On Identifying a Robust Alternative Method for ANOVA with Unequal Variance

O.P. Adebayo[1*]; O.A. Ogunjimi[2]; and I. Ahmed[3]

[1]Department of Statistics, Phoenix University, Agwada, Nasarawa State, Nigeria.
[2]Department of Computer Science, Phoenix University, Agwada, Nasarawa State, Nigeria.
[3]Department of Statistics, Nasarawa State University, Keffi, Nasarawa State, Nigeria.

E-mail: adebayo.patrick@phoenixuniversity.edu.ng*
ogunjimi.olalekan@phoenixuniversity.edu.ng
ibrahimloko@nsuk.edu.ng

ABSTRACT

Six chosen procedures were compared with ANOVA, which was typically used as a backup to ANOVA (Analysis of Variance) in cases where the assumption of equal variance was not met. The tests Alexander-Govern, Brown-Forsythe, James second order, Welch's heteroscedastic F, Kruskal Wallis, Welch's heteroscedastic F with trimmed means, and Winsorized variances tests are the six procedures that were chosen. The six procedures and ANOVA were compared using simulated data, type I error rate, and test power to determine which of the two was the most robust. The criteria were: normally distributed sample; balance and imbalance; small and large sample; and different significant levels (0.01, 0.025, and 0.05).

The simulation's outcome demonstrates that Welch's heteroscedastic F test with trimmed means and Winsorized variances test is preferred over others. The simulation result shows that Welch's heteroscedastic F test, Alexander-Govern test, and Welch's heteroscedastic F test with trimmed means and Winsorized variances test are almost the same.

(Keywords: heteroscedastic, type 1 error rate, power of the test, robust, normal distribution)

INTRODUCTION

The One-way Analysis of Variance (ANOVA) is a procedure for testing the hypothesis that K population means are equal, where K greater than two. The One-way ANOVA compares the means of the samples or groups in order to make inferences about the population means. The One-way ANOVA is also called a single factor analysis of variance because there is only one independent variable or factor. As the name suggests, the one-way ANOVA is suitable for experiments with only one independent variable (factor) with two or more levels. . For example, a researcher may want to compare teaching methods of four different teachers using marks obtained by twenty students in an examination.

The null hypothesis (H_0) tested in the One-way ANOVA is that the population means from which the m samples are selected are equal. Or that each of the group means is equal. H_0: $\mu_1 = \mu_2 = . . . = \mu_m$, Where m is the number of levels of the independent variable. For example: If the independent variable has three levels – we would write H_0: $\mu_1 = \mu_2 = \mu_3$ If the independent variable has five levels – we would write H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ .The subscripts could be replaced with group indicators. From the example cited earlier : H_0: $\mu_{Method1} = \mu_{Method2} = \mu_{Method3} = \mu_{Method4}$ The alternative hypothesis (H_1) is that at least one group mean significantly differs from the other group means. Or that at least two of the group means are significantly different. H_1: $\mu_i \neq \mu_m$ Where i and m simply indicate unique groups.

Analysis of variance (ANOVA) has three basic assumptions which are equality of variances, normality and independence. Lantz (2013) discovered that ANOVA is more powerful if the assumptions of normality and variance homogeneity hold. Non – normality has minimal effect on the type I error when the variances are equal, but when the variance are not equal, ANOVA provides poor control over the type I and II error rates (Bishop, 1976). When the assumption of variance homogeneity was violated, Lix et al. (1996) found that parametric alternatives to the ANOVA F test performed better. The most popular heteroscedastic alternatives to the ANOVA F test include Welch's (1951) test, James' (1951) second-order approach, Brown and Forsythe's (1947a) test, and Alexander and Govern's (1994) test.

Welch (1951) provides one of the most well-known parametric alternatives to the ANOVA. It has been widely used and is incorporated in statistical software. However, simulation studies (Dijkstra and Werter, 1981; Wilcox, 1988, 1989; Alexander and Govern, 1994; Hsiung, *et al*., 1994; Oshima and Algina, 1992; Júlio, 2023) reveal that James's (1951) second-order test appears to be the most accurate method over a wide variety of realistic conditions. One big disadvantage is the computational complexity.

James (1951) presented two approaches for altering the critical value: first and second-order procedures. However, for small sample sizes, James' first-order solution does not control the rate of Type I errors in the present of variance heterogeneity (Brown and Forsythe, 1974b).

