Ilorin Journal of Science
Volume 2, Number 1, 2015, pp.68–79 (Printed in Nigeria)
ISSN: 2408 – 4840 © 2015 Faculty of Physical Sciences, University of Ilorin
https://doi.org/10.54908/iljs.2015.02.01.006

ILORIN
JOURNAL OF SCEINCE

ILJS-15-031

Development of Hypothesis Testing On Type One Error and Power Function

Abidoye*, A. O., Jolayemi, E. T., Sanni, O. O. M. and Oyejola, B. A.

Department of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria.

Abstract

In this study the proposed test statistic for testing hypothesis of equality of means against ordered alternatives under heterogeneous variances was proposed. We compared it with the conventional t – test using the sample pooled variance. The comparisons of the two tests were made using type 1 error and power under non – overlapping situations. The result shows that proposed test performed better with respect to type 1 error rate. Power of the proposed test was found to be consistently higher than the conventional t – test.

Keyword: Type I error, Power, heterogeneous variances, unequal variance, conventional t –test, proposed t – test.

1. INTRODUCTION

In statistical test theory the notion of statistical error is an integral part of hypothesis testing. The test requires an unambiguous statement of a null hypothesis, which usually corresponds to a default “state of nature”. Due to the statistical nature of a test, the result is never, except in very rare cases, free of error. Two types of error are distinguished: type I and type II errors.

A type I error, also known as an error of the first kind, occurs when the null hypothesis (H_0)

Corresponding Author: Abidoye, O. A.
Email: abidoye@unilorin.edu.ng

is true, but is rejected. A type I error may be compared with so called false negative (result that a given condition is absent when actually it is present).

The value of the type I error is called the size of the test and denoted by the Greek letter α (alpha). It usually equals the significance level of a test. In the case of a simple null hypothesis, α is the probability of a type I error. If the null hypothesis is composite, α is the maximum (supremum) of the possible probabilities of a type I error. See Abidoye (2012). A type II error, also known as error of the second kind, occurs when the null hypothesis is false, but it is erroneously accepted as being true. The type II error may be compared with the so – called false negative (where an actual 'hit' is disregarded by the test and seen as a 'miss'). The type II error is committed, if we accept the null hypothesis when it is actually false.

The value rate of the type II error is denoted by the Greek letter β (beta) and related to the power of a test. What is actually called type I or type II error depends directly on the null hypothesis. Therefore, negation of the null hypothesis causes type I and type II errors to switch roles see Shemer (2002) and Abidoye (2012). The goal of the test is to determine if the null hypothesis can be rejected. A statistical test can either reject or fail to reject the null hypothesis. However, the test has not proved the hypothesis to be true. See Fisher (1935), Ott (1984), Johckheere (1954), Montgomery (1981), Dunnett (1964), Dunnett and Tamhane (1997), Gupta *et al.* (2006) and Abidoye (2012).

If β is the probability of making the type II error, then (1- β) is the probability of correctly rejecting the null hypothesis (H_0) when a specific alternative hypothesis (H_0) is provided. A test's ability to correctly reject the null, when an alternate hypothesis is true, is its power.

2. METHODOLOGY

We proposed a suitable test procedure to test the hypothesis

$H_0 : \mu_i = \mu \quad \forall_i$ versus $H_1 : \mu_1 > \mu_2 > \mu_3 > ... > \mu_g$

or $H_0 : \mu_i = \mu \quad \forall_i$ against $H_1 : \mu_1 < \mu_2 < \mu_3 < ... < \mu_g$ See Abidoye (2012).

Using the proposed t - test Statistic defined in Abidoye (2012)

$$Y = \min(\overline{Y}_i - \overline{Y}_{i+1}) \sim \lambda N(\mu_i - \mu_{i+1}, \frac{\sigma_i^2}{n_i} + \frac{\sigma_{i+1}^2}{n_{i+1}}) \tag{2.1}$$

where Y_i is the set of independent and normally distributed random variable.

μ_i is the mean of group i,

σ_i^2 is the variances of group i.

