation*, **28**, 213-226.

[21] Luh, W.M. and Guo, J.H. (2000): Approximate transformation trimmed mean methods to the test of simple linear regression slope equality, *Journal of Applied Statistics*, **27**, 843-858.

[22] Mendeş, M. (2003): The comparison of Levene, Bartlett, Neyman-Pearson and Bartlett 2 tests in terms of actual type I error rates, *Journal of Agriculture Sciences*, **9 (2)**, 143-146.

[23] Micceri, T. (1989): The unicorn, the normal curve, and other improbable creatures, *Psychological Bulletin*, **105 (1)**, 156-166.

[24] Nelson, L. S. (2000): Comparing two variances from normal populations, *Journal of Quality Technology*, **32**, 79–80.

[25] Oshima,T.C. and Algina, J. (1992): Type I error rates James's second-order test and Wilcox's H_m test under heteroscedasticity and non-normality, *British Journal of Mathematical and Statistical Psychology*, **45,** 255-263.

[26] Ott, L. (1998): *An introduction to statistical methods and data analysis*, Third Edition. PWS-Kent Publishing Company.

[27] Pardo, J.A., Pardo, M.C., Vicente, M.L. and Esteban, M.D. (1997): A statistical information theory approach to compare the homogeneity of several variances, *Computational Statistics and Data Analysis*, **24 (4)**, 411-416.

[28] Piepho, H.P. (1997): A Monte Carlo test for variance homogeneity in linear models, *Biometrical Journal*, **38**, 461-473.

[29] 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,** 493-498.

[30] Sawilowsky, S.S. and Blair, R.C. (1992): A more realistic look at the robustness and type II error properties of the t test to departures from population normality, *Psychological Bulletin*, III, 352-360.

[31] Schneider, P.J. and Penfield, D.A. (1997): Alexander and Govern's approximation: Providing an alternative to ANOVA under variance heterogeneity, *The Journal of Experimental Education*, **65,** 271-286.

[32] Sharma, S.C. (1991): A new jackknife test for homogeneity of variances, *Communications in Statistics-Simulation and Computation*, **20(2-3)**, 479-495.

[33] Tan, W.Y. and Tabatabia, M.A. (1986): Some Monte Carlo studies on the comparison of several means under heteroscedasticity and robustness with respect to departure from normality, *Biometrical Journal* , **7**, 801-814.

[34] Tiku, M.L. and Balakrishnan, N.(1984): Testing equality of population variances the robust way. *Communications in Statistics-Theory and Methods*, **13(17)**, 2143-2159.

[35] Tomarken, A. J. and Serlin, R.C. (1986): Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures, *Psychological Bulletin*, **99**, 90-99.

[36] Weerahandi, S. (1995): ANOVA under unequal error variances, *Biometrics*, **51**, 589-599.

[37] Wei-ming, L. (1999): Developing trimmed mean test statistics for two-way fixed effects ANOVA models under variance heterogeneity and non-normality, *The Journal of Experimental Education,* **67 (2),** 243-264.

[38] Wilcox, R.R., Charlin, V.L. and Thompson, K.L. (1986): New Monte Carlo results on the robustness of the ANOVA F, W and F* statistics, *Journal of Statistical Computation and Simulation*, **15**, 33-943.

[39] 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**, 109-117.

[40] Wilcox, R.R. (1989): Adjusting for unequal variances when comparing means in one-way and two-way effects ANOVA models, *Journal of Educational Statistics*, **14**, 69-278.

[41] Wilcox, R.R. (1994): A one-way random effects model for trimmed means, *Psychometrika*, **59**, 289-306.

[42] Wicox, R.R. (1995): ANOVA: A paradigm for low power and misleading measures of effect size? *Review of Educational Research*, **65**, 51-77.

[43] Wilcox, R.R. (2002): Comparing variances of two independent groups, *British Journal of Mathematical and Statistical Psychology*, **55**, 169-175.

[44] Winer, B.J., Brown, D.R. and Michels, K.M. (1991): *Statistical principles in experimental design*, New York: McGraw-Hill Book Company.