Welch's (1951) and James's (1951) tests can be employed when the variance homogeneity assumption is not met, but they should be avoided if the data are moderately to heavily skewed, even in balanced designs (Clinch and Keselman, 1982; Wilcox, *et al*., 1986; Lix, *et al.*, 1996).

Alexander-Govern's (1994) process (Lix, *et al.*, 1996) appears to be a rival of James' (1951) second-order and Welch's (1951) tests, since it is said to have many characteristics with the James method. A second adjustment to the Brown-Forsythe (1974b) test was proposed by Rubin (1983) and later by Mehrotra (1997). A comparison of Alexander-Govern's, ANOVA, Kruskal-Wallis, Welch's, Brown-Forsythe's, and James's second-order tests found that, under variance heterogeneity, Alexander-Govern's approximation was comparable to Welch's test and James's second-order test, and, in some cases, superior (Schneider and Penfield, 1997).

The same study discovered that the Alexander-Govern test is liberal when the distribution is very skewed and conservative when it is platykurtic. Wilcox (1997) also found comparable results. Schneider and Penfield (1997) recommend the Alexander-Govern procedure as the best alternative to the ANOVA F test when variances are heterogeneous for three reasons: (1) it is computationally simpler; (2) it is generally superior under most experimental conditions; and (3) the questionable results of Welch's test when more than four treatment groups are investigated (Dijkstra and Werter, 1981; Wilcox, 1988).

In the light of this, many scholars have come up with alternative procedures that can be used when the assumption of equality of variance failed. Therefore, out of the numerous procedures the following are selected for this study: Alexander-Govern test, Brown-Forsythe test, James second order test, Welch's heteroscedastic F test, Kruskal Walli test, and Welch's heteroscedastic F test with trimmed means and Winsorized variances. The aim of this study is to compare all the aforementioned procedures with ANOVA using type I error rate and power of the test to identify the robust one out of the selected procedures.

ANOVA

The ANOVA test statistic:

$$F = \frac{\frac{\sum_{j=1}^{n} n_j(\bar{X}_{.j}-\bar{X}_{..})^2}{k-1}}{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{ij}-\bar{X}_{.j})^2}{N-K}} \qquad \textbf{(1)}$$

The above follows an F distribution with K – 1 degrees of freedom for the numerator and N – K degrees of freedom for the denominator.

K : is the number of groups

N: is the total number of observations

n_j: is the number of observations in the j^{th} groups.

X_{ij} is the ith observations in the (i = 1,2 , . . . ,n_j) in the jth group (j =1,2 , . . .,k) , $\sum_{j=1}^{n} n_j = N,$

$\bar{X}_{..}$ is the overall mean, where $\bar{x}_{..} = \frac{\sum_{j=1}^{n} n_j(\bar{x}_{.j})}{N}$ and $\bar{x}_{.j}$ is sample mean for the jth group, $X_{.j} = \frac{\sum_{i=1}^{n} X_{ij}}{n_{j.}}$.

Welch's Heteroscedastic F Test

Welch (1951) proposed a heteroscedastic alternative to ANOVA that is robust to the violation of variance homogeneity assumption. Under the null hypothesis $H_0: \mu_1 = \mu_2 = \ldots = \mu_{k,}$ the test statistic F_w,

$$F_w = \frac{\frac{\sum_{j=1}^{n} w_j(\bar{X}_{.j} - X'_{..})^2}{k-1}}{\left[1 + \frac{2}{3}((k-2)v)\right]}, \quad (2)$$

Follows an F distribution with degrees of freedom k – 1 for the numerator and $\frac{1}{v}$ degrees of freedom for the denominator. In Equation (2):

$$w_j = \frac{n_j}{S_j^2},$$

$$S_j^2 = \frac{\sum_{i-1}^{n}(X_{ij} - \bar{X}_{.j})^2}{n_j - 1},$$

$$X'_{..} = \frac{\sum_{j=1}^{n} w_j \bar{X}_{.j}}{\sum_{j=1}^{n} w_j}$$

And $$v = \frac{3\sum_{j=1}^{n}\left[\frac{\left(1 - \frac{w_j}{\sum_{j-1}^{n} w_j}\right)^2}{(n_j - 1)}\right]}{k^2 - 1}$$

Welch's heteroscedastic F test is robust and less sensitive to heteroscedasticity when compared to the ANOVA as within groups variance is based on the relationship between the different sample sizes in the different groups instead of a simple pooled variance estimate (Lantz, 2013).

