In this research paper, we have discussed tests based on adjusted p values such that, if the adjusted p value for an individual hypothesis is less than the chosen significance level $\alpha$, then the hypothesis is rejected, when the entire family of tests is considered. It includes Bonferroni procedure and modification of that procedure by Holm, Holand & Copenhaver, Hommel, Hochberg and Rom. From them some of the methods are single step procedure and others are sequential methods. Further Sequential methods can be categorized in two ways i.e. step up method and step down method.

**Single Step/ Simultaneous Procedure:** It is also called Single Step (SS) procedure. The single step procedure sets a single criterion for testing all individual hypotheses. SS procedure conducts all comparisons regardless of any other comparison is significant or not using a constant critical value (Einot & Gabriel, 1975). SS procedures are valid to use both for hypothesis testing and to calculate confidence intervals.

In SS procedure, the decision about any hypothesis $H_i$ does not depend on the decision about any other hypothesis $H_j$ therefore the hypotheses can be tested without reference to one another.

**Sequential Procedure:** It is also called Step Wise (SW) procedure. A step wise procedures consider either the significance of the omnibus test or the significance of other comparisons or both in evaluating the significance of a particular comparison.

In SW procedure, the hypotheses are tested in a specific order, generally determined by the magnitudes of the test statistics or the associated p-values, $p_i$ and the decisions on them are made in a stepwise manner. The decisions on the earlier hypotheses in the order may affect those on the later hypotheses in the order.

A major disadvantage of stepwise (multi-stage) method is that it does not allow the construction of confidence interval, which is extremely useful for the interpretation of the results.

SW procedures can be further subdivided into 2 categories.

i. Step Up procedures (SD)

ii. Step Down procedures (SU)

**Step Up procedures:** In SU procedure, the hypotheses are tested beginning with the least significant one and testing continues until a hypothesis is rejected at which point all the remaining hypotheses are rejected by implication without actually testing them (Tamhane & Dunnett, 1999).

SU procedure begins by testing all minimal hypotheses and then steps up through the hierarchy of implying hypotheses. If any hypothesis is rejected, then all of its implying hypotheses are rejected without further tests; thus a hypothesis is tested if and only if all of its implied hypotheses are retained.

**Step Down Procedure:** A step down procedure begins by testing the overall intersection hypothesis and then steps down through the hierarchy of implied hypotheses. If any hypothesis is not rejected, all of its implied hypotheses are retained without further tests; thus a hypothesis is tested if and only if all of its implying hypotheses are rejected, in which case the decision continues until a hypothesis is not rejected at which point all the remaining hypotheses are accepted by implication without actually testing them.
2. Non-parametric Tests:
These methods can be classified on the basis of the either simultaneous or sequential procedure.

2.1. Bonferroni Test (1961):
The Bonferroni method applies to both continuous and discrete data. This method is flexible because it controls the FWE for tests of joint hypotheses about any subset of m separate hypotheses (including individual contrasts). The procedure will reject a joint hypothesis H0 if any p-value for the individual hypotheses included in H0 is less than α/c. Bonferroni method, however, yields conservative bounds on Type I error hence it has low power. This procedure controls the FWE at α without any further assumption on the dependence structure of the p value.

This method is designed for comparisons involving pair wise comparisons as well as combinations of means, provided the number of comparisons to be made is fixed in advance. It is recommended for non-orthogonal contrast because it splits the type I error rate equally among all comparisons.

The Bonferroni procedure is used for evaluating a small number of contrast that is selected prior to observing the data while preserving a selected family wise type I error rate (Ott & Longnecker, 2010). The researcher must have sufficient theory about the phenomena of interest in order to know which contrasts to specify. It is appropriate when the number of comparisons exceeds the number of degrees of freedom between groups. This method controls type I error but it will increase type II error. The purpose of Bonferroni procedure is to reduce the probability of identifying significant results that do not exist, that is, to guard against making Type I error in the testing process (Mchugh, 2011). These potential for error increases with an increase in the number of tests being performed in a given study and is due to the multiplication of probabilities across the multiple tests (Mchugh, 2011).

Test statistics:

\[ t_{i,j} = \frac{\bar{X}_i - \bar{X}_j}{MS_e \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \quad \ldots(1) \]

Where,
- \( MS_e \) = Error mean square from the ANOVA.
- \( \bar{X}_i \) and \( \bar{X}_j \) are the two means being compared.
- \( n_i \) and \( n_j \) are the respective sample sizes from population i and j.

Critical Value:

\[ t_{\alpha/2, df_e} \quad \ldots(2) \]

Where,
- \( t \) is the value from the t distribution.
- \( c \) is the number of pair wise comparisons in the family.

For complex comparison:

\[ t_{cal} = \left[ \sum \frac{c_i^2 \bar{X}_i}{SE} \right] \quad \ldots(3) \]

Decision procedure:
Reject the null hypothesis if \( t_{cal} \geq t_{\alpha/2, df_e} \); do not reject \( H_0 \) otherwise.

Confidence Interval:

\[ \bar{X}_i - \bar{X}_j \pm t_{\alpha/2, df_e} \sqrt{MS_e \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \quad \ldots(4) \]

Here, margin of error depends on the number of comparisons.

