A REVIEW ON TESTING FOR EQUALITY OF MEANS AGAINST ORDERED MEANS

1SAADU,Y. O, 2ADEOYE, A.O AND 3YUSUF, G. A

1DEPARTMENT OF LIBRARY INFORMATION KWARA STATE POLYTECHNIC ILORIN
2DEPARTMENT OF MATHEMATICS/STATISTICS FEDERAL POLYTECHNIC OFFA
3DEPARTMENT OF STATISTICS KWARA STATE POLYTECHNIC ILORIN

ABSTRACT
This research a review on testing for equality of means against ordered means. Statistical hypothesis testing is a key technique of frequent statistical inference, and it is widely used, but also much criticized. The critical region of a hypothesis test is the set of all possible outcomes which, if they occur, will lead us to reject the null hypothesis in favour of alternative hypothesis.

Keywords: testing, equality , means, ordered ,hypothesis testing

1. INTRODUCTION
Hypothesis testing is a method of making statistical decision using experimental data. One use of hypothesis testing is in deciding whether experimental result contains enough information to cast doubt on conventional wisdom. Hypothesis testing are performed by many researcher in various fields of inquiry, usually to discover something about particular process. Literally, hypothesis testing is a method of testing a claim about a parameter in a population, using data measured in a sample[3,4]. In this method, we test some claim by determine a likelihood that a sample statistic could have been selected if the hypothesis regarding the population parameter were true. The null hypothesis is a conservative statement about a population parameter, and it is so termed because it is most in variably states that the given sample comes from a population. The purpose of hypothesis testing is to test the variability of the null hypothesis in the light of experimental data.

2. METHODOLOGY
Depending on the data, the null hypothesis either will or will not be rejected as a variable possibility,[6] worked on hypothesis testing of homogeneity of the form

\[ H_0 : \mu_1 = \mu_2 = \ldots = \mu_g \quad \text{against the ordered alternative}, \quad H_0 : \mu_1 = \mu_2 < \ldots < \mu_g \]

Let

\[ \delta_i = \mu_{i+1} - \mu_i, \quad 1 \leq i \leq g-1, \text{then for each pair of means } \mu_i \text{ and } \mu_{i+1}, \mu_i < \mu_{i+1} \text{if only } \delta_i > 0 \]

Hence the test for hypothesis set above becomes

\[ H_0 : \delta_i = 0 \quad \text{against} \quad H_1 : \min \delta_i > 0 \quad \forall \ i = 1 \rightarrow g - 1 \]

an unbiased estimate of \( \delta_i \) is

\[ \hat{\delta}_i = \bar{Y}_{i+1} - \bar{Y}_i \]

then

\[ \hat{\delta}_i \sim N[\delta_i, \nu(\hat{\delta}_i)] \quad \text{and} \]

\[ \nu(\hat{\delta}_i) = \nu(\bar{Y}_{i+1} - \bar{Y}_i) \]

\[ = \nu(\bar{Y}_{i+1}) + \nu(\bar{Y}_i) \]

\[ = \frac{\sigma^2}{n_{i+1}} + \frac{\sigma^2}{n_i} \quad \text{under the assumption or homogeneity of variances} \]
Thus

\[ v (\hat{\delta}) = \left[ \frac{n_i + n_{i+1}}{n_i n_{i+1}} \right] \sigma^2 \]  

(1)

Now consider the quantity \( m_i \) such that

\[ m_i = \sqrt{\frac{n_i}{n_i + n_{i+1}}} \text{ and let} \]

\[ \hat{\gamma}_i = m_i \hat{\delta} = m_i (\bar{Y}_{i} - \bar{Y}) \]

then

\[ \hat{\gamma}_i \sim N (\gamma_i, \sigma^2) \]

Where \( \gamma_i = m_i \left( \mu - \mu_i \right) \)

The result of the equation (2.3) above is so because:

\[ E(\hat{\gamma}_i) = E(m_i \hat{\delta}) \]

\[ = E(m_i (\bar{Y}_{i} - \bar{Y})) \]

\[ = m_i E(\bar{Y}_{i} - \bar{Y}) \]

Thus, \( E(\hat{\gamma}_i) = m_i \left( \mu - \mu_i \right) = \gamma_i \)  

