A Multiple Comparison Procedure with Control under Unequal Variances

E. Alih*; S. E. OnaloI; S. O. Ige

Department of Mathematics and Statistics,
Federal Polytechnic Idah, Idah, Nigeria.
e-mail: ekelson2002@yahoo.com*

Abstract - In **this article, test procedures for multiple comparisons with a control under unequal variance was investigated. An algorithm for estimating the sample variances in the presence of heteroscedasticity. In this algorithm, a weight that is a function of Mahalanobis distances is used to down weight the influence of heteroscedasticity before conducting the multiple comparisons test. The merits of the proposed procedures alongside other alternatives is examined through a Monte Carlo Simulation experiment by computing the family-wise error rate at a specific nominal level, α. The result of the experiment indicates that conventional procedures assuming equal variance will have inflated error rate and may yield misleading inferences when the assumption of equal variance fails. The Monte Carlo simulation experiment also reveals that the proposed procedures always control the family-wise error rate at a specific nominal level α, while some established procedures are liberal and have inflated error rate in some scenarios.**

Keywords - *Multiple comparisons, Heteroscedasticity, Family-wise error rate, Mahalanobis Distance.*

I. INTRODUCTION

In regression analysis, the usual properties and assumptions of the Ordinary Least Squares (OLS) alongside its ease of computation and close form solution have made it so nice and attractive that it has become the common practice over the years to use it (OLS) as inferential tools in regression analysis. Among these attractive properties is the homogeneity of error variance known as homoscedasticity.

Homoscedasticity is an important assumption for which the OLS estimators enjoy minimum variance property. According to Carrol and Ruppert [1], there are many occasions where the assumption of homoscedastic error variance breakdown and heteroscedasticity sets in. Such occasions are:

***Cross sectional study*:** If one is examining a cross section of firms in one industry, error terms associated with very large firms might have large variance and vice versa. The inconsistency of the error variance from one firm to another in one industry leads to heteroscedasticity.

***Income growth*:** As income grows, people have more discretionary income and more scope for choice about the disposition of their income, hence σ_i^2 is likely to increase with income growth.

Model specification: When regression model is not correctly specified and a wrong functional form of regression analysis is adopted, σ_i^2 tend to diverge thereby setting in heteroscedasticity.

Skewness: Skewness in the distribution of one or more repressors' in the model leads to non- constant error variance. This occurs more frequently in data where the distribution of income and wealth is uneven.

Phase II clinical study: In a phase II clinical study, multiple groups with different dose levels are usually of interest to be compared with a control group to detect the effect size of the dose level. According to Li and Wei [8], one of the main goals of a phase II clinical study is to estimate the minimum effective dose (MED) and the maximum safe dose (MSD).

Under the usual normality and equality of variance assumption, Dunnett's method has been widely used to estimate the MED and/or MSD in dose–response studies. In the practical clinical study, the homogeneity of variance assumption is seldom satisfied since the variation of response under different dose levels is usually different with the change of dose levels because patients in different groups tend to respond differently due to some biological factors or the toxicity effect at various dose levels.

Methods that assume equal variances may lead to incorrect decision. The danger of homoscedasticity assumption is that an ineffective dose may be claimed as effective or an unsafe dose may be claimed as safe when the assumption is violated. Therefore, more flexible methods that do not assume equal variances are in great need.

II. LITERATURE

A. *Multiple comparisons*:

Under the usual normality and equality of variances assumption, many multiple comparison procedures have been developed and widely applied. Dunnett [2] proposed a well-known test for the multiple comparisons with a control (MCC). Tukey [12] provided the simultaneous confidence intervals (SCIs) for all-pairwise comparisons (MCA). Scheffe [9] pioneered the pivoting of the F-statistic to give SCI for all-contrast comparisons (ACC).

Tamhane [10] proposed two approximated approaches for the MCC and all-pairwise comparisons when the variances are unequal. Games and Howell [3] provided the approximate SCIs for all-pairwise differences under the heteroscedasticity. Hochberg [5] proposed an approximate approach for all-pairwise comparisons.

Tamhane [11] gave a brief review and comparison of these procedures. Hasler and Hothorn [4] proposed a procedure for the multiple contrast tests that provided approximated SCIs for ACC. Li [7] proposed an approximate method for all-pairwise comparisons with unequal variances.

In all of these works, effort has been based on conducting multiple comparison tests through construction of simultaneous confidence interval for all pairwise comparisons. However, heteroscedasticity is a property of the variance estimate which requires down weighting approach prior to the comparison test and hence in this paper, we propose the minimum Mahalanobis distance algorithm for computing variance estimate based on weights that are function of the data set itself.