Confidence interval for the contrasts:

\[ \sum_{i=1}^{k} c_i \bar{X}_i \pm t_{\alpha/2, df_e} \sqrt{MS_e} \sqrt{\sum_{i=1}^{k} \frac{c_i^2}{n_i}} \quad \ldots(5) \]

Advantages:

1. This method is highly flexible because it can be applied to test any subset of hypotheses for continuous, discrete data and even correlated tests.
2. This method is used for any design.
3. When number of comparisons are small (i.e. number of comparison less than or equal to number of groups-1) it gives smaller confidence interval.
4. This method is useful in confirmatory research when a family of selected pair wise comparisons is specified prior to data collection; it reduces the problem of alpha inflation (Mchugh, 2011).
5. Bonferroni method can also test complex pairs. (Mchugh, 2011).
6. This method has relatively good power for small sets of planned comparison (Toothaker, 1993).

Disadvantages:

1. It cannot be used for data snooping because the tests of interests are specified prior to data analysis.
2. This method has lower power to reject an individual hypothesis and it lacks power if several highly correlated tests are undertaken (Simes; 1986; Li, 2009; Hommel, 1988).
3. It has power to quickly decline as the number of comparisons increases (Olejnik et al., 1997; Toothaker, 1993).
4. It is not a tool for exploratory data analysis.
5. The test does not take into account whether the findings are consistent with theory and past research. If consistent with previous findings and theory, an individual result should be less likely to be a Type I error.

2.2. Holm Test (1979):
It is a modification of Bonferroni procedure that yields a more powerful test. The goal of Holm method is to increase the power of the statistical tests while keeping under control the FWE. It is a step down procedure. It is also called a sequential rejection method because it examines each hypothesis in an ordered sequence and the decision to accept or reject the null hypothesis depends on the results of the previous hypothesis tests (Tamhane et al., 1998). He was the first to formally introduce a sequentially rejective Bonferroni procedure. Bonferroni method does not account for the correlations between the test statistics, the Holm procedure can be improved.

Holm method can be applied to almost any data because of its non-parametric nature. This test can be applied in any pair wise comparison where the classical Bonferroni test is usually applied. It is applicable when pair wise comparisons of median or linear combinations or non linear combinations of median are used. It is used to perform priori comparison. For several a priori contrasts, not necessarily pair wise, it controls FWE while at the same time maximizes the power (Howell, 2007).

Assumptions:
There are no restrictions on the type of test; the only requirement is that it should be possible to calculate the obtained level for each separate test. Further, there are no problems to include in the analysis only for the a priori interesting hypotheses, while more special multiple tests usually include on all hypotheses of a certain kind.

Holm's procedure may be used either as a protected test or as an unprotected test but the protected version is preferred due to the additional power gains. But when there exist logical implications among the hypotheses, problems arise which we have to take in to consideration (Holm, 1979). So, Holm's procedure makes no distributional assumptions, logical assumptions about the hierarchy of the hypotheses to be tested and does not assume independence of comparisons.
Critical Value:
\[
\frac{\alpha}{c - i + 1}
\] ...

Decision procedure: Reject \(H_i\) to \(H_c\) if
\[
P(i) \leq \frac{\alpha}{c - i + 1}
\]
\(u\) will change at all stages because of its step down nature.

The critical value of this method is based on the Bonferroni inequality.

Advantages:
1. This method is flexible and simple to implement.
2. It controls the FWE in the strong sense, i.e. it guarantees control of generalized Type I error probability to be at most \(\alpha\) (Hochberg, 1988; Schochet, 2008; Ekenetierna, 2004; Hochberg & Benjamini, 1990; De Muth, 2006).
3. This archives lower Type II error while keeping the Type I error rate at level of rejection using \(\alpha' = \alpha/j'\), where \(j'\) is the largest number in \(J\).
4. It becomes very conservative when the numbers of comparisons are large and when tests are not independent (De Muth, 2006).

2.3. Holland & Copenhaver Test (1987):
It uses the Sidak (1967) inequality to set the criterion for each hypothesis test. It is a step down procedure. When there is need for further research in situations, where there is no logical inter relationship among the hypotheses, this method is useful.

Assumptions:
Positive orthant dependence of the test statistics is satisfied.

Procedure:
Let \(p_{i1},...,p_i\) be the ordered p values (smallest to largest) and \(H_{i1},...,H_{ic}\) be the corresponding hypotheses. Suppose \(i\) is the smallest integer from 1 to \(m\) such that \(p(I) > 1 - (1 - \alpha)^{m-c}\). The Holland-Copenhaver procedure rejects \(H_{i1}\) to \(H_{ic}\) and retains \(H_{i1}\) to \(H_{ic}\) (Olejnik et al., 1997).

Test Statistics:
For unequal sample size, the test statistics is same as Bonferroni given by ...

Critical value:
\[
1 - (1 - \alpha)^{m-c}
\] ...

Decision procedure:
Reject \(H_i\) to \(H_c\) if
\[
p(i) < 1 - (1 - \alpha)^{m-c}
\]

Advantages:
This method is conservative under the condition that the test statistics are positive orthant dependent.