Welch's Heteroscedastic F Test with Trimmed Means and Winsorized Variances

Welch's heteroscedastic F test with trimmed means and Winsorized variance (welch, 1951) is a robust procedure that tests the equality of means by substituting trimmed means and Winsorized variances for the usual means and variances. Therefore, this test statistic is relatively insensitive to the combined effects of non – normality and variance heterogeneity (Keselman, *et al.*, 2008).

Let $X_{(1)j} \le X_{(2)j} \le \ldots \le X_{(nj)j}$ be the ordered observations in the jth group and $g_j = \|\epsilon n_j\|$, ϵ is the proportion to be trimmed in each tail of the distribution. After trimming, the effective sample size for the jth group becomes $h_j = n_j - 2g_j$. Then jth sample trimmed mean is:

$$\bar{X}_{ij} = \frac{1}{h_j}\sum_{i=g_i+1}^{n_j - g_j} X_{(i)j'} \quad (3)$$

and jth sample Winsorized mean is:

$$\bar{X}_{wj} = \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ij;} \quad (4)$$

Where,

$$Y_{ij} = \begin{cases} X_{(g_i+1)_{i'}} & \text{if } X_{ij} \le X_{(g_i+1)_{i'}} \\ X_{ij} & \text{if } X_{(g_i+1)_{i'}} < X_{ij} < X_{(n_j-g_i)_{i'}} \\ X_{(n_j-g_i)_{i'}} & \text{if } X_{ij} \ge X_{(n_j-g_i)_{i'}} \end{cases}$$

The sample Winsorized variance is:

$$S_{wj}^2 = \frac{1}{(n_j - 1)}\sum_{i=1}^{n_j}(Y_{ij} - \bar{X}_{wj})^2 \quad (5)$$

Let,

$$q_j = \frac{(n_j - 1)S_{wj}^2}{h_j(h_j - 1)},$$

$$w_j = \frac{1}{q_j},$$

$$\mu = \sum_{j=1}^{n} w_j,$$

$$\bar{X} = \frac{1}{\mu}\sum_{j=1}^{n} w_j \bar{X}_{ij},$$

$$A = \frac{1}{k-1}\sum_{j=1}^{n} w_j(\bar{X}_{ij} - \bar{X})^2,$$

$$B = \frac{2(k-2)}{k^2 - 1}\sum_{j=1}^{n}\frac{(1 - \frac{w_j}{\mu})^2}{h_j - 1}$$

Under $H_o: \mu_{i1} = \mu_{i2} = \ldots \mu_{ik,}$ Welch's heteroscedastic F test with trimmed means and Winsorized variances $F_{wi,}$

$$F_{wi} = \frac{A}{B+1}, \quad \textbf{(6)}$$

Follows an approximately F distribution with $k-1$ and v' degrees of freedom, where v' is:

$$v' = \left(\frac{3}{k^2-1}\sum_{j=1}^{n}\frac{(1-\frac{w_j}{\mu})^2}{h_j-1}\right)^{-1} \quad \textbf{(7)}$$

F_{wi} is not less sensitive to heteroscedasticity and non – normality but also robust to the negative effects of outlier as it utilizes the trimmed means and Winsorised variances (Keselman, *et al*., 2008).

Brown – Forsythe Test

Brown and Forsythe (1974 a,b.) proposed the following test statistic:

$$F_{BF} = \frac{\sum_{j=1}^{n} n_{.j}\left(\bar{X}_{.j}-\bar{X}_{..}\right)^2}{\sum_{j=1}^{n}\left(1-\frac{n_j}{N}\right)S_j^2} \quad \textbf{(8)}$$

F_{BF} statistic has an approximately F distribution with $k-1$ and f degrees of freedom, $\left(F_{k-1,f}\right)$ where f is obtained as:

$$f = \left(\sum_{j=1}^{n}\frac{c_j^2}{(n_j-1)}\right)^{-1},$$

and

$$c_j = \frac{\left(1-\frac{n_j}{N}\right)S_j^2}{\left[\sum_{j=1}^{n}\left(1-\frac{n_j}{N}\right)S_j^2\right]}$$

F_{BF} is a modification of ANOVA which has the same numerators as the ANOVA but an altered denominator. This test is more powerful when some variances appear unusually low (Brown and Forsythe, 1974a)

The Alexander – Govern Test

The Alexander-Govern test is a test introduced by Alexander-Govern (1994) and it uses mean as its central tendency measure. For a normal data, it possesses good control of Type I error rates and gives high power under variance heterogeneity but is not robust to non-normal data. This test is used for comparing two or more groups and its test statistic is obtained by using the following procedure.