III. METHODOLOGY

A. *Simultaneous Confidence Intervals*

Let X_{ij} be the observed measurement on j^{th} experimental material upon which i^{th} treatment is applied, and X_{ij}^{t} are identically and independently distributed as (μ_i, σ_i^2). Consider the one-way Analysis of Variance (ANOVA) model given as

$$X_{ij} = \mu_i + \varepsilon_{ij} \qquad (1)$$

where $i = 0,1, \dots, k$ and $j = 1,2, \dots, n_i$ $\varepsilon_i \sim N(0, \sigma_i^2)$ with $i = 0$ being the control group which can be either a placebo or an active control and $i = 1,2, \dots k$ is the active experimental groups. Denote the sample location and scale estimator of μ_i and σ_i^2 as $\bar{X}_i$and S_i^2 respectively,

$$\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \qquad (2)$$

and

$$S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(X_{ij} - \bar{X}_i\right)^2 \qquad (3)$$

where $S_i^2 \sim \left(\frac{\sigma_i^2}{v_i}\right)\chi_{v_i}^2, \quad v_i = n_i$

Note that Equations (2)-(3) are the least-squares estimator for μ_i and σ_i^2 respectively. Consider testing the differences between each new treatment group mean and the control group mean, $H_{0i}: \mu_i - \mu_0 \leq \delta \; vs \; H_{ai}: \mu_i - \mu_0 > \delta$, with $i = 1, \dots, k$ where δ is some given threshold constant. The test statistic is given by

$$t_i = \frac{\bar{X}_i - \bar{X}_0 - \delta}{\sqrt{\frac{S_i^2}{n_i} + \frac{S_0^2}{n_0}}} \qquad (4)$$

where S_0^2 and S_i^2 are the sample variances of the control and the new treatment group.

Several methods such as the ones in Tamhane [10] herein referred as TM, Li and Wei [8] herein referred to as LW and Dunnet [1] herein referred to as DM, and Hasler and Hothorn [4] usually construct the $100(1-\alpha)\%$ simultaneous confidence lower bound for $\mu_i - \mu_0$ is given by

$$\mu_i - \mu_0 > \bar{X}_i - \bar{X}_0 - \boldsymbol{\eta}\sqrt{\frac{S_i^2}{n_i} + \frac{S_0^2}{n_0}} \; for \; i = 1, \dots, k$$

where the critical value $\boldsymbol{\eta}$ is the solution of the equation

$$p\left(\frac{\bar{X}_i - \bar{X}_0 - (\mu_i - \mu_0)}{\sqrt{\frac{S_i^2}{n_i} + \frac{S_0^2}{n_0}}} < \boldsymbol{\eta}, i = 1, \dots, k\right) = 1 - \alpha \qquad (5)$$

However the simultaneous confidence interval may not be able to overcome the problem of heteroscedasticity since it affects the variances of each of the independent group and not the mean and hence this article proposed a method for estimating the joint covariance matrix of the treatment group through an algorithm referred to as minimum Mahalanobis distance algorithm, (MMD).

The methodology of the algorithm is that of developing resampling algorithm capable of partitioning data into a homoscedastic and heteroscedastic cluster through the Mahalanobis distance. Observations that are homoscedastic will have a small Mahalanobis distance while those that are heteroscedastic will have a large Mahalanobis distance. The inverse of this Mahalanobis

distance serves as weights that downweights the influence of heteroscedasticity in the multiple comparisons test. The algorithm is presented below:

B. *The Minimum Mahalanobis Distance Algorithm*

The algorithm below constructs a minimum Mahalanobis distance, MMD which becomes a metrics used to identify observations that are either homoscedastic or heteroscedastic in a set of multivariate data.

Let $\boldsymbol{Z}_{ki}$ be the $n \times k$ matrix containing the k active experimental groups in n sample points. Also, let h be the integer part of $(n+k+1)/2$ and $(n-1)^{-1}\sum_{i=1}^{n}(\boldsymbol{Z}_{ki}-\boldsymbol{m})(\boldsymbol{Z}_{ki}-\boldsymbol{m})^{T}$ where $\boldsymbol{m}$ is the vector of coordinatewise median such that $c_f = 1+\frac{k+1}{n-k}+\frac{1}{n-h-k}$ is a correction factor for the distance

$$MMD_i(\overline{\boldsymbol{Z}}_{kb},\boldsymbol{S}_b) = \sqrt{(\boldsymbol{Z}_{ki}-\overline{\boldsymbol{Z}}_{kb})^{T}\boldsymbol{S}_b^{-1}(\boldsymbol{Z}_{ki}-\overline{\boldsymbol{Z}}_{kb})},\quad i=1,\dots,n \qquad (6)$$

Where $\overline{\boldsymbol{Z}}_{kb}$ and $\boldsymbol{S}_b$ are the mean vector and covariance matrix of the basic subset.

