

Application of Harmonic Mean of Variances for Testing Ordered Alternative Hypothesis under Variance Heterogeneity

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

Department of Statistics, University of Ilorin, Ilorin, Nigeria

Abstract In this paper, we examined a test statistic for testing ordered alternative hypothesis under unequal group variances. The proposed modified t – test statistic obtained by replacing the pooled sample variance with the corresponding harmonic mean of variances in order to avoid the Behren – Fisher’s problem. The distribution of the sample of harmonic mean of variances which is known to be generalized Beta is approximated by the chi – square using a condition that is easily satisfied. The result shows that the modified t – test statistic is found to be appropriate for the data set obtained from Kwara Agricultural Development Project (KWADP) where the crop yields of the treatment means were known to be in a particular order because of ecological conditions.

Keywords Ordered alternative, ANOVA, Variances heterogeneity, Harmonic mean of variances

1. Introduction

This work primarily concerns itself with the application of test hypothesis with directional alternatives. This has application in many fields. For examples, there may be a number of cures for a particular ailment. Orthodox chemotherapy, non-orthodox herbal and “body cure” (allowing the body to perform the cure) may be three methods of cure. The best, in terms of duration of treatment before cure could be chemotherapy, followed by the herbal. Thus, if time to realize the cure is the concerned variable, then the means would be ordered from the shortest (obtained from orthodox) to the one obtained by herbs, while the time for body cure is the last. Several authors have provided adequate literature in this area of Analysis of Variance such as Ott (1984), Montgomery (1981) amongst. Many authors have also developed method of testing of homogeneity of means. Such authors include Jonckheere (1954), Bartholomew (1959), Yahya and Jolayemi (2003), Dunnett (1964), Barlow et. al (1971), Dunnett (1980), Dunnett and Tamhane (1997), Keselman and Wilcox (1999), Gupta et al (2006), Abidoye et. al (2013a), (2013b), (2013c) and several others.

2. Development of the Test Procedure

In this section we are interested in developing a suitable test procedure to test the hypothesis against ordered means which is given by

$$H_0 : \mu_i = \mu \text{ vs } H_1 : \mu_i - \mu_{i+1} > 0 \quad (2.1)$$

[The situation in which the hypothesis set is given as

$$H_0 : \mu_i = \mu \text{ vs } H_1 : \mu_1 < \mu_2 < \dots < \mu_g \text{ i.e } H_1 : \mu_{i+1} - \mu_i > 0 \text{ can also be handled similarly}]$$

When $\min(\mu_i - \mu_{i+1}) > 0$ then the alternative hypothesis is sustained. In this formulation then,

$$H_0 : \mu_i = \mu \quad \forall_i \quad \text{Vs} \quad H_1 : \min(\mu_i - \mu_{i+1}) > 0 \quad (2.2)$$

Now assume that $Y_{ij} \sim N(\mu_i, \sigma_i^2)$ and identically and independently distributed where $\sigma_i^2 \neq \sigma^2$ for at least one $i, i = 1, 2, \dots, g, j = 1, 2, \dots, n_i$

The unbiased estimate of

$$\min(\mu_i - \mu_{i+1}) = \min(\bar{Y}_i - \bar{Y}_{i+1}) \quad (2.3)$$

and that

$$V(Y) = \text{Var}[\min(\bar{Y}_i - \bar{Y}_{i+1})] = \frac{\sigma_i^2}{n_i} + \frac{\sigma_{i+1}^2}{n_{i+1}} \quad (2.4)$$