[45] Wludyka, P. and Nelson, P.R. (1999): Two non-parametric analysis-of-means type tests for homogeneity of variances, *Journal of Applied Statistics*, **26**, 243–256.

[46] Zar, J.H. (1999): *Biostatistical analysis*, New Jersey: Prentice –Hall Inc. Simon and Schuster/A Viacom Company.

Groups		$k=2$							$k=3$					
Distr	n_i	**Bart**	**Lev1**	**Lev2**	**Lev3**	**Lev4**	**MP**	n_i	**Bart**	**Lev1**	**Lev2**	**Lev3**	**Lev4**	**MP**
$\chi^2(3)$	4	14.57	11.83	7.35	13.93	15.63	4.90	4	14.90	15.92	8.79	17.28	20.65	5.46
	6	18.03	12.70	7.71	12.86	14.56	6.85	6	19.08	16.36	9.38	15.94	19.20	6.81
	8	21.11	12.55	7.11	12.45	14.18	6.37	8	22.01	15.74	8.21	15.38	18.42	6.07
	10	23.25	12.38	6.76	12.25	13.98	6.05	10	24.33	15.63	7.41	15.19	18.14	5.41
	12	24.31	12.45	6.59	12.21	13.94	5.52	12	25.50	15.48	7.12	15.03	18.03	5.27
	14	25.17	12.34	6.34	12.12	13.99	5.43	14	26.68	15.07	6.71	14.75	17.68	5.04
	16	25.71	12.25	6.09	12.17	13.93	5.23	16	27.81	15.25	6.46	14.94	17.95	4.81
	18	26.16	12.07	5.94	11.95	13.71	4.99	18	28.47	15.03	6.16	14.84	17.83	4.90
	20	26.84	11.93	5.77	11.77	13.62	5.03	20	28.99	14.73	5.85	14.59	17.58	5.09
	4:8	12.22	11.82	6.30	12.93	14.06	7.32	4:8:10	19.42	15.67	9.79	16.55	18.96	6.40
	10:15	18.20	11.50	6.27	11.73	13.44	7.83	10:15:25	26.87	14.48	6.40	14.56	17.51	5.07
	20:30	21.13	11.70	5.64	11.61	13.49	6.71	20:40:60	31.35	14.26	5.26	14.39	17.57	4.78
$\beta(6.5{,}1)$	4	6.54	10.48	7.12	10.36	12.13	6.52	4	6.98	12.46	8.29	11.93	14.67	5.30
	6	7.23	10.30	7.08	8.97	10.79	6.19	6	8.08	12.07	8.11	10.19	12.75	6.54
	8	7.76	9.67	6.74	8.40	9.99	6.01	8	8.83	11.40	7.64	9.63	11.97	6.12
	10	7.98	9.40	6.73	8.13	9.60	5.92	10	8.92	11.09	7.28	9.35	11.60	5.62
	12	8.18	9.35	6.58	8.26	9.60	5.82	12	9.31	11.11	7.03	9.47	11.55	5.41
	14	8.17	9.26	6.48	8.18	9.51	5.62	14	9.47	10.74	6.82	9.14	11.28	5.42
	16	8.27	9.29	6.52	8.39	9.60	5.59	16	9.50	10.62	6.64	9.29	11.28	5.06
	18	8.39	9.00	6.28	7.99	9.20	5.26	18	9.49	10.41	6.48	8.99	10.97	4.95
	20	8.44	9.00	6.30	8.03	9.31	5.23	20	9.68	10.57	6.50	9.01	11.12	5.04