The technique to obtain the test statistic for the Alexander-Govern test begins by ordering the data distribution, with population $(j = 1, \ldots, J)$. For each of the data distribution, the mean is obtained using:

$$\bar{X} = \frac{\sum_j X_{ij}}{n_j} \quad \textbf{(9)}$$

Where X_{ij} represents the observed ordered random samples of the data set and n_j is the sample sizes of the observations. The mean represents the measure of central tendency in the Alexander-Govern (1994) method. After the mean is obtained, the estimate of the usual unbiased variance can be calculated by using:

$$S_{.j}^2 = \frac{\sum\left(X_{ij}-\bar{X}_{.j}\right)^2}{\left(n_j-1\right)} \quad \textbf{(10)}$$

Where $\bar{X}_{.j}$ is used for estimating μ_j for the population (j). The standard error of the mean is obtained for each of the groups by using:

$$S_{ej} = \left[\frac{S_{.j}^2}{n_j}\right]^{\frac{1}{2}} \quad \textbf{(11)}$$

The weight (w_j) for the group sizes with j population of the ordered sample data is defined where $\sum w_j$ must be equal to 1. Therefore, the weight (w_j) for each of the group is obtained by using the formula below:

$$w_j = \frac{\frac{1}{S_{ej}{}^2}}{\sum_j \frac{1}{S_{ej}{}^2}} \quad \textbf{(12)}$$

The null hypothesis testing for the Alexander-Govern (1994) procedure, for the equality of the mean, under variance heterogeneity is defined as:

$H_0: \mu_1 = \mu_2 = \ldots = \mu_j$

Vs

$H_0: \mu_1 \neq . . . \mu_2, j = 1, . . . , J$

The alternative hypothesis disclaims the statement made by the null hypothesis. The variance weighted estimate of the total means for all the groups in the ordered data set, is obtained by using:

$$\hat{\mu} = \sum_{j=1}^{j} w_j \bar{X}_j \quad \textbf{(13)}$$

where w_j represent the weight for each of the group in the data distribution and $\bar{\bar{X}}_j$ is the mean of each of the groups in the ordered data set. The t statistic for each group is obtained by using:

$$t_j = \frac{\bar{X}_j - \hat{\mu}}{S_{ej}} \quad \textbf{(14)}$$

Where $\bar{X}_j$ is the mean for each of the groups

$\hat{\mu}$ is the grand mean for all the groups under analysis

and S_{ej}denotes the standard error of the mean for each of the groups in the population j.

The t statistic is distributed as a t variable with $n_j - 1$ degrees of freedom for . where v is the degree of freedom for each of the groups in the ordered data set. The t statistic is obtained for each of the groups and is transformed to a standard, normal deviates by using the Hill's (1970) normalization approximation in the Alexander-Govern (1994) method. The formula is defined as:

$$Z_j = c + \frac{[c^3 + 3c]}{b} - \frac{[4c^7 + 33c^5 + 240c^3 + 855c]}{[10b^3 + 8bc^4 + 1000b]},$$

$$c = \left[a \times \log_e \left(1 + \frac{t_j^2}{v_j}\right)\right]^{\frac{1}{2}}$$

Where $v_j = n_j - 1, a = v_j - 0.5, b = 48a^2$, the test statistic for the Alexander-Govern method is defined as:

$$A = \sum_{j=1}^{J} Z_j^2 \quad \textbf{(15)}$$

The level of significance used for the Alexander-Govern, test statistic is $\alpha = 0.05$, with $j - 1$ chi-square degree of freedom. The critical value is obtained by using the chi-square distribution table. Thereafter, we compare our chi-square value with the test statistic value of the Alexander-Govern test. Supposed if A obtained is greater than the value of the test from the chi-square distribution table, we reject H_o and conclude that the means of the different groups are different from each other, otherwise we do not reject H_o.