* Corresponding author:

abidoye@unilorin.edu.ng (Abidoye A. O.)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2014 Scientific & Academic Publishing. All Rights Reserved

so that

$$\min(\bar{Y}_i - \bar{Y}_{i+1}) \sim \lambda N(\mu_i - \mu_{i+1}, \frac{\sigma_i^2}{n_i} + \frac{\sigma_{i+1}^2}{n_{i+1}}) \quad (2.5)$$

where

$$\lambda = g(1 - \Phi(\bar{Y}_i - \bar{Y}_{i+1}))^{g-1}, 0 < \lambda < 1 \quad (2.6)$$

obtained from order statistics, from equation (2.5), S_p^2 will be appropriate to be used as the estimate if the group of variances are equal, i.e $\sigma_i^2 = \sigma_{i+1}^2 \forall_i$ where

$$S_p^2 = \frac{\sum (n_i - 1)s_i^2}{\sum (n_i - 1)}$$

Consider the observations X_1, X_2, \dots, X_n in which there exists at least one outlier. Within this situation it is a well known fact that sample mean \bar{X} given as

$$\bar{X} = \frac{1}{n} \sum X_i$$

is not a good location summary. Instead, the Harmonic mean \bar{X}_H given by

$$\bar{X}_H = \left[\frac{1}{n} \sum \frac{1}{X_i} \right]^{-1}$$

is a better choice. Similarly let $\sigma_1^2, \sigma_2^2, \dots, \sigma_g^2$ be the variances of the g groups σ_H^2 which is the Harmonic mean of the variances is a better choice whenever at least one $\sigma_i^2 \neq \sigma^2$.

Note that

$$\begin{aligned} \hat{\sigma}_H^2 &= S_H^2 \\ &= \left[\frac{1}{g} \sum \frac{1}{S_i^2} \right]^{-1} \end{aligned}$$

see abidoye et. al (2012, 2013a and 2013b)

But in a situation where the group variances are not equal, that is, $\sigma_i^2 \neq \sigma^2$ for at least one i . S_p^2 (pooled variance which is the weighted mean of variances) can not be used. This is the Behren's – Fisher problem.

Thus

$$\begin{aligned} Var(\bar{Y}_i - \bar{Y}_{i+1}) &= \frac{\sigma_i^2}{n_i} + \frac{\sigma_{i+1}^2}{n_{i+1}} \\ &= \sigma_H^2 \left(\frac{1}{n_i} + \frac{1}{n_{i+1}} \right) \end{aligned}$$

The test statistic for the hypotheses set in (2.2) is therefore

$$t^* = \frac{\lambda Y}{Z} \quad (2.7)$$

where

$$Y = \min(\bar{y}_i - \bar{y}_{i+1}) \quad (2.8)$$

and

$$Z = \sqrt{S_H^2 \left(\frac{1}{n_i} + \frac{1}{n_{i+1}} \right)} \quad (2.9)$$

where λ is the appropriate normalization from order statistics. The null hypothesis is then rejected if

$$\begin{aligned} P(t_r^* = \lambda \frac{Y}{Z} > t_0^*) &= P(t_r^* = \lambda t_r > t^*) \\ &= P(t_r > \frac{t^*}{\lambda}) < \alpha \end{aligned} \quad (2.10)$$

where t_r is the regular t – distribution with r d.f is given by

$$t_r = \frac{(\bar{Y}_i - \bar{Y}_{i+1})}{S_H \left(\frac{1}{n_i} + \frac{1}{n_{i+1}} \right)^{\frac{1}{2}}} \quad (2.11)$$

because S_H^2 has approximately the χ^2 - distribution with the degree of freedom r to be determined from $r = 22.096 + 0.266(n-g) - 0.000029(n-g)^2$. See Abidoye (2012) and Abidoye et. al 2013a. If the alternative hypothesis (H_1) in equation (2.10) is reversed.

Then

$$t_r = \frac{(\bar{Y}_{i+1} - \bar{Y}_i)}{S_H \left(\frac{1}{n_i} + \frac{1}{n_{i+1}} \right)^{\frac{1}{2}}} \quad (2.12)$$

3. Application of the Test Procedure

Practical situations where this test procedure is applicable are presented in this section. The data on Maize was found in Kwara Agricultural Development Project (KWADP), Ilorin, Kwara State for the period of 1998 to 2007. By KWADP standard, Kwara State has been categorized into four major unique ecological zones based on the climatic and ecological differences.

