APPLICATION OF A T - STATISTIC FOR TESTING EQUALITY OF MEANS WITH DIRECTIONAL ALTERNATIVE WHEN POPULATION VARIANCES ARE UNEQUAL

Supported by tetfund TERTIARY EDUCATION TRUST FUND

A. O. Abidoye and O. O. M. Sanni
Department of Statistics, University of Ilorin, Kwara State, Nigeria
*Corresponding author: solatayo2003@gmail.com, abidoye@unilorin.edu.ng, solatayo2003@yahoo.com

Received: December 12, 2017 **Accepted:** June 19, 2018

Abstract: In this study, we proposed a test statistic for testing equality of means when variances are not equal. When variances of different groups are significantly different from one another it is not proper to use the pooled sample variance (S_P^2) as a single value for the variances. In this work we are interested in testing directional hypothesis, since the variances are unequal then we make use of harmonic mean variance (S_H^2). The means are ranked such that the problem reduces to a two sample situations. Data set from Kwara State Ministry of Agriculture on the yield of maize (kilograms) in four different locations was used to demonstrate directional hypothesis testing.

Keywords: Harmonic mean of variances, chi-square distribution, directional alternative hypothesis

Introduction

A t-test is often used to compare the difference between two means or more that are based on samples. The samples come from populations. In that context, the test's statistical power is the probability that you will conclude that more than two population means are different when they are different. It can also represent the probability of correctly deciding that one population mean is not just different from but larger than the other. Every hypothesis test requires that analyst to state a null hypothesis and an alternative hypothesis. The hypothesis is stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false; and vice versa.

This work primarily concern itself with the application of testing hypothesis with directional alternatives, this has application in many fields such as Agriculture, Medicine, Orthodox chemotherapy, non – orthodox herbal and body cure. The hypothesis of homogeneity of means $H_0 : \mu_1 = \mu_2 = ... = \mu_g = \mu$ to be tested against the ordered alternative, H₁: $\mu_1 \neq \mu_2 \neq ... \neq \mu_g$, H₁: $\mu_1 < \mu_2 < ... < \mu_g$ or H₁: $\mu_1 > \mu_2 > ... > \mu_g$.

The interest of this work is to develop a suitable test procedure to address heterogeneity of variances if present and we propose a test statistic for testing equality of means against directional alternative in the presence of heterogeneity of variances. See Abidoye 2012. Also Abidoye *et al.* (2015), Abidoye *et al.* (2016a, 2016b), mention the use of Agricultural research where the interest is to investigate the effectiveness of certain brands of fertilizer meant for a particular crop there might be a pre- conceived belief that certain brand(s) are more effective than others; indeed following an ordered form of performance.

Adegboye and Gupta (1986) discussed testing equality of means under common but unknown variance (σ^2) using ordered alternative with strict inequality. Bartholomew (1959) duel on testing k normal variates having some mean against the alternative hypothesis $H_1 : \mu_1 \neq \mu_2 \neq ... \neq \mu_k$. Gupta *et al*. (2006) in consideration of multivariate mixed models, suggested that the distributional assumptions of the errors are not required but only assumed that the random sample from large population of levels. Cochran (1964) investigated the test of equality of means in Behrens – Fisher problem and compared his test with the test developed by Benerjee (1960) and McCullough *et al.* (1960). Levene (1960) proposed a test criterion for testing equality of variances for specified significance level. In this paper we are proposing a t– statistic for testing equality of means when the variances are unequal.

Methodology

The unbiased estimate of

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

and the unbiased estimate of

$$\max(\mu_i - \mu_{i+1}) = \max(\bar{Y}_i - \bar{Y}_{i+1}) = Y^* \qquad (2)$$

Where $\min(\mu_i - \mu_{i+1})$ is the ordered means for minimum and $\max(\mu_i - \mu_{i+1})$ is the ordered means for maximum.