James's (1951) Test

An alternative test to ANOVA was proposed by James (1951). This test statistic (J) is:

$$J = \sum_j t_j^2, \quad \textbf{(16)}$$

where $t_j = \frac{\bar{X}_j - \hat{\mu}}{S_{ej}}$

The test statistic, J, is compared to a critical value, $h(\alpha)$, where:

$$h(\alpha) = r + \frac{1}{2}(3\chi_4 + \chi_2)T + \frac{1}{16}(3\chi_4 + \chi_2)^2\left(1 - \frac{k-3}{r}\right)T^2$$
$$+\frac{1}{2}(3\chi_4 + \chi_2)(8R_{23} - 10R_{22} + 4R_{21} - 6R_{12}^2 + 8R_{12}R_{11} - 4R_{11}^2)$$
$$+(2R_{23} - 4R_{22} + 2R_{21} - 2R_{12}^2 + 4R_{12}R_{11} - 2R_{11}^2)(\chi_2 - 1)$$

$$+\frac{1}{4}(-R_{12}^2+4R_{12}R_{11}-2R_{12}R_{10} -4R_{11}^2+4R_{11}R_{10} -R_{10}^2)(3\chi_4-2\chi_2 -1)$$

$$+(R_{23}-3R_{22}+3R_{21}-R_{20})(5\chi_6 +2\chi_4+\chi_2)$$

$$+\frac{3}{16}(R_{12}^2-4R_{23}+6R_{22}-4R_{21} +R_{20})(35\chi_9+15\chi_6 +9\chi_4+5\chi_2)$$

$$+\frac{1}{16}(-2R_{22}+4R_{21}-R_{20} +2R_{12}R_{10}-4R_{11}R_{10} +R_{10}^2)(9\chi_8-3\chi_6 -5\chi_4-\chi_2)$$

$$+\frac{1}{4}(-R_{22}+R_{11}^2)(27\chi_8+3\chi_6+\chi_4+\chi_2) +\frac{1}{4}(R_{23}-R_{12}R_{11})(45\chi_8 +9\chi_6+7\chi_4+3\chi_2).$$

with r denoting the $1-\alpha$ quantile of a χ_{k-1}^2 distribution and with:

$$\chi_{2s} = \frac{r^s}{(k-1)(k+1)(k+3)\ldots(k+2s-3)} \quad \textbf{(17)}$$

$$R_{st}=\sum_{i=1}^{k}\frac{\left(\frac{w_i}{W}\right)^t}{(n_i-1)^s} \quad \textbf{(18)}$$

The first order-approximation to the critical value for U, $h1(\alpha)$, is given by (J1 test):

$$h1(\alpha)=r+\frac{1}{2}(3\chi_4+\chi_2)\sum_{i=1}^{k}\frac{\left(1-\frac{w_i}{W}\right)^2}{n_i-1} \quad \textbf{(19)}$$

this can be rewritten as:

$$h1(\alpha)=r+\frac{r}{2(k-1)}\left(1+\frac{3r}{k+1}\right)\sum_{i=1}^{k}\frac{\left(1-\frac{w_i}{W}\right)^2}{n_i-1} \quad \textbf{(20)}$$

Simulation Study

The stimulation study will be carried out by using R package to compare the type I error rate and power of the test of the selected procedures namely: ANOVA, Alexander-Govern test, Brown-Forsythe test, James second order test, Welch's heteroscedastic F test, Kruskal Walli test and Welch's heteroscedastic F test with trimmed means and Winsorized variances test. The simulations are performed when.

(i) there is unequal variance $\left(\Sigma_1 \neq \Sigma_2 \neq \ldots \neq \Sigma_j\right)$

(ii.) the $H_o\left(\mu_1=\mu_2=\ldots=\mu_j\right)$ is true

(iii.) the data are normal

(iv) sample size are the same $(n_1=n_2=\ldots=n_k)$ and not the same$(n_1 \neq n_2 \neq \ldots \neq n_k)$.

(v) significant level α 0.05.

All these were replicated one thousand times (1,000) and the result obtained were presented in both graphical and tabular forms.

Table 1: Sample Size Considered in the Simulation.

Sample Size			
$n_1 = n_2 = \ldots = n_k$		$n_1 \neq n_2 \neq \ldots \neq n_k$.	
Three level	Five level	Three level	Five level
10,10,10	10,10,10,10,10	20,15,10	20,10,15,15,30
50,50,50	50,50,50,50,50	50,30,25	50,30,25,25,70
100,100,100	100,100,100,100,100	100,70,50	100,70,50,50,200
400,400,400	400,400,400,400,400	500,300,250	400,300,200,200,350

RESULTS

Examine the Performances of the Selected Procedures

Table 2: Type 1 Error Rate when the Sample Size are not equal at $\alpha = 0.05$.