3.1. Example on Maize Yields in Kwara State

Consider the data on yields of maize (in Tonnes per hectare) given by KWADP who believed that crop yield in

the area under consideration, zone D has the highest yield followed by zone A, then zone C and zone B with the least yield.

If μ_i represents the mean population yield of maize in zone i where $1 \leq i \leq g = 4$.

Let Y_{ij} represents the crop (maize) yield (in Tonnes per hectare) in zone i for year j , $i = 1, 2, 3, 4$; $j = 1, 2, \dots, 10$.

Then, assume

$$Y_{ij} \sim N(\mu_i, \sigma_i^2) \quad (2.13)$$

First it is necessary to examine the values of the group variances. This was done in Table 3.1, confirming unexqual variances.

Table 3.1. Spss Result Output on Test of Homogeneity of Variances For Maize

	Levene STATISTIC	df ₁	df ₂	P-value
Response	7.017	3	36	.001

The hypothesis to be tested is

$$H_0: \mu_D = \mu_A = \mu_C = \mu_B \text{ Vs } H_1: \mu_D > \mu_A > \mu_C > \mu_B ;$$

$$\sigma_i^2 \neq \sigma^2 \text{ for at least one } i$$

The Proposed Test Statistic is

$$t^* = \frac{\min(\bar{X}_{i+1} - \bar{X}_i)}{S_H \sqrt{\left(\frac{1}{n_{i+1}} + \frac{1}{n_i}\right)}} \sim \lambda t_r,$$

$$\text{where } S_H^2 = \left(\frac{1}{g} \sum \frac{1}{s_i^2}\right)^{-1}$$

and

$$n = \sum_{i=1}^g n_i$$

The null hypothesis, H_0 is rejected at specified α level of significance if

$$\text{P-value} = P\left(t_r > \frac{t^*}{\lambda}\right) < \alpha, \text{ where } t \sim t_r$$

COMPUTATION ON MAIZE

From the data on Maize yield the following summary statistics were obtained:

$$\text{Zone A: } \bar{Y}_A = 1.014, S_A^2 = 0.056, n_A = 10$$

$$\text{Zone B: } \bar{Y}_B = 0.618, S_B^2 = 0.0126, n_B = 10$$

$$\text{Zone C: } \bar{Y}_C = 0.8, S_C^2 = 0.0111, n_C = 10$$

$$\text{Zone D: } \bar{Y}_D = 2.776, S_D^2 = 0.027, n_D = 10$$

Therefore,

	D	A	C	B
Means	2.776	1.014	0.80	0.618
Variances	0.027	0.056	0.0111	0.0126

Then consider the differences given below:

$$\bar{Y}_D - \bar{Y}_A = 2.776 - 1.014 = 1.762$$

$$\bar{Y}_A - \bar{Y}_C = 1.014 - 0.80 = 0.214$$

$$\bar{Y}_C - \bar{Y}_B = 0.80 - 0.618 = 0.182$$

$$\min(\bar{Y}_{i+1} - \bar{Y}_i) = (\bar{Y}_C - \bar{Y}_B) = 0.182$$

$$\text{with } n_{i+1} = n_i = 10, g = 4$$

$$n = \sum_{i=1}^4 n_i = 40$$

$$S_H^2 = \left(\frac{1}{g} \sum \frac{1}{s_i^2}\right)^{-1}$$

$$S_H^2 = \left(\frac{1}{4}(0.056^{-1} + 0.0126^{-1} + 0.0111^{-1} + 0.027^{-1})\right)^{-1}$$

$$= 0.01783$$

$$S_H = \sqrt{S_H^2} = 0.1335$$

$$S_H = 0.1335$$

$$t^* = \frac{\min(\bar{Y}_{i+1} - \bar{Y}_i)}{S_H \sqrt{\left(\frac{1}{n_{i+1}} + \frac{1}{n_i}\right)}}$$