	4:8	7.03	9.73	5.67	9.80	10.97	5.67	4:8:10	8.14	11.18	7.63	11.13	12.68	5.67
	10:15	7.81	9.25	6.35	8.37	9.73	5.81	10:15:25	9.15	10.45	6.61	9.27	11.39	5.11
	20:30	8.57	9.46	6.04	8.26	9.45	5.28	20:40:60	9.07	9.69	5.59	8.74	10.36	4.75
$\omega(1.5{,}1)$	4	7.00	10.30	6.94	10.46	12.26	4.40	4	8.17	12.69	8.13	12.46	15.21	5.10
	6	8.46	10.14	6.77	9.06	10.90	6.34	6	7.94	12.64	8.03	12.31	15.16	6.08
	8	9.31	9.65	6.50	8.67	10.12	6.04	8	10.06	11.97	7.94	10.36	13.00	5.91
	10	9.92	9.48	6.35	8.34	9.85	5.95	10	11.31	11.72	7.45	10.03	12.44	5.55
	12	10.28	9.33	6.23	8.37	9.74	5.80	12	12.14	11.19	6.96	9.70	11.86	5.27
	14	10.72	9.16	5.93	8.29	9.58	5.54	14	12.90	11.08	6.70	9.56	11.82	5.11
	16	10.98	9.26	6.08	8.38	9.66	5.42	16	13.10	10.80	6.51	9.50	11.42	4.97
	18	11.24	8.93	5.85	7.98	9.32	5.08	18	13.66	10.71	6.31	9.43	11.39	4.86
	20	11.40	9.14	5.90	8.32	9.59	5.12	20	14.11	10.57	5.92	9.24	11.31	4.86
	4:8	7.61	9.22	5.20	9.36	10.64	5.72	4:8:10	10.41	11.63	7.77	11.09	13.13	5.89
	10:15	9.80	8.93	5.90	8.37	9.58	5.70	10:15:25	13.18	10.40	6.18	9.61	11.41	4.86
	20:30	11.30	8.76	5.73	8.00	9.11	4.89	20:40:60	15.68	10.43	5.77	9.41	11.19	4.83
$t(5)$	4	7.75	8.17	4.87	9.08	10.51	3.61	4	9.45	10.08	5.59	10.47	12.64	4.65
	6	9.86	7.08	4.25	7.01	8.12	5.38	6	12.80	8.30	5.09	7.84	9.58	5.71
	8	11.40	6.57	3.90	6.51	7.32	5.55	8	15.00	7.24	4.44	6.75	8.16	5.58
	10	12.66	6.32	4.06	6.18	6.91	5.33	10	16.80	6.78	4.48	6.39	7.54	5.17
	12	13.69	6.02	3.86	5.93	6.53	5.19	12	18.29	6.50	4.23	6.13	7.13	5.08
	14	14.17	5.90	3.93	5.87	6.44	5.13	14	19.23	6.04	3.94	5.84	6.64	4.95
	16	15.02	5.80	4.04	5.67	6.19	5.04	16	20.19	5.98	3.99	5.67	6.40	5.10
	18	15.66	5.63	4.09	5.60	5.96	5.02	18	20.94	5.96	3.98	5.78	6.50	4.96
	20	16.15	5.69	4.12	5.65	5.99	5.06	20	21.69	5.88	3.99	5.61	6.23	4.96
	4:8	9.10	7.93	4.53	7.92	8.97	5.26	4:8:10	13.03	8.11	6.37	8.21	9.23	6.26
	10:15	13.11	5.74	3.85	5.69	6.32	5.08	10:15:25	19.79	6.44	4.91	5.81	6.77	5.75
	20:30	16.77	5.17	4.07	5.38	5.65	4.74	20:40:60	21.19	5.75	4.38	5.49	5.91	5.07