(2)

Also,

\[ v(\hat{\gamma}_i) = v(m_i \hat{\delta}) = m_i^2 v(\hat{\delta}) \]

\[ = \frac{n_i + n_{i+1} + 1}{n_i + n_{i+1}} \delta^2 \]

Now let

\[ \hat{\gamma}^1 = (\hat{\gamma}_1, \hat{\gamma}_2, ..., \hat{\gamma}_{g-1}) \]

and

\[ \hat{\gamma}^2 = (\gamma_1, \gamma_2, ..., \gamma_{g-1}) \]

Then

\[ \hat{\gamma} = \begin{bmatrix} m_1 (\bar{Y}_1 - \bar{Y}_g), m_2 (\bar{Y}_2 - \bar{Y}_g), ..., m_{g-1} (\bar{Y}_{g-1} - \bar{Y}_g) \end{bmatrix} \]

and

\[ V = \begin{bmatrix} m_1 (\mu_1 - \mu_g), m_2 (\mu_2 - \mu_g), ..., m_{g-1} (\mu_{g-1} - \mu_g) \end{bmatrix} \]

Therefore, the hypothesis set can be written in terms of \( \gamma_i \) as
\[ H_0 : \gamma_i = 0 \ \forall \ i, \text{ against } H_i : \min \gamma_i > 0, 1 \leq i \leq g - 1 \]

Now let

\[ S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_j)^2 \]

and

\[ S^2 = \frac{1}{n - g} \sum_{i=1}^{g} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \]

Then

\[ S^2 = \frac{1}{n - g} \sum_{i=1}^{g} (n_i - 1)S_i^2 \]

Where

\[ S_i^2 \] is the sample variance for the \( i^{th} \) population, \( 1 \leq i \leq g - 1 \) and \( S^2 \) is the pooled sample variance for all the \( g \) population.

Define

\[ T' = (T_1, T_2, \ldots T_{g-1}) \]

\[ T = \left[ m_1(\bar{Y}_2 - \bar{Y}_1), m_2(\bar{Y}_3 - \bar{Y}_2), \ldots, m_{g-1}(\bar{Y}_g - \bar{Y}_{g-1}) \right] \]

and \( U \) as

\[ U^1 = T^1 \frac{S}{S^2} = \left[ \frac{T_1}{S}, \frac{T_2}{S}, \ldots, \frac{T_{g-1}}{S} \right] \]

Where \( T_i \) is defined as

\[ T_i = m_i(\bar{Y}_{i+1} - \bar{Y}_i) \]

Then, we have

\[ U^1 = \left[ m_1 \frac{(\bar{Y}_2 - \bar{Y}_1)}{S}, m_2 \frac{(\bar{Y}_3 - \bar{Y}_2)}{S}, \ldots, m_{g-1} \frac{(\bar{Y}_g - \bar{Y}_{g-1})}{S} \right] \]

\[ = (u_1, u_2, \ldots, u_{g-1}) \]

Where

\[ U_i = m_i \frac{(\bar{Y}_{i+1} - \bar{Y}_i)}{S}, 1 \leq i \leq g - 1 \]

Then, statistic \( U \) has a multivariate \( t \) - distribution with mean \( \gamma \) and \( n - g \) degree of freedom.

\[ MT_{g-1} (\gamma, n - g) \] Independent of statistics [1]

Let \( \nu = \min u_i \]

\[ 1 \leq i \leq g - 1 \]

Where \( \nu \) is non- decreasing in \( T \) and distributed independently of \( S \).

Therefore, in testing the hypothesis,
In term of $\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_g,$ and $S^2$ test $\varphi$ becomes

$$\phi(\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_g, S^2) = \begin{cases} 1, & \text{if } \min_{1 \leq i \leq g-1} mi (\overline{Y}_{i+1} - \overline{Y}_i) / S \geq t_{g,n,a,i=1,\ldots,g-1} \\ 0, & \text{otherwise} \end{cases}$$

That is $H_0$ is rejected if,

$$\hat{T} = \min_{1 \leq i \leq g-1} mi (\overline{Y}_{i+1} - \overline{Y}_i) / S > t_{g,n,a,i=1,\ldots,g-1}$$