$$= \frac{0.182}{0.1335 \sqrt{\frac{1}{10} + \frac{1}{10}}}$$

$$= \frac{0.182}{0.05970}$$

$$= 3.0484$$

$$\text{where } r = 22.096 + 0.266(n-g) - 0.000029(n-g)^2$$

$$= 31.634$$

$$\lambda = g(1 - \Phi(\bar{Y}_i - \bar{Y}))^{g-1} = 0.3149, 0 < \lambda < 1$$

from equation (2.1)

$$\begin{aligned}
\text{Now p-value} &= P(t_r > t) = P\left(t_r > \frac{t_{cal}}{\lambda}\right) \\
&= P\left(t_r > \frac{3.0484}{0.3149}\right) \\
&= P(t_r > 9.6805) \\
&= 1 \times 10e^{-8} < 0.05
\end{aligned}$$

Which led to the rejection of H_0 and conclude that the mean yields of maize in all the four zones were indeed ordered from D, A, C, B.

4. Conclusions

In this study, we examined a test statistic for testing ordered alternative hypothesis under unequal variances. We proposed a modified t – test statistic, the harmonic mean of the variances to replace the sample pooled variance so as to avoid the Behrens – Fisher’s problem. Because the sample harmonic mean of variances has the chi – square distribution, then the modified t – statistic is found to be appropriate for the data set used from Kwara Agricultural development project (KWADAP).

REFERENCES

- [1] 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.
- [2] Abidoye, A.O, Jolayemi, E.T, Sanni, O.O.M and Oyejola, B.A (2013a): On Hypothesis Testing Under Unequal Group Variances: The Use of the Harmonic variance. The International Journal of Institute for science, Technology and Education. Vol. 3, No 7, Page 11 – 15.
- [3] Abidoye, A.O, Jolayemi, E.T, Sanni, O.O.M and Oyejola, B.A (2013b): Development of a Test statistic for Testing Equality of Two Means Under Unequal Population Variances. The International Journal of Institute for science, Technology and Education. Vol. 3, No 10, page 10 -14.
- [4] Abidoye, A.O, Jolayemi, E.T, Sanni, O.O.M and Oyejola, B.A (2013c): Development of a Test Statistic for Testing Equality of Means Under Unequal Population Variances. The International Journal of Mathematics and Statistics Invention. Vol. 1, page 38 -43.
- [5] Barlow, R.E, Bartholomew, D.J, Bremner, J.M and Brunk, H.D (1971): “Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression.” Wiley, New York.
- [6] Bartholomew, D.J (1959): “A test of homogeneity for ordered alternative.” Biometrika, 46, 3648.
- [7] 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.
- [8] Dunnett, C. W (1964): New tables for multiple comparison with a control. Biometric, 20, 482-491.
- [9] Dunnett, C.W (1980): “Pairwise multiple comparison in the homogeneous variance, unequal sample size case” Journal of the American Statistical Association. 75, 789 – 795.
- [10] 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, pp 148-178.
- [11] Jonckheere, A.R (1954): “A distribution-free k-sample test against ordered alternative.” Biometrical, 41, 133-145.
- [12] Keselman, H.J and Wilcox R. (1999): The Improved Brown and Forsythe test for mean equality: Some things can’t be fixed. Communications in Statistics – Simulation and Computations, Vol. 28, Pages 687- 698.
- [13] Montgomery, D.C (1981): “Design and Analysis of Experiment.” Second Edition. John Wiley and sons Inc. New York.
- [14] Ott, L (1984): “An Introduction To Statistical Methods and Data Analysis”. Second edition. P.W.S Publisher. Boston.
- [15] Yahya, W.B and Jolayemi, E.T (2003): Testing Ordered Means against a Control. JNSA, 16, 40-51.