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

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^2_p) 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^2_h). 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 : μ_1 = μ_2 = ... = μ_g = μ to be tested against the ordered alternative, H_1: μ_1 ≠ μ_2 ≠ ... ≠ μ_g, H_1: μ_1 < μ_2 < ... < μ_g or H_1: μ_1 > μ_2 > ... > μ_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. Adegbeyo 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 : μ_1 ≠ μ_2 ≠ ... ≠ μ_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
\[ \min(\mu_i - \mu_{i+1}) = \min(\bar{Y}_i - \bar{Y}_{i+1}) = Y \] (1)
and the unbiased estimate of
\[ \max(\mu_i - \mu_{i+1}) = \max(\bar{Y}_i - \bar{Y}_{i+1}) = Y^* \] (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,
\[ Var(Y) = Var[\min(\bar{Y}_i - \bar{Y}_{i+1})] \] (3)
and
\[ Var(Y^*) = Var[\max(\bar{Y}_i - \bar{Y}_{i+1})] \] (4)

\[ Var(Y) = \frac{\sigma_i^2}{n_i} + \frac{\sigma_{i+1}^2}{n_{i+1}} \] (5)
\[ Y = \min(\bar{Y}_i - \bar{Y}_{i+1}) \sim 2N(\mu_i - \mu_{i+1}, \frac{\sigma_i^2}{n_i} + \frac{\sigma_{i+1}^2}{n_{i+1}}) \] (6)
also
\[ Y^* = \max(\bar{Y}_i - \bar{Y}_{i+1}) \sim \lambda N(\mu_i - \mu_{i+1}, \frac{\sigma_i^2}{n_i} + \frac{\sigma_{i+1}^2}{n_{i+1}}) \] (7)

where \( \lambda \) is a scaling factor from normal population

Consequently, the test statistic for the hypotheses set in equation (1) is
\[ t = \frac{2Y}{Z} \] (8)
where
\[ Y_i = \min(\bar{Y}_i - \bar{Y}_{i+1}) \] (9)
and
\[ Z = \sqrt{\frac{S^2_h}{n} \left( \frac{1}{n_i} + \frac{1}{n_{i+1}} \right)} \] (10)
where \[ S^2_h \sim g\beta(\alpha, \beta, \lambda) \] which has approximately the \( \chi^2 \) distribution with the degree of freedom to be determined and \[ Y_i = \min(\bar{Y}_i - \bar{Y}_{i+1}) \] follow normal distribution.
Now p-value = \( P(t_r > t) = P(t_r^* > \frac{t}{\lambda}) \)  
\( \lambda \) has defined earlier above, it can be \( \lambda_1 \) or \( \lambda_2 \) 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.

<table>
<thead>
<tr>
<th>Years</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone A</td>
<td>30</td>
<td>72</td>
<td>63</td>
<td>44</td>
<td>55</td>
<td>36</td>
<td>65</td>
<td>49</td>
<td>69</td>
<td>56</td>
</tr>
<tr>
<td>Zone B</td>
<td>34</td>
<td>29</td>
<td>22</td>
<td>31</td>
<td>13</td>
<td>33</td>
<td>45</td>
<td>20</td>
<td>31</td>
<td>24</td>
</tr>
<tr>
<td>Zone C</td>
<td>25</td>
<td>28</td>
<td>31</td>
<td>29</td>
<td>27</td>
<td>34</td>
<td>13</td>
<td>25</td>
<td>18</td>
<td>23</td>
</tr>
<tr>
<td>Zone D</td>
<td>31</td>
<td>29</td>
<td>30</td>
<td>25</td>
<td>19</td>
<td>26</td>
<td>26</td>
<td>14</td>
<td>19</td>
<td>27</td>
</tr>
</tbody>
</table>

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

**Table 2:** Levene test for variance equality

<table>
<thead>
<tr>
<th>Levene Statistic</th>
<th>df1</th>
<th>df2</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>12.367</td>
<td>3</td>
<td>36</td>
</tr>
</tbody>
</table>

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:

- **Zone A:** \( X_A = 53.9, S_A^2 = 197.88, n_A = 10 \)
- **Zone B:** \( X_B = 28.2, S_B^2 = 78.84, n_B = 10 \)
- **Zone C:** \( X_C = 25.3, S_C^2 = 38.01, n_C = 10 \)
- **Zone D:** \( X_D = 24.8, S_D^2 = 22.62, n_D = 10 \)

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

\[
Y_1' = 53.9 - 33.1 = 20.8 \\
Y_2' = 28.2 - 33.1 = -4.9 \\
Y_3' = 25.3 - 33.1 = -7.8 \\
Y_4' = 24.8 - 33.1 = -8.3
\]

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

\[
\lambda_2 = g(\Phi(\overline{Y_i} - \overline{Y}))^{\gamma - 1} = 0^+
\]

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

\[
Y'' = \max((X_i - \overline{X})) = (X_A - \overline{X}) = 20.8
\]

In the above data set, \( n_i = 10, g = 4, n = \sum n_i = 40 \).

\[
S_B^2 = \left( \frac{1}{g} \sum \frac{1}{S_i^2} \right)^{-1} = 45.32
\]

\[
S_H^2 = \left( \frac{1}{n} \sum \frac{1}{S_i^2} \right)^{-1} = 45.32
\]

The hypothesis to be tested is

\[
H_0: \mu_1 - \mu = 0 \text{ vs } H_1: \mu_i - \mu > 0
\]

\[
t = \frac{\max((X_i - \overline{X}))}{S_H \sqrt{\left( \frac{1}{n_i} + \frac{1}{n} \right)}} \sim \lambda t_r
\]

\[
= \frac{20.8}{6.73 \sqrt{\left( \frac{1}{10} + \frac{1}{40} \right)}} = \frac{20.8}{2.3794} = 8.74
\]
\[ \lambda_i = g(1 - \Phi(\bar{Y}_i - \bar{Y}))^{x^{-1}} = 0 \quad 0 < \lambda_i < 1 \quad \text{from equation (1)} \]

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

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

\[ \approx P(t_r < +\infty) < 0.016 \]

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, Jolayemi ET, Sanni OOM & Oyejola BA 2016a. On application of modified \( F \) – Statistic: An example of sales distribution of pharmaceutical drug.