Where $mi = n_i n_{i+1} / (n_i + n_{i+1})$

For equal sample sizes $mi = m \forall i$, the test statistic $T$ becomes

$$\hat{T} = \sqrt{\frac{m}{2}} \min_{1 \leq i \leq g-1} (\overline{Y}_{i+1} - \overline{Y}_i) / S$$

It should be noted that test $\varphi$ above reduces to the two-sample one tail t-test when $g = 2$ and it also inherits the three optimality properties of Uniformly Most Powerful Unbiased (UMPU), Uniformly Most Powerful Invariant (UMPI) and Asymptotically Consistency of the two sample one tailed t-test.[1]

However, since the $\varphi$ reduces to the two samples one tailed t-test when $g = 2$, it then shows that statistic $T$ of 5 and 6 follow the multivariate T-distribution. Therefore, the decision rule adopted with unequal sample sizes taken from all the g population is to reject the null hypothesis $H_0$, if

$$\hat{T} = \min \sqrt{\frac{n_i n_{i+1}}{n_i + n_{i+1}}} (\overline{Y}_{i+1} - \overline{Y}_i) / S > t_{g,n,a}$$

With equal sample sizes, this reduce to

$$\hat{T} = \sqrt{\frac{m}{2}} (\overline{Y}_{i+1} - \overline{Y}_i) / S > t_{g,n,a}$$

Where $ni = m \forall i$ and $t_{g,n,a}$ is obtained from table of critical values developed to test $\varphi$ by [1] which is obtained using ordered statistics and its distribution. However, these critical values are generally smaller than those obtained using
the t-distributions. The implication of this is that the students t-distribution when in use for the ordered alternative provides a more conservative test than the actual [2]. Furthermore, if the test statistic rejects using the percentage points of the t-distribution it would be rejected using the actual critical values developed. However, the rejection or acceptance of the null hypothesis Ho by the proposed test procedure is on the strength of the alternative set H1. In this connection, if Ho is rejected at a specified \( \alpha \) level, it shows that the data being analyzed supports the ordered alternative set of H1.[6] developed the test procedure for testing Ho against ordered means with one of them being the control, assuming homogenous variances. For the general case, we can let \( \delta = \mu_i - \mu_{i+1} \) if and only if \( \mu_i = \max (\mu) \) where \( i = 1, 2, 3, \ldots, g \) but for the control \( \delta = \mu - \mu_{g} \) where \( g \) is fixed and \( \mu_g = \max (\mu_i), i = 1, 2, \ldots, g \) in particular, if \( \mu_{1} \) represents the mean of population \( i, i = 1, 2, \ldots, g \), the problem is to test the hypothesis.

\[
H_0 : \mu = \mu_1 = \ldots = \mu_{g-1}
\]

against

\[
H_1 : \mu \leq \mu_1 \leq \ldots \leq \mu_{g-1}
\]

Where \( \mu = \text{Max}_{1 \leq i \leq g} \mu_i \) is control mean from the \( g \)th population. By the term control, we mean that all other \( g-1 \) treatment means are compared against \( \mu \) and serves as the basis of comparison other.

\[
\mu_i's, \ 1 \leq i \leq g-1
\]

Define \( \delta = \mu - \mu_g \)

Then

\[
\mu < \mu \text{ if and only if } \delta_i < 0 \ \forall i, i = 1, \ldots, g-1.
\]

Therefore, the hypothesis of equation (2.9) becomes

\[
H_0 : \delta_i = 0 \ \forall i, i = 1, \ldots, g-1
\]

against

\[
H_1 : \max_{1 \leq i \leq g-1} \delta_i < 0
\]

The unbiased estimate of \( \delta_i \) is

\[
\hat{\delta}_i = \bar{Y}_i - \bar{Y}_g \quad (5)
\]

Therefore,

\[
\hat{\delta}_i \sim N[\delta_i, \nu(\delta_i)]
\]

Where

\[
\nu(\hat{\delta}_i) = \nu(\bar{Y}_i - \bar{Y}_g)
\]

\[
= \nu(\bar{X}_i) + \nu(\bar{X}_g)
\]

\[
= \frac{\delta^2}{n_i} + \frac{\delta^2}{n_g}
\]