Therefore,

$$Var(Y) = Var[\min(\bar{Y}_i - \bar{Y}_{i+1})] \qquad (3)$$

and

$$Var(Y^*) = Var[\max(\bar{Y}_i - \bar{Y}_{i+1})] \qquad (4)$$

$$Var(Y) = \frac{\sigma_i^2}{n_i} + \frac{\sigma_{i+1}^2}{n_{i+1}} \qquad (5)$$

$$Y = \min(\bar{Y}_i - \bar{Y}_{i+1}) \sim \lambda N(\mu_i - \mu_{i+1}, \frac{\sigma_i^2}{n_i} + \frac{\sigma_i^2}{n_{i+1}}) \qquad (6)$$

also

$$Y^* = \max(\bar{Y}_i - \bar{Y}_{i+1}) \sim \lambda\ N(\mu_i - \mu_{i+1}, \frac{\sigma_i^2}{n_i} + \frac{\sigma_{i+1}^2}{n_{i+1}}) \qquad (7)$$

where λ is a scaling factor from normal population

Consequently, the test statistic for the hypotheses set in equation (1) is

$$t = \frac{\lambda Y_i}{Z} \qquad (8)$$

where

$$Y_i = \min\left(\bar{Y}_i - \bar{Y}_{i+1}\right) \qquad (9)$$

and

$$Z = \sqrt{S_H^2\left(\frac{1}{n_i} + \frac{1}{n}\right)} \qquad (10)$$

where $S_H^2 \sim gb(\alpha, \beta, \lambda)$ which has approximately the χ^2 - distribution with the degree of freedom to be determined and $Y_i = \min(\bar{Y} - \bar{Y}_{i+1})$ follow normal distribution.

Now p- value = $P(t_r > t) = P(t_r^* > \frac{t}{\lambda})$ (11)

where λ has define earlier above, it can be λ_1 or λ_2 and t_r^* is regular t – distribution and r is the appropriate degrees of freedom for the t – test.

Data Analysis

The data used in this study are secondary data, collected primarily by Kwara State Ministry of Agriculture, Ilorin, Kwara State, Nigeria.

Table 1: The yield of maize (kilograms) in four different locations in Ministry of Agriculture, Ilorin, Kwara State

Years	1	2	3	4	5	6	7	8	9	10
Zone A	30	72	63	44	55	36	65	49	69	56
Zone B	34	29	22	31	13	33	45	20	31	24
Zone C	25	28	31	29	27	34	13	25	18	23
Zone D	31	29	30	25	19	26	18	24	19	27

By the application of Levene test of equality of variances, the test is given in Table 2.

Table 2: Levene test for variance equality

	Levene Statistic	df_1	df_2	P-value
Response	12.367	3	36	0.000

From above test result, the variances are significantly different from location to location. See Abidoye *et al.* (2016b).
From the data in Table 1 the following summary statistics were obtained:

$$ZoneA: \bar{X}_A = 53.9, S_A^2 = 197.88, n_A = 10$$
$$ZoneB: \bar{X}_B = 28.2, S_B^2 = 78.84, n_B = 10$$
$$ZoneC: \bar{X}_C = 25.3, S_C^2 = 38.01, n_C = 10$$
$$ZoneD: \bar{X}_D = 24.8, S_D^2 = 22.62, n_D = 10$$

$\bar{X}_A$	$\bar{X}_B$	$\bar{X}_C$	$\bar{X}_D$	$\bar{\bar{X}}$
53.9	28.2	25.3	24.8	33.1

Therefore, we consider the minimum and maximum differences of means respectively as given below:

$$Y_1 = 53.9 - 33.1 = 20.8$$
$$Y_2 = 28.2 - 33.1 = -4.9$$
$$Y_3 = 25.3 - 33.1 = -7.8$$
$$Y_4 = 24.8 - 33.1 = -8.3$$

$$S_H^2 = \left(\frac{1}{g}\sum \frac{1}{s_i^2}\right)^{-1}$$
$$= 45.32$$

Then, the minimum difference of means is

$$Y^* = \min(\bar{X}_i - \bar{\bar{X}}) = (\bar{X}_C - \bar{\bar{X}}) = -8.3$$

In the above data set, $n_i = 10$, g = 4 , $n = \sum_{i=1}^{4} n_i = 40$,

$$S_H^2 = \left(\frac{1}{4}\sum_{i=1}^{4}\frac{1}{s_i^2}\right)^{-1}, \quad S_H^2 = 45.32$$

The main hypothesis is;

$H_0: \mu_A = \mu_B = \mu_C = \mu_D = \mu$ against H1: $\mu_i \neq \mu$, for at least one i, i.e $i = A, B, ..., D$

The hypothesis to be tested is

$H_0: \mu_i - \mu = 0$ vs H1: $\mu_i - \mu < 0$

$$t = \frac{\min(\bar{X}_i - \bar{\bar{X}})}{S_H\sqrt{\left(\frac{1}{n_i} + \frac{1}{n}\right)}} \sim \lambda_2 t_r$$

$$= \frac{-8.3}{6.73\sqrt{\left(\frac{1}{10} + \frac{1}{40}\right)}} = \frac{-8.3}{2.3794}$$