Step 1:

(a) Compute $d_i(\boldsymbol{m},\boldsymbol{A}) = \sqrt{(\boldsymbol{Z}_{ki}-\boldsymbol{m})^{T}\boldsymbol{A}^{-1}(\boldsymbol{Z}_{ki}-\boldsymbol{m})}$ and rearrange $\boldsymbol{Z}_{ki}$ in ascending order $d_i(\boldsymbol{m},\boldsymbol{A})$

(b). Compute $d_i(\overline{\boldsymbol{Z}}_{kh},\boldsymbol{S}_h)$ where $\overline{\boldsymbol{Z}}_{kh}$ and $\boldsymbol{S}_h$ are the mean and covariance matrix of the observations with the h^{th} smallest values of $d_i(\boldsymbol{m},\boldsymbol{A})$

(c). Rearrange the n observations in ascending order according to $d_i(\overline{\boldsymbol{Z}}_{kh},\boldsymbol{S}_h)$ and divide the observations into two initial subsets: one subset includes the first $(k+1)$ observations and another subset containing the remaining $(n-k-1)$ observation. Refer to these subsets as the initial "Basic" and "Non-basic" subsets respectively.

Step2: Compute $MMD_i(\overline{\boldsymbol{Z}}_{kb},\boldsymbol{S}_b) = \sqrt{(\boldsymbol{Z}_{ki}-\overline{\boldsymbol{Z}}_{kb})^{T}\boldsymbol{S}_b^{-1}(\boldsymbol{Z}_{ki}-\overline{\boldsymbol{Z}}_{kb})}$ and rearrange the observations in ascending order according to $MMD_i(\overline{\boldsymbol{Z}}_{kb},\boldsymbol{S}_b)$. Let r be the number of observations in the current basic subset. Divide the observations into two subsets: a basic subset which includes the first $(r+1)$ observations and a non-basic subset containing the remaining $(n-k-1)$ observations

Step 3: Repeat Step 2 until the basic subset contains h observations.

Step 4: Let r continue to be the number of observations in the current basic subset then

(a). Compute $MMD_i(\overline{\boldsymbol{Z}}_{kb},\boldsymbol{S}_b)$ and let $d_{(r+1)}$ be the $(r+1)^{th}$ ordered statistic of $MMD_i(\overline{\boldsymbol{Z}}_{kb},\boldsymbol{S}_b)$

(b). If $d_{(r+1)}^2 \geq c_f\chi_{k,\alpha}^2$, then declare all observations with $MMD_i^2 \geq c_f\chi_{k,\alpha}^2$ as outliers and stop, otherwise go to step 5.

Step 5: Divide the observations into two subsets: a basic subset which includes observations with the smallest $(r+1)$ values of $MMD_i(\overline{\boldsymbol{Z}}_{kb},\boldsymbol{S}_b)$ and a non-basic subset containing the remaining $(n-k-1)$ observations. If $n = r+1$ stop, declare no outliers in the data; otherwise go to step 4.

C. *Simulation Design*

We carried out several simulations to compare the family-wise error rate of different methods for MCC. Let $\lambda_i = \frac{\sigma_i}{\sigma_0}, i = 1,\dots,k$, be the ratios of standard deviation of the i^{th} treatment to that of the control group. Three different settings are considered, and λ_i are assumed to be the same for $i = 1,\dots,k$ in settings one and two.

Setting one: Three random samples are generated from a normal distribution with mean $\mu = 1$ and different standard deviations as below.

$$n_0 = 20,\quad n_1 = 15,\quad n_2 = 10,$$
$$\sigma_0 = 1,\ \lambda_i = 3,5,7,10 \text{ for } i = 1,2.$$

Setting two: Four random samples are generated from a normal distribution with mean $\mu = 1$ and different standard deviations

$$n_0 = 20,\ n_1 = 15,\ n_2 = 10,\ n_3 = 5,$$
$$\sigma_0 = 1,\ \lambda_i = 3,5,7,10 \text{ for } i = 1,2,3$$

Setting three: Four random samples are generated from a normal distribution with mean $\mu = 1$ and different standard deviations

$$n_0 = 20,\quad n_1 = 15,\quad n_2 = 10,\quad n_3 = 5,$$
$$\sigma_0 = 1,\ \sigma_1 = 2,\ \sigma_2 = 4,\quad 2 \leq \sigma_3 \leq 10.$$

The family-wise error rate is defined as the probability of rejecting at least one null hypothesis given all the null hypotheses are true according to Hochberg and Tamhane [6]. The estimated error rates are given in Tables 1–3, respectively, for settings one, two and three at the

significance level $\alpha = 0.05$ based on 5000 simulations, and the standard errors of the estimate are given in the parentheses.