This occurs under equality of variance assumption and independent sampling. Thus,
Consider the quantity \( m_i \) which is defined as
\[
m_i = \left[ \frac{n_i + n_g}{n_i n_g} \right]^{\frac{1}{2}} \quad 1 \leq i \leq g-1
\]

and let
\[
\hat{\gamma}_i = m_i \hat{\delta}_i = m_i (\bar{X}_i - \bar{X}_g)
\]
Then, we have that,
\[
\hat{\gamma}_i \sim N(\gamma_i, \sigma^2)
\]

Where \( \gamma_i = m_i (\mu_i - \mu_g) \)

Since \( E(\hat{\gamma}_i) = m_i E(\hat{\delta}_i) = m_i (\mu - \mu_g) \)

and \( \nu(\hat{\gamma}_i) = \nu [m_i \hat{\delta}_i] = m_i^2 \nu [\hat{\delta}_i] = \sigma^2 \)

Now, let \( \hat{\gamma}^1 = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_{g-1}) \)

and \( \hat{\gamma}^1 = (\gamma_1, \gamma_2, \ldots, \gamma_{g-1}) \)

Then
\[
\hat{\gamma}^1 = \left[ m_i (\bar{X}_i - \bar{X}_g), m_2 (\bar{X}_2 - \bar{X}_g), \ldots, m_{g-1} (\bar{X}_{g-1} - \bar{X}_g) \right]
\]

and
\[
\hat{\gamma} = \left[ m_1 (\mu_1 - \mu_g), m_2 (\mu_2 - \mu_g), \ldots, m_{g-1} (\mu_{g-1} - \mu_g) \right]^T
\]

Therefore, equation (2.11) becomes
\[ H_0 : \gamma_i = 0 \quad \forall i, \quad i = 1, \ldots, g-1 \]

against \( H_i : \max_{1 \leq i \leq g-1} \beta_i < 0 \)

Suppose we define statistic T as
\[
T = \left[ m_1 (\bar{Y}_{i1} - \bar{Y}_1), m_2 (\bar{Y}_{i2} - \bar{Y}_2), \ldots, m_{g-1} (\bar{Y}_{i,g-1} - \bar{Y}_{g-1}) \right]
\]

\( = (T_1, T_2, \ldots, T_{g-1}) \) and \( S^2 \), the pooled sample variance for all the g populations as
\[
S^2 = \frac{1}{n-g} \sum_{i=1}^{g} \left[ \sum_{j=1}^{m_i} (X_{ij} - \bar{X}_i) \right]^2
\]

That is
\[
S^2 = \frac{1}{n-g} \sum_{i=1}^{g} (n_i - 1)S_i^2
\]

Where, \( n = \sum_{i=1}^{g} n_i \)

\( S^2 \) is the sample variance for the i population, independent of \( T \).
Therefore, statistics $U$ is distributed $MT_{g-1}(\gamma, n-g)$ multivariate $T$–distribution.

Define statistic $H$ as, $H = h(u) = \max_{1 \leq i \leq k-1} u_i$. Then $H$ is non-decreasing in $T$ and distributed independent of $S^2$. Thus, we propose a test procedure for testing hypothesis set up as

$$
\phi(u) = \begin{cases} 
1, & \text{if } \max_{1 \leq i \leq k-1} u_i > t_0 \\
0, & \text{otherwise}
\end{cases}
$$

where $\max u_i$ shows that the preconceived order $\mu_i < \mu_{i+1}$ $\forall i$, $i = 1, 2, \ldots, g$ and $t_0$ is determine such that $E_{\mu_0}(\phi(u)) = \alpha, \alpha \in (0,1)$ is predetermined.