Then

$$t_H = \frac{\min(\overline{Y}_i - \overline{Y}_{i+1})}{S_H \sqrt{\frac{1}{n_i} + \frac{1}{n_{i+1}}}} \sim \lambda t_r \tag{2.2}$$

where

$$S_H^2 = \hat{\sigma}_H^2 \tag{2.3}$$

then $\overline{Y}_i$ *and* $\overline{Y}_{i+1}$ are from ordered statistic

The conventional t – test Statistic defined in Yahya and Jolayemi (2003)

$$t = \frac{\min(\overline{Y}_i - \overline{Y}_{i+1})}{S_p \sqrt{\frac{1}{n_i} + \frac{1}{n_{i+1}}}} \sim t_{n-k,\alpha} \tag{2.4}$$

where S_P^2 is the conventional pooled variance. This is the Statistic closest to the proposed statistic even though it is inappropriate as contained in Behrens – Fisher problem see Behrens (1964) and Tamhane (1979). Since for most practical purposes $S_P^2 \geq S_H^2$ with equality when $\sigma_i^2 = \sigma^2 \ \forall_i$, $t_H \geq t$, and t would be a more conservative test compared to t_H.

3. SIMULATION OF TYPE I ERROR AND POWER

Experiments were carried out to obtain Type I error rate for proposed test statistic on each set of groups at different sample sizes but with equal sets of means $(\mu_1 = \mu_2 = ... = \mu_g)$ and different variances $(\sigma_1^2 \neq \sigma_2^2 \neq ... \neq \sigma_g^2)$. Experiments were also carried out to obtain power for both proposed and conventional test statistics. Two different situations were used; the first situation was for set of groups of different means $(\mu_1 < \mu_2 < ... < \mu_g)$ and varied variances $(\sigma_1^2 \neq \sigma_2^2 \neq ... \neq \sigma_g^2)$. We examined for each set of groups different sample sizes. Experiments were carried out with varied sample sizes and data for each sample size were generated 1000 times in order to obtain value for Type I error and power of the two tests.

In these experiments, varied values of means and the variances are considered. The results obtained are shown in the tables below:

Table 3.1: Computed type I error rates using the proposed statistic under the following groups: N(2,3), N(2,4), N(2,15), N(2,18) $\sigma_P^2 = 10.0$ $\sigma_H^2 = 5.7$ $\alpha = 0.05$

	Replicates										
Sample size	1	2	3	4	5	6	7	8	9	10	Average
10	0.054	0.0450	0.054	0.053	0.049	0.049	0.050	0.051	0.049	0.047	0.051
15	0.05	0.049	0.055	0.053	0.044	0.050	0.048	0.054	0.048	0.049	0.050
20	0.052	0.047	0.056	0.054	0.045	0.047	0.049	0.054	0.046	0.049	0.050
25	0.050	0.05	0.053	0.054	0.045	0.049	0.049	0.058	0.046	0.049	0.051
30	0.052	0.05	0.056	0.054	0.049	0.047	0.049	0.054	0.046	0.049	0.051
40	0.054	0.062	0.055	0.060	0.044	0.048	0.049	0.052	0.046	0.048	0.051
50	0.063	0.06	0.057	0.054	0.064	0.068	0.057	0.059	0.056	0.058	0.051
100	0.063	0.06	0.057	0.054	0.064	0.068	0.057	0.059	0.056	0.058	0.051

From Table 3.1, it is observed that on the average, the proposed test performs uniformly good even for small sample sizes since the type I error rates are close to the preselected significance level (α = 0.05). From Levene's test Statistic, P – value is obtained to be 0.034 which shows that the variances are unequal.

Table 3.2: Computed type I error rates using the following groups N(10,3), N(10,4), N(10,6), N(10,8), N(10,10), N(10,16), N(10,20), N(10,75) $\sigma_H^2 = 7.3$ $\sigma_P^2 = 17.8$ α = 0.05

		Replicates										
Sample size	Statistic	1	2	3	4	5	6	7	8	9	10	Average
10	Proposed	0.054	0.049	0.056	0.053	0.049	0.047	0.049	0.056	0.042	0.048	0.050
15	Proposed	0.05	0.05	0.057	0.053	0.045	0.047	0.049	0.055	0.047	0.049	0.050
20	Proposed	0.052	0.049	0.057	0.054	0.045	0.046	0.051	0.053	0.046	0.048	0.050
25	Proposed	0.05	0.051	0.053	0.054	0.044	0.047	0.05	0.053	0.048	0.049	0.050
30	Proposed	0.051	0.052	0.056	0.053	0.045	0.049	0.048	0.050	0.046	0.049	0.050
40	Proposed	0.055	0.049	0.056	0.053	0.046	0.047	0.05	0.050	0.047	0.048	0.050
50	Proposed	0.051	0.051	0.055	0.053	0.047	0.048	0.049	0.051	0.047	0.049	0.050
100	Proposed	0.051	0.051	0.055	0.053	0.047	0.048	0.049	0.051	0.047	0.049	0.050

From table 3.2 above, it is observed that on the average, that the proposed test performs uniformly good even for small sample sizes since the type I error rates are close to the preselected significance level (α = 0.05). From Levene's test Statistic, P – value is obtained to be 0.046 which shows that the variances are unequal.