Sample n	Anova	AlexG	BrownF	James	Welch	WelchT	Kruskal
Level = 3 rnorm(n1,3,8), rnorm(n2,3,38), rnorm(n3,3,8) ,							
10	0.080	0.041	0.058	0.082	0.044	0.048	0.058
50	0.087	0.054	0.082	0.060	0.053	0.050	0.078
100	0.074	0.056	0.073	0.064	0.057	0.052	0.083
400	0.090	0.051	0.088	0.053	0.052	0.063	0.081

Table 3: Power of the Test when the Sample Size are not equal at $\alpha = 0.05$.

Sample n	Anova	AlexG	BrownF	James	Welch	WelchT	Kruskal
Level = 3 rnorm(n1,3,8), rnorm(n2,3,38), rnorm(n3,3,8) ,							
10	0.895	0.955	0.895	0.956	0.955	0.947	0.945
50	0.908	0.948	0.908	0.948	0.948	0.949	0.938
100	0.903	0.943	0.903	0.944	0.944	0.939	0.948
400	0.886	0.938	0.886	0.938	0.938	0.939	0.955

Table 4: Type 1 Error Rate when the Sample Size are not equal at $\alpha = 0.05$.

Level =5	Anova	AlexG	BrownF	James	Welch	WelchTr	Kruskal
rnorm(n,3,8), rnorm(n,3,8), rnorm(n,3,78) , rnorm(n,3,28), rnorm(n,3,8)							
n=10	0.103	0.043	0.062	0.291	**0.043**	0.056	0.063
n=50	0.113	**0.056**	0.098	0.205	0.055	0.045	0.075
n=100	0.112	**0.037**	0.106	0.199	0.037	0.040	0.077
n=400	0.107	0.053	0.107	0.192	0.053	**0.049**	0.083

Table 5: Power of the Test when the Sample Size are not equal at $\alpha = 0.05$.

Level =5	Anova	AlexG	BrownF	James	Welch	WelchTr	Kruskal
rnorm(n,3,8), rnorm(n,3,8), rnorm(n,3,78) , rnorm(n,3,28), rnorm(n,3,8)							
n=10	0.791	0.960	0.797	0.999	**0.961**	0.955	0.942
n=50	0.784	**0.947**	0.784	0.999	0.947	0.955	0.945
n=100	0.780	**0.946**	0.781	0.998	0.946	0.945	0.947
n=400	0.798	0.943	0.799	1.000	0.944	**0.958**	0.942

Table 6: Type 1 Error Rate when the Sample Size are not equal at $\alpha = 0.05$.

Sample n	Anova	AlexG	BrownF	James	Welch	WelchTm	Kruskal
Level = 3 rnorm(n1,3,38), rnorm(n2,3,8), rnorm(n3,3,8)							
20,15,10	0.219	0.044	0.073	0.081	**0.049**	0.053	0.110
50,30,20	0.266	0.063	0.094	0.084	0.067	**0.069**	0.160
100,70,50	0.185	0.052	0.078	**0.058**	0.051	0.049	0.115
500,300,250	0.190	0.064	0.097	**0.064**	0.063	0.053	0.122

Table 7: Power of the Test when the Sample Size are not equal at $\alpha = 0.05$.

Sample n	Anova	AlexG	BrownF	James	Welch	WelchTm	Kruskal
Level = 3 rnorm(n1,3,38), rnorm(n2,3,8), rnorm(n3,3,8)							
20,15,10	0.908	0.956	0.857	0.957	**0.956**	0.953	0.945
50,30,20	0.922	0.965	0.871	0.966	0.965	**0.956**	0.971
100,70,50	0.882	0.931	0.827	**0.932**	0.931	0.940	0.954
500,300,250	0.892	0.953	0.853	**0.953**	0.953	0.967	0.957

Table 8: Type 1 Error Rate when the Sample Size are not equal at $\alpha = 0.05$.

Sample n	Anova	AlexG	BrownF	James	Welch	WelchT	Kruskal
Level = 5 rnorm(n1,3,8), rnorm(n2,3,8), rnorm(n3,3,78) , rnorm(n4,3,28), rnorm(n5,3,8)							
20,10,15,15,30	0.151	0.054	0.078	0.231	0.056	0.059	0.092
50,30,25,25,70	0.256	0.064	0.108	0.242	0.061	0.066	0.132
100,70,50,50,200	0.275	0.044	0.116	0.210	0.047	0.045	0.152
400,300,200,200,350	0.174	0.048	0.092	0.195	0.048	0.044	0.135

Table 9: Power of the Test when the Sample Size are not equal at $\alpha = 0.05$.