Table 2: Type I error rates (%) for different tests. distributions and sample sizes for $k=2,3$.

Variance ratio		1:2						1:4					
Distr	n_i	**Bart**	**Lev1**	**Lev2**	**Lev3**	**Lev4**	**MP**	**Bart**	**Lev1**	**Lev2**	**Lev3**	**Lev4**	**MP**
$\chi^2(3)$	4	17.22	13.28	8.02	16.15	17.77	7.99	24.70	18.39	10.30	23.05	24.98	16.91
	6	23.51	16.31	9.81	16.46	18.66	13.93	36.58	26.58	15.75	26.98	30.21	31.57
	8	27.43	18.16	10.34	17.62	20.27	15.91	44.21	33.51	19.32	31.35	36.16	39.55
	10	31.11	19.93	11.23	18.64	21.74	17.21	50.30	39.78	22.99	35.41	41.78	46.46
	12	33.72	21.49	11.99	19.55	23.22	18.82	55.96	45.64	26.55	39.63	47.31	52.89
	14	36.05	23.33	12.85	20.78	24.86	20.76	60.61	51.01	30.15	43.27	52.38	58.67
	16	37.95	25.01	13.61	21.93	26.69	22.62	64.85	55.66	33.51	46.71	56.73	63.77
	18	39.54	26.85	14.61	23.06	28.36	24.39	67.96	60.13	36.79	49.99	60.89	69.05
	20	41.61	28.70	15.56	24.41	30.22	26.09	71.13	64.44	40.11	53.42	64.97	73.21
$\beta(6.5,1)$	4	9.06	12.41	8.26	12.97	14.82	6.72	17.22	17.82	11.21	20.00	22.22	14.57
	6	12.73	14.70	10.10	12.92	15.36	13.10	29.36	27.58	18.40	24.65	28.80	31.33
	8	16.01	16.87	11.81	14.14	17.04	15.82	40.40	35.65	24.76	29.53	35.77	41.82
	10	19.23	19.00	13.64	15.11	18.78	18.49	50.02	44.08	31.96	34.23	42.74	50.86
	12	22.34	21.20	15.69	16.37	20.59	20.61	57.68	51.39	38.86	38.47	49.02	58.74
	14	24.77	23.41	17.52	17.59	22.43	23.29	64.98	58.31	45.39	42.76	55.03	65.42
	16	27.65	26.07	19.56	19.11	24.81	25.77	70.88	64.63	51.63	46.85	60.67	71.52
	18	30.41	28.20	21.59	20.24	26.67	28.20	76.39	69.92	57.77	50.66	65.48	76.95
	20	33.02	30.48	23.49	21.33	28.47	31.02	80.73	74.64	63.11	54.24	69.81	81.31
$\omega(1.5,1)$	4	9.61	11.90	7.85	12.64	14.56	6.95	18.04	17.24	10.47	19.90	22.17	14.60
	6	14.03	14.48	9.61	13.09	15.41	13.10	30.19	26.38	17.06	24.13	28.08	30.53
	8	17.71	16.32	10.83	14.28	16.91	15.70	40.69	34.92	23.36	29.20	35.22	40.48
	10	21.24	18.58	12.81	15.34	18.62	17.69	49.56	41.99	29.27	33.51	41.38	48.50
	12	24.00	20.79	14.47	16.51	20.60	19.88	57.52	49.63	35.52	38.10	48.02	56.05
	14	26.90	22.93	15.94	17.84	22.40	21.90	63.86	55.83	40.98	42.08	53.59	62.72
	16	29.59	24.91	17.38	18.92	24.18	24.18	69.76	62.15	46.68	46.20	59.40	68.90
	18	32.13	27.01	19.19	19.93	25.95	26.29	74.35	67.22	51.76	49.83	64.10	74.02
	20	34.59	29.21	20.94	21.18	27.94	28.81	78.48	71.70	56.67	53.36	68.24	78.46
$t(5)$	4	10.33	9.66	5.70	11.09	12.55	5.39	18.47	14.66	8.04	17.83	19.76	11.47
	6	15.13	10.82	6.37	10.53	12.23	9.28	31.42	21.27	12.46	20.77	23.91	21.60
	8	19.38	12.12	7.30	10.93	12.88	11.22	41.35	28.00	16.82	24.72	29.53	28.10
	10	23.09	13.48	8.76	12.12	14.27	12.79	49.58	34.30	21.60	28.76	35.08	35.12
	12	26.23	15.09	10.02	12.80	15.54	14.53	56.89	40.76	26.97	32.57	40.64	41.04
	14	29.37	16.75	11.33	13.85	16.95	16.01	63.24	46.77	32.59	36.35	46.04	46.80
	16	31.92	17.98	12.44	14.67	18.02	17.27	68.18	52.66	37.43	40.66	51.71	52.63
	18	34.49	19.91	14.01	15.44	19.59	19.21	72.56	58.11	42.55	44.00	56.71	57.60
	20	37.04	21.78	15.56	16.83	21.33	21.23	76.03	63.02	47.12	47.67	61.23	62.66

Table 3: Test power (%) for different tests, distributions, sample sizes and variance ratios for $k = 2$.