In term of $\bar{X}_1$, $\bar{X}_2$, $\ldots$, $\bar{X}_g$, and $S^2$, becomes

$$
\phi(\bar{Y}, S^2) = \begin{cases} 
1, & \text{if } \max_{1 \leq i \leq g-1} \sqrt{n_i n_g} \frac{(Y_i - \bar{Y}_g)}{S} > t_0 \\
0, & \text{otherwise}
\end{cases}
$$

That is in testing (2.9) the rule is to reject the null hypothesis $H_0$ at a specified type I error if

$$
\hat{T} = \max_{1 \leq i \leq g-1} \sqrt{n_i n_g} \frac{(Y_i - \bar{Y}_g)}{S} < t_0
$$

For equal sample sizes $n_i = ng = m \forall i$, then the rule is to reject $H_0$ if

$$
\hat{T} = \max_{1 \leq i \leq g-1} \sqrt{m} \frac{(Y_i - \bar{Y}_g)}{S} < t_0
$$

For $\delta_i$ (2.10) define as $\delta_i = \mu - \mu_i$ $i = 1, \ldots, g-1$ Test $\lambda$ reduces to the two sample one side t-test with $g = 2$ and inherits the three optimality properties of Uniformly Most Powerful Unbiased (UMPU), Uniformly Most Powerful Invariant (UMPI) and Asymptotical Consistency of the two sample one tailed test, [11]

In this connection, test $\varphi$ reduces to the two samples one tail $t$-test when $g - 1 = 1$ and it possesses the three optimality properties of UMPU, UMPI and consistency using $H_0$ against $H_1$.

Now, if $\tilde{\delta_i}$ of (2.10) is redefined as $\tilde{\delta_i} = \mu - \mu_i$ then, for $i = 1, \ldots, g-1$, $\mu < \mu_i$, if and only if $\tilde{\delta_i} > 0 \forall i$, and hypothesis set of (2.11) will become

$$
H_o : \tilde{\delta_i} = 0 \forall i, \quad i = 1, \ldots, g-1
$$

against

$$
H_1 : \min_{1 \leq i \leq g} \tilde{\delta_i} > 0
$$

Following other procedure all through, test $\varphi$ of (18) becomes
\[
\phi(\bar{Y}, S^2) = \begin{cases} 
1, & \text{if } \min_{1 \leq i \leq k - 1} \left( \frac{n_i n_g}{n_i + n_g} \frac{\bar{Y}_i - \bar{Y}_g}{S} \right) > t_0 \\
o, & \text{otherwise}
\end{cases}
\]

For some \(t_0\). Hence, (2.19) and (2.20) respectively become

\[
\hat{\phi} = \min_{1 \leq i \leq k - 1} \left( \frac{n_i n_g}{n_i + n_g} \frac{\bar{Y}_i - \bar{Y}_g}{S} \right) < t_0
\]

\[
\hat{\phi}_g = \min_{1 \leq i \leq k - 1} \left( \frac{m}{2} \frac{\bar{Y}_g - \bar{Y}_i}{S} \right) > t_0
\]

REFERENCES


Saadu, Yusuf olanrewaju was born on 23rd of December, 1984. He finished his National Diploma and Higher National Diploma in Statistics from Kwara State Polytechnic, Ilorin in 2005 and 2008 respectively and he just concluded his Post Graduate Diploma in Statistics from University of Ilorin July 2013. Currently am working at Kwara State Polytechnic, Ilorin as system Analyst in Library Department.

Adeoye Akeem O was born on October 7th 1976 in Owu Isin, Kwara State of Nigeria. He obtained his Higher National Diploma (H.N.D) in Statistics from the Kwara State Polytechnic Ilorin in 1999, PGD Statistics in 2008 from University of Ilorin, PGDE Education in 2007 from University of Ado-Ekiti, Teacher training Certificate (TTC) in 2010 from Federal College of Education Akoka Lagos State, Masters degree (M.Sc) in Statistics in 2010 from University of Ilorin, Nigeria and currently on his Ph.D in Statistics at the University of Ilorin, Nigeria. He has been lecturing Statistics in Kwara State Polytechnic as an Instructor from April 2005 to March 2013. He joined Federal Polytechnic Offa from April 2013 to date as a lecturer. He has attended many workshops and conferences. He has thirteen journals for both national and international.

Yusuf Gbemisola A (Mrs) was born on 1984 from Ajase in Ifelodun Local Government Area of Kwara State of Nigeria. She finished his National Diploma and Higher National Diploma in Statistics from Federal Polytechnic, Offa in 2004 and 2007 respectively and she just concluded her Post Graduate Diploma in Education from Kwara State College of Education Ilorin August 2013.