Sample n	Anova	AlexG	BrownF	James	Welch	WelchT	Kruskal
Level = 5 rnorm(n1,3,8), rnorm(n2,3,8), rnorm(n3,3,78) , rnorm(n4,3,28), rnorm(n5,3,8)							
20,10,15,15,30	0.829	0.955	0.797	0.998	0.956	0.955	0.956
50,30,25,25,70	0.865	0.946	0.781	0.999	0.946	0.948	0.961
100,70,50,50,200	0.887	0.950	0.792	0.994	0.950	0.949	0.973
400,300,200,200,350	0.861	0.944	0.786	0.998	0.944	0.949	0.961

DISCUSSION OF THE FINDINGS

When the sample size is very small (say, 10), the James test is liberal when the design is balanced or unbalanced and when the level is not greater than three (3), but when the sample size is very large (say, 400), the James test performed well in terms of power and type I error rate from Tables 2-9. James test type L error rate and power are quite high at all significant levels α when the level is increased to five.

When the design is balanced or unbalanced, as well as when the sample size is small or big, the Alexander-Govern test, Welch's heteroscedastic F test, and Welch's heteroscedastic F test with trimmed means and Winsorized variances type I error rate are conservative. When the sample size is very small, the Welch's heteroscedastic F test with trimmed means and Winsorized variances type I error rate is a little bit higher than the Alexander-Govern test and the Welch's heteroscedastic F test. Both the Alexander-Govern test and the Welch's heteroscedastic F test fared exceptionally well when the levels (3 and 5) of consideration, as well as sample size (small and large), balance, and unbalance, were taken into account.

When the sample size is very small and the level is three, the type I error rate for the Brown-Forsythe test is lower than that for the ANOVA; however, as the sample size grows, the type I error rates for both tests become almost equal when the dataset is balanced. However, if the design is unbalanced more effective than ANOVA is the Brown-Forsythe test. The power of the Brown Forsythe test and the ANOVA are equal at level 5 when the design is balanced.

In terms of type I error rate and test power, the Kruskal-Wallis test outperformed ANOVA (which is more conservative than ANOVA) at levels three and five (3, 5), significant levels 0.01 and 0.025 (α = 0.01, 0.025), and when the design is balanced. However, when the level is increased to five, the Kruskal-Wallis test outperforms ANOVA at all sample sizes. When the design is unbalanced, the test likewise performed fantastically better than ANOVA at = 0.01, 0.025, and 0.05.

ANOVA When the level is three for both balance and imbalance at all significant levels (0.01, 0.025, and 0.05), and when the sample is small or large, the test's power can be taken into account or successfully compete with other tests.

Alexander-Administer test Welch's heteroscedastic F and James test execution are practically a similar in term of power of the test when the level is three (3) for balance and unbalance design however when the level is increment to five (5) for both design James has most noteworthy power followed by Kruskal Walli test.

CONCLUSION

Another drawback of James' (1951) second-order test is its computational complexity. James' (1951) second-order test can only be used with very small sample sizes (say 10), whether the design is balanced and unbalanced under the unequal variance at all significant levels (Clinch and Keselman, 1982; Wilcox, *et al.,* 1986; Lix, *et al*., 1996).

The simulation indicates several similarities between the Alexander-Govern (1994) and Welch (1951) techniques. Mehrotra (1997). The Alexander-Govern's process, however, outperforms the Welch technique when the sample size is small, while the two procedures effectively compete when the sample size is large. Schneider and Penfield (1997) recommend the Alexander-Govern method as the best alternative to the ANOVA F test when variances are different.

REFERENCES

1. Alexander, R.A. and D.M. Govern. 1994. “A New and Simpler Approximation and ANOVA under Variance Heterogeneity”. *J Educ Stat*. 19:9.

2. Bathke, A. 2004. “The ANOVA F Test can still be used in some Balanced Designs with Unequal Variances and Nonnormal Data”. *J Stat Plan Inference*. 126:413–422.

3. Bishop, T.A. 1976. “Heteroscedastic ANOVA, MANOVA, and Multiple-Comparisons”. PhD thesis. The Ohio State University: Columbus, OH. [p176].

4. Brown, M.B. and A.B. Forsythe. 1974a. “The Small Sample Behavior of Some Statistics Which Test the Equality of Several Means”. *Technometrics*. 16(1):129–132. URL https://doi.org/10.2307/1267501. [p175, 176, 178]

5. Brown, M.B. and A.B. Forsythe. 1974b. “Robust Tests for the Equality of Variances”. *Journal of the American Statistical Association*, 69 (346):364–367,. [p175, 176, 178].