IV. Results And Discussion

For all three settings, DM is very liberal and has inflated error rates (range from 0.07 to 0.2). TM is very conservative with most error rates around 0.04. The proposed methods, MMD, has error rates around 0.05 for all settings, and performs the best among all comparison methods. The performance of LM is similar in terms of the error rate for all settings (around 0.052).

Table 1. Estimated error rate for $k = 2$, $\alpha = 0.05$. (Setting one)

λ_i	3.0	5.0	7.0	10.0
MMD	**0.050(0.0030)**	**0.051 (0.0030)**	**0.050(0.0032)**	**0.051 (0.0030)**
TM	0.047 (0.0031)	0.046 (0.0031)	0.048 (0.0031)	0.048 (0.0031)
LW	0.056 (0.0031)	0.058 (0.0031)	0.058 (0.0030)	0.061 (0.0031)
DM	0.070 (0.0036)	0.068 (0.0036)	0.072 (0.0037)	0.079 (0.0038)

Table 2. Estimated error rate for $k = 3$, $\alpha = 0.05$. (Setting two)

λ_i	3.0	5.0	7.0	10.0
MMD	**0.051 (0.0029)**	**0.050 (0.0029)**	**0.051 (0.0029)**	**0.049 (0.0028)**
TM	0.046 (0.0031)	0.046 (0.0031)	0.045 (0.0031)	0.048 (0.0030)
LW	0.055 (0.0032)	0.054 (0.0031)	0.056 (0.0031)	0.057 (0.0032)
DM	0.084 (0.0039)	0.090 (0.0040)	0.089 (0.0040)	0.089 (0.0040)

Table 2. Estimated error rate for $k = 3$, $\alpha = 0.05$. (Setting three)

λ_i/n_3	0.4	0.8	1.2	1.6	2.0
MMD	**0.050 (0.0030)**	**0.050 (0.0038)**	**0.051 (0.0032)**	**0.050 (0.0028)**	**0.051 (0.0029)**
TM	0.047 (0.0030)	0.047 (0.0031)	0.049 (0.0031)	0.053 (0.0032)	0.047 (0.0030)
LW	0.054 (0.0032)	0.053 (0.0032)	0.049 (0.0031)	0.049 (0.0031)	0.051 (0.0031)
DM	0.092 (0.0041)	0.137 (0.0048)	0.184 (0.0055)	0.205 (0.0057)	0.218 .0058)

V. Conclusion

In this paper, we propose the MMD algorithm procedures for MCC when the equal variance assumption fails. Compared with other approximations, the proposed method controls the family-wise error rate for different sample sizes and variances.

In summary, how to control the family-wise error rate is a central issue in the area of the multiple comparisons. The plausibility of the equal variance condition should always be considered and verified. When the assumption of the equal variances is not satisfied, the methods with more flexible restrictions, such as the method proposed in this article, may be considered as a more reasonable candidate for the MCC.

Acknowledgment

The authors are grateful to anonymous reviewers for their valuable comments on the original draft of this manuscript

References

[1] Carrol, R.J. and D. Ruppert. (1998). *Transformation and weighting in Regression*. John Wiley and Sons New York

[2] Dunnett, C.W. (1955). A Multiple Comparison Procedure for Comparing Several Treatments with a Control, *Journal of the American Statistical Association*, 50, 1096-1121.

[3] Games, P.A. and Howell, J. F. (1976). Pairwise Multiple Comparison Procedures with Unequal and/or Variances: A Monte Carlo study. *Jour. Educ. Statistics*, 1, 113-125.

[4] Hasler, M. and Hothorn, L. A. (2008). Multiple Contrast Tests in the Presence of Heteroscedasticity. *Biometrical Journal*, 50, 793-800.

[5] Hochberg, Y. (1976) . A Modification of the T -Method of Multiple Comparisons for a One Way Layout with Unequal Variances. *Journal of the American Statistical Association*, 71, 200-203.

[6] Hochberg, Y. and Tamhane, A. C. (1987). A*Multiple Comparison Procedure*, New York: Wiley.

[7] Li, H. (2011). All Pairwise Comparison for population with unequal variances. *Far East Journal of Theoretical Statistics*, 34, 1-16.

[8] Li, H. and Wei. N. (2012) . Multiple comparisons with a control under heteroscedasticity. *Journal of Applied Statistics*, 39, 2275 2283.

[9] Scheff, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, 40, 87-110.

[10] Tamhane, A.C. (1977). Multiple comparisons in model I one-way ANOVA with unequal variances. *Communications in Statistics – Theory and Methods*, 6, 15-32.

[11] 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.

[12] Tukey, J.W. (1953). *The Problem of Multiple Comparisons.*. Dittoed Manucript of 396 pages, Department of Statistics, Princeton University