Variance ratio		1:2:3						1:1:4					
Distr	n_i	**Bart**	**Lev1**	**Lev2**	**Lev3**	**Lev4**	**MP**	**Bart**	**Lev1**	**Lev2**	**Lev3**	**Lev4**	**MP**
$\chi^2(3)$	4	19.48	20.05	10.51	21.92	25.89	12.28	26.67	25.14	12.71	26.46	31.17	16.98
	6	29.12	23.53	13.28	23.08	27.52	18.10	40.83	32.94	19.71	30.21	36.63	28.34
	8	36.29	26.55	13.92	25.16	30.29	20.89	51.33	40.67	24.46	34.75	43.00	37.05
	10	42.22	29.67	15.02	26.97	33.23	23.83	58.94	46.71	28.20	38.78	48.49	45.90
	12	47.09	32.63	15.91	29.33	36.23	26.73	65.42	52.97	32.18	43.39	54.32	53.27
	14	51.42	36.25	17.08	31.79	39.73	30.66	70.59	58.55	36.35	47.14	59.43	60.15
	16	54.72	38.87	18.29	33.70	42.19	33.71	75.00	63.38	40.49	50.97	64.03	66.37
	18	58.09	41.94	19.57	35.66	45.30	37.29	78.95	68.21	44.09	54.89	68.66	71.88
	20	61.27	45.11	21.06	38.26	48.37	40.95	81.78	72.10	48.05	58.19	72.10	76.63
$\beta(6.5,1)$	4	12.07	16.81	10.59	16.35	20.11	12.18	15.09	18.97	11.83	18.06	22.18	18.14
	6	18.94	21.16	14.17	17.58	22.21	19.38	24.42	25.56	18.39	19.92	25.93	32.55
	8	24.62	24.69	16.46	19.70	25.05	24.14	32.75	32.04	24.00	23.04	30.90	43.53
	10	30.66	29.05	19.43	21.89	28.84	28.52	40.27	37.84	28.95	26.12	35.56	54.03
	12	36.15	33.12	22.27	24.31	32.36	32.79	47.17	43.71	34.70	29.35	40.57	62.80
	14	41.47	37.78	25.52	26.91	36.43	37.51	53.43	49.24	39.97	32.46	45.42	70.05
	16	46.77	41.93	28.75	29.51	40.17	41.85	59.23	54.55	45.19	35.64	49.96	76.44
	18	51.19	45.87	32.05	31.64	43.61	46.41	64.20	59.22	50.37	38.43	54.17	81.31
	20	55.87	50.10	35.50	34.20	47.33	50.49	68.64	63.52	55.13	41.48	58.02	85.86
$\omega(1.5,1)$	4	13.20	16.75	10.17	16.83	20.33	11.89	21.68	22.89	13.24	21.61	26.90	17.14
	6	20.72	20.56	13.30	17.70	22.09	18.75	35.37	32.30	22.91	25.37	32.76	30.40
	8	27.35	24.00	15.35	19.82	25.05	22.81	46.63	40.90	29.79	30.16	39.92	40.93
	10	33.25	27.81	17.65	21.91	28.29	26.25	55.60	49.05	36.99	35.04	46.90	50.88
	12	38.89	31.83	19.88	24.29	32.04	30.27	64.46	56.69	43.54	39.52	53.43	59.03
	14	43.80	35.79	22.57	26.51	35.34	34.66	71.12	63.32	50.10	44.20	59.45	67.11
	16	48.92	39.85	25.03	28.96	39.17	38.84	76.49	69.30	56.21	48.37	65.04	73.16
	18	53.26	43.67	27.79	31.42	42.71	42.92	80.92	74.49	61.82	52.23	69.98	78.21
	20	57.27	47.43	30.56	33.70	46.10	47.61	84.68	78.80	66.70	56.45	74.41	83.03
$t(5)$	4	14.38	13.62	7.31	14.15	17.15	9.35	22.16	19.56	10.39	18.74	23.39	11.13
	6	23.12	14.98	8.88	13.78	16.99	14.18	36.60	26.12	16.83	20.93	27.41	21.58
	8	30.47	17.14	10.43	15.17	18.75	16.90	48.18	32.97	22.75	24.50	32.90	29.84
	10	36.47	19.77	12.09	16.64	21.04	19.83	57.74	40.27	29.06	28.51	38.95	38.04
	12	42.27	22.77	13.94	18.60	23.80	22.64	64.97	46.97	34.95	32.77	44.95	45.27
	14	47.12	25.72	15.79	20.42	26.59	25.39	70.96	53.53	40.93	36.76	50.83	52.20
	16	51.42	28.79	17.98	22.43	29.57	28.25	76.28	59.59	46.56	40.90	56.31	58.42
	18	55.88	31.97	20.13	24.50	32.34	31.28	80.23	64.99	51.60	45.24	61.62	64.29
	20	59.39	35.33	22.48	26.53	35.52	35.06	83.72	69.61	56.12	48.71	66.02	69.52