6. Clinch, J.J. and H.J. Keselman. 1982. “Parametric Alternatives to the Analysis of Variance”. *J Educ Stat*. 7: 207–214.

7. Dijkstra, J.B. and P.S. Werter. 1981. “Testing the Equality of Several Means when the Population Variances are Unequal”. *Commun Stat B-Simul*. 10: 557–569.

8. Hill, G.W. 1970. “Algorithm 395. Student’s t-Distribution”. *Communications of the ACM*. 13: 617-619.

9. Hsiung, T., S. Olejnik, and C.J. Huberty. 1994. “Comment on a Wilcox Test Statistic for Comparing Means when Variances are Unequal”. *J Educ Stat*. 19:111–118.

10. James, G.S. 1951. “The Comparison of Several Groups of Observations when the Ratios of the Population Variances are Unknown”. *Biometrika*. 38:324–329.

11. Júlio, A.M. 2023. “Alternatives Test in ANOVA with Unequal Variance and Unequal Sample Size”. *IJISRT*. 8(11).

12. Keselman, H., J. Algina, L.M. Lix, R.R. Wilcox, and K.N. Deering. 2008. “A Generally Robust Approach for Testing Hypotheses and Setting Confidence Intervals for Effect Sizes”. *Psychological Methods*. 13(2): 110–129. URL https://doi.org/10.1037/1082-989X.13.2.110. [p177]

13. Lantz, B. 2013. “The Impact of Sample Non-Normality on ANOVA and Alternative Methods”. *British Journal of Mathematical and Statistical Psychology*. 66(2): 224–244. URL https://doi.org/10.1111/j.2044- 8317.2012.02047.x. [p175, 176].

14. Lee, S. and C.H. Ahn. 2003. “Modified ANOVA for Unequal Variances”. *Commun. Statist. Simul.Computa*. 32: 987–1004.

15. Linnik, Y.A.V. 1968. “Statistical Problems with Nuisance Parameters”. *Translations of Mathematical Monographs*, No. 20. American Mathematical Society: New York, NY.

16. Lix, L.M., J.C. Keselman, and H.J. Keselman. 1996. “Consequences of Assumption Violations Revisited, A Quantitative Review of Alternatives to the One-Way Analysis of Variance F Test”. *Rev Educ Res*. 66: 579–619.

17. Mehrotra, D.V. 1997. “Improving the Brown-Forsythe Solution to the Generalized Behrens-Fisher Problem”. *Commun Stat-Simul*. 26: 1139–1145.

18. Oshima, T.C. and J. Algina. 1992. “Type I Error Rates for James’s Second-Order Test and Wilcox’s Hm Test under Heteroscedasticity and Non-Normality”. *Br J Math Stat Psychol*. 45:255–263.

19. Rubin, A.S. 1983 “The Use of Weighted Contrasts in Analysis of Models with Heterogeneity of Variance”. *P Bus Eco Stat Am Stat Assoc*. 347–352.

20. Schneider, P.J. and D.A. Penfield. 1997. “Alexander and Govern’s Approximation, Providing an Alternative to ANOVA under Variance Heterogeneity”. *J Exp Educ*. 65: 271–286.

21. Welch, B.L. 1947. “The Generalization of Student’s Problem when Several Different Population Variances are Involved”. *Biometrika*. 34: 28–35.

22. Welch, B.L. 1951. “On the Comparison of Several Mean Values: An Alternative Approach”. *Biometrika*. 38: 330–336.

23. Wilcox, R.R., V.I. Charlin, and K.L. Thompson. 1986. “New Monte Carlo Results on the Robustness of the ANOVA F, W and F-Statistics”. *Commun Stat Simul C*. 15: 933–943.

24. Wilcox, R.R. 1988. “A New Alternative to the ANOVA F and New Results on James’ Second-Order Method”. *Br J Math Stat Psychl*. 41: 109–117.

25. Wilcox, R.R. 1989. “Adjusting for Unequal Variances when Comparing Means in One-Way and Two-Way Fixed Effects ANOVA Models”. *J Educ Stat* 14: 269–278.

SUGGESTED CITATION

Adebayo, O.P., O.A. Ogunjimi, and I. Ahmed. 2024. “On Identifying a Robust Alternative Method for ANOVA with Unequal Variance”. *Pacific Journal of Science and Technology*. 25(2): 45-54.

Pacific Journal of Science and Technology