Power Function

We want to consider a situation where the means are non-overlapping where

$H_0 : \mu_1 = \mu_2 = ... = \mu_g$ vs $H_1 : \mu_1 < \mu_2 < ... < \mu_g$ and when the variances are not equal that is, $\sigma_i^2 \neq \sigma_j^2, i \neq j$

Table 3.3: computed power is when some of the confidence intervals of the group means are non-overlapping.
N(3,3) N(7,4) N(11,6) N(17,75) $S_H^2 = 5.2$ $S_P^2 = 21.5$

		Replicates										
Sample size	Statistic	1	2	3	4	5	6	7	8	9	10	Average
10	Proposed	0.778	0.774	0.772	0.744	0.668	0.772	0.669	0.675	0.714	0.687	0.7245
	conventional	0.321	0.331	0.308	0.327	0.363	0.228	0.347	0.329	0.355	0.336	0.3245
15	Proposed	0.754	0.841	0.792	0.824	0.837	0.726	0.828	0.804	0.817	0.802	0.8025
	conventional	0.344	0.302	0.347	0.367	0.283	0.297	0.302	0.379	0.388	0.359	0.3368
20	Proposed	0.805	0.813	0.77	0.801	0.84	0.803	0.823	0.814	0.707	0.867	0.8043
	conventional	0.304	0.43	0.36	0.388	0.314	0.364	0.301	0.305	0.301	0.302	0.3369
25	Proposed	0.77	0.801	0.746	0.851	0.858	0.831	0.801	0.827	0.834	0.743	0.8062
	conventional	0.355	0.302	0.349	0.337	0.316	0.35	0.37	0.386	0.356	0.472	0.3593
30	Proposed	0.801	0.818	0.888	0.886	0.808	0.817	0.802	0.909	0.905	0.811	0.8445
	conventional	0.405	0.406	0.423	0.401	0.425	0.411	0.406	0.403	0.346	0.41	0.4036
50	Proposed	0.801	0.818	0.888	0.895	0.908	0.817	0.993	0.809	0.805	0.811	0.8545
	conventional	0.405	0.373	0.423	0.401	0.425	0.411	0.406	0.403	0.407	0.41	0.4064
100	Proposed	0.801	0.818	0.888	0.895	0.908	0.817	0.993	0.809	0.805	0.811	0.8545
	conventional	0.405	0.373	0.423	0.401	0.425	0.411	0.406	0.403	0.407	0.41	0.4064

It is observed from Table 3.3 that as the sample sizes increase the power for the proposed and conventional test also increase on the average, the power of our proposed test is more than twice that of conventional test, for all sample sizes.

The computed power rates using the groups: N(3,3), N(7,4), N(11,6) , N(17,75) is as reported in the figure below:

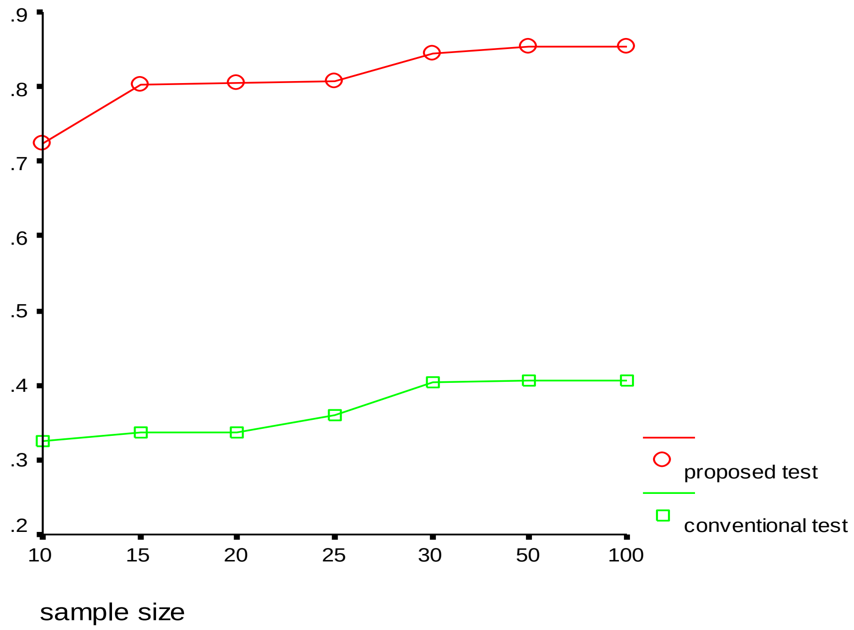

Fig.3.1: Shows the graph of both powers for the proposed and conventional tests.

Table 3.4: Computed power when some of the confidence intervals of the group means are non-overlapping. N(3,3) N(7,4) N(9,6) N(11,8) N(13,10) N(15,16) N(18,20) N(24,75) $\sigma_H^2 = 7.3$ $\sigma_P^2 = 17.8$

	Replicates										
Sample size	Statistic	1	2	3	4	5	6	7	8	9	10	Average
10	Proposed	0.792	0.794	0.793	0.799	0.702	0.782	0.788	0.791	0.831	0.721	0.779
	Conventional	0.354	0.369	0.329	0.342	0.341	0.361	0.338	0.321	0.314	0.301	0.337
15	Proposed	0.643	0.741	0.834	0.807	0.839	0.834	0.824	0.811	0.831	0.826	0.799
	Conventional	0.362	0.351	0.342	0.366	0.373	0.376	0.355	0.382	0.381	0.362	0.363
20	Proposed	0.807	0.839	0.832	0.862	0.746	0.844	0.836	0.831	0.844	0.839	0.828
	Conventional	0.386	0.372	0.391	0.399	0.377	0.412	0.392	0.376	0.399	0.386	0.389
25	Proposed	0.842	0.742	0.862	0.871	0.883	0.853	0.845	0.846	0.86	0.846	0.845
	Conventional	0.392	0.381	0.393	0.399	0.376	0.439	0.399	0.382	0.399	0.392	0.395
30	Proposed	0.873	0.767	0.892	0.881	0.796	0.872	0.881	0.892	0.87	0.866	0.859
	Conventional	0.396	0.384	0.397	0.401	0.381	0.463	0.399	0.392	0.406	0.402	0.402
50	Proposed	0.884	0.867	0.793	0.884	0.796	0.872	0.883	0.892	0.877	0.866	0.860
	Conventional	0.392	0.384	0.397	0.403	0.381	0.463	0.399	0.393	0.406	0.402	0.403
100	Proposed	0.884	0.867	0.793	0.884	0.796	0.872	0.883	0.892	0.877	0.866	0.860
	Conventional	0.392	0.384	0.397	0.403	0.381	0.463	0.399	0.393	0.406	0.402	0.403

It is observed that from table 3.4 that as the sample sizes increase the power for the proposed and conventional tests also increase even for large group. It is also observed that the power of our proposed test is on the average more than twice that of conventional test in virtually all cases considered.

The computed power rates using the groups: N(3,3), N(7,4), N(9,6) , N(11,8), N(13,10), N(15,16), N(18,20), N(24,75) is reported in the Figure 3.2 below:

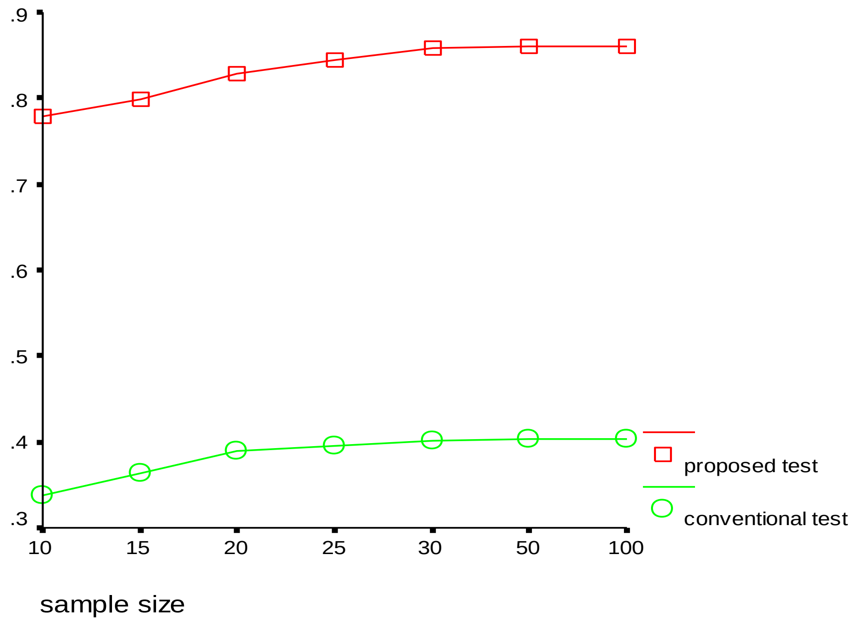

Fig. 3.2: Shows the graph of both powers for the proposed and conventional tests.

Table 3.5: Computed power when some of the confidence intervals of the group means are Overlapping.
N(3,3) N(7,4) N(9,6) N(11,8) N(13,10) N(19,75) $\sigma_H^2 = 6.1$ $\sigma_P^2 = 17.7$

		Replicates										
Sample size	Statistic	1	2	3	4	5	6	7	8	9	10	Average
10	Proposed	0.746	0.776	0.739	0.814	0.771	0.679	0.782	0.783	0.703	0.791	0.7584
	conventional	0.336	0.367	0.345	0.344	0.354	0.33	0.34	0.329	0.351	0.34	0.3436
15	Proposed	0.712	0.798	0.704	0.802	0.7	0.823	0.695	0.804	0.818	0.808	0.7656
	conventional	0.326	0.333	0.358	0.336	0.371	0.375	0.345	0.379	0.375	0.359	0.3545
20	Proposed	0.731	0.725	0.723	0.721	0.893	0.83	0.789	0.814	0.735	0.821	0.7782
	conventional	0.454	0.349	0.385	0.392	0.375	0.374	0.387	0.372	0.397	0.385	0.3688
25	Proposed	0.742	0.75	0.844	0.755	0.86	0.824	0.766	0.821	0.736	0.731	0.7829
	conventional	0.393	0.399	0.393	0.325	0.375	0.4	0.357	0.372	0.375	0.369	0.3758
30	Proposed	0.884	0.784	0.754	0.799	0.756	0.788	0.779	0.782	0.781	0.778	0.7885
	conventional	0.365	0.393	0.367	0.397	0.384	0.375	0.382	0.374	0.381	0.385	0.3803
50	Proposed	0.884	0.784	0.754	0.799	0.756	0.788	0.779	0.782	0.781	0.778	0.7885
	conventional	0.393	0.393	0.367	0.369	0.384	0.375	0.382	0.374	0.381	0.385	0.3803
100	Proposed	0.884	0.784	0.754	0.799	0.756	0.788	0.779	0.782	0.781	0.778	0.7885
	conventional	0.393	0.393	0.367	0.369	0.384	0.375	0.382	0.374	0.381	0.385	0.3803

The computed powers reported in table 3.5indicate that as the sample sizes increase the power for the proposed and conventional test increase under overlapping of confidence intervals. The power of our proposed test is on the average more than twice that of conventional test.

The computed power rates using the groups: N(3,3), N(7,4), N(9,6) , N(11,8), N(13,10), N(19,75) are show in Figure 3.3 that follows:

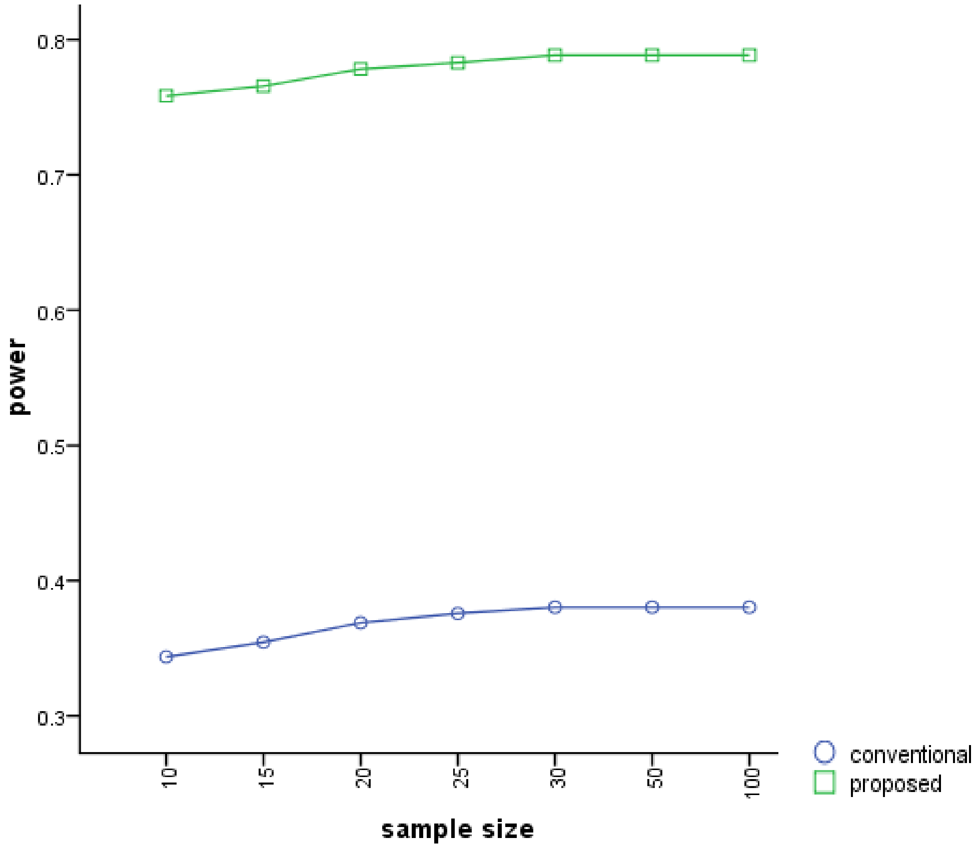

Fig.3.3: Shows the graph of both powers for the proposed and conventional tests.

4. CONCLUSION

In real life situations, we do come across situations where we have heterogeneous variances. Our proposed test is design to capture such situation. Since the conventional test do not seem to perform well. The proposed statistic performs better because the estimates are closer to α – level (0.05) of significance under type 1 error also we compared our proposed test with the conventional test using type 1 error rates as well as power of the test under overlapping and non – overlapping of confidence interval.

Empirically, our proposed test outperform the conventional test using both type 1 error rates and powers under both small and large sample sizes. Similar results are obtained under small group means and large group means. Our results, as would be expected, indicate that on the average the power increases as the sample size increases. The proposed test is therefore recommended for both type 1 error and power.

REFERENCE

Abidoye, A. O. (2012): *Development of Hypothesis Testing Technique for Ordered Alternatives under heterogeneous variances.* Unpublished Ph.D. Thesis submitted to Dept. of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria.

Behrens, W. U. (1964): The comparison of means of independent normal distribution with different variance. Biometrics, **20**, 16 – 27.

Dunnett, C. W. and Tamhane, A. C. (1997): Multiple testing to establish superiority/ equivalence of a new treatment compare with k standard treatments. Statistics in Medicine. **16**, 2489 – 2506.

Dunnett, C. W. (1964): New tables for multiple comparisons with a control. Biometric, **20**, 482-491.

Fisher, R. A. (1935): *The design of experiments.* Oliver and Boyd, Edinburgh, Scotland.

Gupta, A. K., Solomon, W. H. and Yasunori, F. (2006): Asymptotics for testing hypothesis In some Multivariate variance components model Under non-normality. Journal of Multivariate Analysis Archive. Vol. **97**, 148-178.

Jonckheere, A. R. (1954): A distribution-free k-sample test against ordered alternative. Biometrical, **41**, 133-145.

Montgomery, D. C. (1981): *Design and Analysis of Experiment.* Second Edition. John Wiley and sons Inc. New York.

Ott, L. (1984): *An Introduction to Statistical Methods and Data Analysis.* Second Edition. P. W. S Publisher. Boston.

Tamhane, A. C. (1979): A comparison of procedures for multiple comparisons of means with unequal variances. Journal of the American Statistical Association. **74**, 471 – 480.

Yahya, W. B. and Jolayemi, E. T. (2003): Testing Ordered Means against a Control. Journal of Nigerian Statistical Association, **16**, 40- 51.