Table 4: Test power (%) for different tests, distributions, sample sizes and variance ratios for $k = 3$.

Variance ratio		1:2						1:4					
Distr	n_i	**Bart**	**Lev1**	**Lev2**	**Lev3**	**Lev4**	**MP**	**Bart**	**Lev1**	**Lev2**	**Lev3**	**Lev4**	**MP**
$\chi^2(3)$	4:8	16.20	12.87	4.73	18.89	17.98	12.95	27.79	18.26	5.64	29.06	27.21	25.06
	10:15	30.36	20.59	9.22	20.89	23.81	20.37	56.01	41.84	18.36	40.14	46.34	47.09
	20:30	42.98	30.00	12.69	26.61	32.62	27.17	80.52	71.21	36.98	61.04	72.55	71.75
	8:4	17.12	18.62	12.77	15.17	18.91	19.29	30.46	32.04	24.78	23.26	30.51	39.25
	15:10	29.08	22.33	15.35	18.72	23.24	23.13	56.11	48.90	35.45	39.02	48.72	53.04
	30:20	41.27	32.88	21.89	25.74	33.30	30.70	78.22	72.86	55.72	58.61	71.75	72.91
$\beta(6.5,1)$	4:8	10.76	11.99	5.38	15.60	15.54	8.99	21.57	19.64	6.83	27.19	26.27	19.29
	10:15	20.88	19.88	12.00	17.54	20.57	17.84	57.10	49.12	29.98	40.16	49.33	53.47
	20:30	37.46	33.80	23.18	24.23	32.17	32.00	87.78	81.69	65.03	62.05	77.99	86.60
	8:4	11.66	15.53	12.19	11.44	15.04	15.00	27.06	27.06	25.08	20.06	27.24	38.50
	15:10	22.14	21.88	19.00	15.04	20.32	23.50	59.08	59.08	47.35	36.72	49.10	63.47
	30:20	38.75	36.16	31.86	23.22	32.47	39.68	87.28	87.28	78.50	60.39	77.36	90.01
$\omega(1.5,1)$	4:8	11.42	11.23	4.21	15.47	14.97	8.12	23.13	17.99	6.30	26.26	25.33	18.59
	10:15	23.69	19.56	10.96	17.66	20.62	17.23	56.30	46.23	26.70	39.22	47.24	50.14
	20:30	23.98	32.05	19.70	24.46	31.66	30.38	84.93	78.89	57.09	60.54	76.12	83.36
	8:4	12.81	15.96	12.24	11.10	15.34	15.90	27.94	29.85	25.74	19.70	27.29	37.38
	15:10	24.52	21.98	18.34	15.50	20.90	23.10	57.41	51.88	44.16	36.90	48.30	60.81
	30:20	39.64	34.54	28.44	22.26	31.58	36.54	84.90	80.37	72.66	59.14	75.78	87.60
$t(5)$	4:8	12.88	7.38	2.52	11.88	10.96	5.42	24.00	12.35	3.19	21.83	19.62	10.66
	10:15	26.18	12.95	6.42	13.08	14.67	11.68	57.05	36.15	17.64	33.34	39.07	35.35
	20:30	42.23	23.13	14.04	18.74	23.61	22.15	82.00	68.19	44.76	53.78	67.51	66.69
	8:4	13.59	13.88	11.47	9.65	13.22	13.43	27.76	27.29	23.68	17.44	24.80	30.48
	15:10	25.71	17.08	14.22	12.39	16.34	17.12	56.68	45.62	37.38	31.81	42.55	46.64
	30:20	39.66	27.05	22.76	18.34	24.86	26.81	81.35	74.35	64.22	54.23	70.42	74.00