= -3.388

$$\lambda_2 = g(\Phi(\bar{Y}_i - \bar{\bar{Y}}))^{g-1} = 0^+$$

Now p- value = $P(t_r > t) = P\left(t_r < \frac{t_{cal}}{\lambda_2}\right)$

$$= P\left(t_r < \frac{-3.388}{0^+}\right)$$
$$= P(t_r < -\infty)$$
$$\approx 0$$

< 0.016

In this regard, we reject H_0 and conclude that the mean of the yield of maize (kilograms) in four different zones are significantly different at 5% level of significance.
Next we consider the maximum difference of means

$$Y^{**} = \max(\bar{X}_i - \bar{\bar{X}}) = (\bar{X}_A - \bar{\bar{X}}) = 20.8$$

In the above data set, $n_i = 10$, g = 4 , $n = \sum_{i=1}^{4} n_i = 40$,

$$S_H^2 = \left(\frac{1}{4}\sum_{i=1}^{4}\frac{1}{s_i^2}\right)^{-1}, \quad S_H^2 = 45.32$$

The hypothesis to be tested is

$H_0: \mu_i - \mu = 0$ vs H1: $\mu_i - \mu > 0$

$$t = \frac{\max(\bar{X}_i - \bar{\bar{X}})}{S_H\sqrt{\left(\frac{1}{n_i} + \frac{1}{n}\right)}} \sim \lambda_1 t_r$$

$$= \frac{20.8}{6.73\sqrt{\left(\frac{1}{10} + \frac{1}{40}\right)}} = \frac{20.8}{2.3794}$$

= 8.74

$\lambda_1 = g(1-\Phi(\bar{Y}_i - \bar{\bar{Y}}))^{g-1} = 0 \quad ,0 < \lambda_1 < 1$ from equation (1)

Now p- value = $P(t_r > t) = P\left(t_r < \frac{t_{cal}}{\lambda_1}\right)$

$$= \left(t_r < \frac{8.74}{0^+}\right)$$

$$= P(t_r < +\infty)$$

$$\approx 0$$

$$< 0.016$$

Which led to the rejection of H_0 and we therefore conclude that the mean of the yield of maize (in kg) in the four different zones are significantly different at 5% level of significance.

Conclusion

In this application we have demonstrated testing equality of means with directional alternative when the population variances are not equal. Because the sample harmonic mean of variances has approximately chi – square distribution, the t – statistic is found to be appropriate and it help in overcoming the Beheren- Fisher's problem.

References

Abidoye AO 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.

Abidoye AO, Jolayemi ET, Sanni OOM & Oyejola BA 2016a. On application of modified F – Statistic: An example of sales distribution of pharmaceutical drug. *Journal of Science World*, 11(2): 23 – 26. Available at www.scienceworldjournal.com

Abidoye AO, Jolayemi ET, Sanni OOM & Oyejola BA 2016b. Development of testing ordered mean against a control under heterogeneous variance. *J. Nig. Assoc. Math. Phy.*, 33: 125 – 128.

Abidoye AO, Jolayemi ET, Sanni OOM & Oyejola BA 2015. Development of hypothesis testing on type one error and power function. *Ilorin J. Sci., Faculty of Physical Sci.*, 2(1) 68 – 79.

Adegboye SO & Gupta AK 1986. On testing against restricted alternative about the mean of Gaussian models with common unknown variance. *Sankhya, the Indian J. Stat., Series B*, 48: 333.

Bartholomew DJ 1959. A test of homogeneity for ordered alternative. *Biometrica*, 46: 36 – 48.

Benerjee SK 1960. Approximate confidence interval for linear functions of means of K populations when the population variance are not equal. *Sankhya*, 22: 357 – 358.

Cochran WG 1964. Approximate significance levels of the Behrens – Fisher test. *Biometrics*, 20: 191 – 195.

Gupta AK, Solomon WH & Yasunori F 2006. Asymptotics for testing hypothesis in some multivariate variance components model under non – normality. *J. Multivariate Anal. Archive*, 97: 148 – 178.

Levene H 1960. In contribution to probability and statistics: Essays in honor of harold hotelling, I Olkin *et al.* editions, Standford University Press, pp. 278 – 292.

McCullough Roger S, Gurland J & Rosen-berg L 1960. Small sample behavior of certain tests of the hypothesis of equal means under variance heterogeneity. *Biometrika*, 47: 345 – 353.