Table 5: Power of tests (%) for unbalanced data, $k = 2$

Variance ratio		1:2:3						1:1:4					
Distr	n_i	**Bart**	**Lev1**	**Lev2**	**Lev3**	**Lev4**	**MP**	**Bart**	**Lev1**	**Lev2**	**Lev3**	**Lev4**	**MP**
$\chi^2(3)$	4:8:10	27.97	18.01	7.31	25.78	25.88	11.84	48.54	34.43	14.54	37.54	41.87	30.10
	10:15:25	52.46	31.50	8.92	34.28	39.84	22.96	80.12	64.49	27.46	58.74	69.47	65.56
	20:40:60	75.03	55.00	16.84	50.10	61.52	49.94	97.69	95.26	66.88	86.69	95.33	97.80
	10:8:4	30.91	29.22	22.67	21.73	29.51	26.47	37.46	36.63	33.47	24.83	34.88	35.50
	25:15:10	54.25	43.40	29.56	31.84	42.90	43.95	65.39	57.04	46.30	41.06	55.11	62.12
	60:40:20	77.46	68.17	47.10	52.84	68.20	69.05	87.33	82.37	71.84	64.22	80.46	90.58
$\beta(6.5,1)$	4:8:10	16.82	16.91	7.77	21.24	22.10	13.70	42.63	37.50	20.77	34.78	40.61	35.86
	10:15:25	40.00	34.70	16.34	30.02	36.87	30.92	82.23	74.78	51.55	56.95	72.32	77.40
	20:40:60	72.32	65.44	40.72	47.28	63.65	64.06	99.57	98.97	96.12	88.30	97.63	99.62
	10:8:4	21.57	25.70	22.86	16.02	23.56	28.24	30.89	35.19	37.10	19.14	30.32	39.71
	25:15:10	49.63	46.50	40.32	27.46	41.44	51.64	65.70	63.94	62.64	38.60	56.57	71.04
	60:40:20	82.60	78.66	73.33	51.22	71.82	81.19	91.70	89.92	90.40	63.63	83.81	94.32
$\omega(1.5,1)$	4:8:10	18.15	15.58	7.11	20.56	20.98	12.36	44.07	35.20	18.41	34.02	39.47	32.30
	10:15:25	42.92	31.67	13.50	28.58	35.37	27.35	81.44	71.46	44.20	55.80	70.28	72.08
	20:40:60	74.31	62.42	32.24	47.23	62.40	59.71	99.12	98.34	89.71	87.22	97.33	99.25
	10:8:4	23.78	25.62	22.72	16.28	23.78	25.98	31.26	34.73	35.73	19.45	30.58	38.04
	25:15:10	49.42	44.52	36.54	27.11	40.44	47.74	65.50	61.80	58.20	37.96	55.22	67.94
	60:40:20	81.08	75.54	65.39	49.98	70.00	77.16	90.89	88.30	86.50	63.11	83.21	93.16
$t(5)$	4:8:10	22.34	10.18	4.26	15.68	14.65	10.04	44.56	25.85	11.38	26.52	30.60	22.32
	10:15:25	48.11	19.88	7.95	21.89	24.50	19.40	80.74	58.30	32.29	46.91	59.78	55.79
	20:40:60	74.70	44.01	20.30	36.81	47.21	42.80	98.34	95.35	76.65	82.06	94.80	95.08
	10:8:4	25.26	20.46	19.01	12.95	19.17	20.79	32.65	30.31	33.05	16.24	26.00	30.83
	25:15:10	50.46	35.44	30.58	21.34	32.22	36.75	65.88	53.98	51.68	31.40	47.92	55.44
	60:40:20	78.26	65.08	54.16	43.49	61.68	63.08	89.74	83.24	77.94	58.15	78.27	87.11

Table 6: Power of tests (%) for unbalanced data, $k = 3$