Disadvantages:
Applicability of this method is slightly less than the Holm procedure because of the requirement of positive orthant dependent condition for test statistics.

2.4. Hommel Test (1988):
Hommel (1988) employs the closure principle to extend Simes test and developed a stepwise multiple testing procedure controlling FWE. It is based on the Simes (1986) equality. This is a step up method and it is protected test. This procedure is conservative only when the test statistics are independent, because it based on the Simes equality for independent p values. It is not always necessary to test every possible combination of hypothesis i.e. it can also be used for few comparisons.

The work of Hommel's who generalized Simes procedure that it gives strong control of FWE whenever Simes original procedure does achieve weak control (e.g. with independent tests).

Assumptions:
Test statistics are independent.

Procedure:
Reject all hypothesis that have a p value \(\leq \alpha/j'\) where \(j'\) is defined as
\[
j = \max\{i \in \{1,...,c\}; p(i) > \frac{k \alpha}{i'} \text{ for } k = 1,...,i'\}
\]
\(j\) is non empty, reject \(H_i\) whenever \(P(i) \leq \alpha/j'\) with \(j' = \max\). If \(j\) is empty, reject all \(H_{i1},...,H_{ic}\).

This procedure includes two stages. The first stage uses the obtained \(p\) values to compute the number of members in \(J\). The second stage obtains the significance level of rejection using \(\alpha = \alpha/j'\), where \(j'\) is the largest number in \(J\).

Test statistics is same as Holm given in ...

Critical Value:
\[
\alpha/j'
\]

Advantages:
1. It protects the FWE only when test statistics are independent (Dmitrienko et al., 2009; Olejnik et al., 1997).
2. The uniqueness of the Hommel procedure is that it not only considers the order of the tests but also takes the obtained \(p\) values into the calculation while computing the \(j'\).

Disadvantages:
1. This method is relatively complicated.
2. When correlations between variables are negative, the test can sometimes...
2.5. Hochberg Test (1988)

It is a modification of Dunn procedure. This procedure uses critical values identical to those used in Holm's procedure. This procedure starts by examining the largest p-value p(c). If p(c) ≤ α, then H(c) and all other hypotheses are rejected. If not, H(c) is not rejected and one proceeds to compare p(c-1) with α/2. If the former is smaller, then H(c-1) and all hypotheses with smaller p-values are rejected. Generally, one proceeds from highest to lower p-values, retaining H0, if its p-value satisfies p(i) > α(c-i+1). One stops the procedure at the first ordered hypothesis when that inequality is reversed. This hypothesis is rejected and all hypotheses with lower or equal p-values. This is always a sharper procedure than Holm's.

Critical Value:

\[ \alpha' = \frac{\alpha}{c-i+1} \]  

…(13)

Decision procedure:

Reject H_i to H_c for any i=c,c-1,...,1 if

\[ p(i) \leq \frac{\alpha}{c-i+1} \]  

…(14)

Advantages:

1. This procedure has strong control over the FWE even if the freedom combination condition is not satisfied (Holm, 1979; Holland & Copenhaver, 1987; Olejnik et al, 1997).

2. It controls the FWE under the same conditions for which the Simes global test control the Type I error rate.

3. This method always achieves the same type I FWE control and lower type II error rates (Hochberg & Benjamini, 1999).

4. It has nice characteristic that no adjusted p value can be larger than the largest of the unadjusted P values (Wright, 1992).

5. This method is able to reject at least one individual hypothesis when the null hypothesis is rejected. This property of consonance makes Hochberg procedure easy to interpret (Rom, 1990).

Disadvantages:

1. It lacks the stability under certain conditions, for example, when the test statistics are dependent or correlated (Schochet, 2008).

2. It only can be applied in the independent hypotheses tests (Olejnik, et al. 1997; Schochet, 2008).

2.6. Rom Test (1990):

It is a modification of Hochberg procedure to increase the statistical power. It is a step up procedure. Increased power is achieved by identifying the appropriate adjusted significance levels that control the Type I error rate at exactly the nominal level when test statistics are independent (Olejnik et al, 1997).

Assumptions:

Test statistics are independent.

Procedure:

The Rom procedure differs from the Hochberg procedure when the adjusted significance level is obtained. Both procedures set \( \alpha' = \alpha \) equal to \( \alpha \) and \( \alpha' = \alpha/2 \) equal to \( \alpha/2 \), but the remaining \( m - 2 \) adjusted significance levels differ. The adjusted significance levels are determined recursively as

\[ \alpha'_{m-i} = \left[ \sum_{i=1}^{m-1} \alpha'_{m-i} \right] \left| \sum_{j=1}^{i-1} \alpha'_{m-j} \right| \left( \frac{\alpha}{m-i+1} \right) \]  

…(15)

where \( \alpha_{i} = \alpha \) and \( \alpha_{1} = \alpha/2 \).

It is step up procedure with different critical value of \( c_{1}=\alpha, c_{2}=\alpha/2, c_{3}=\alpha/3 + \alpha/12 \) etc.
Because step up sequential multiple comparisons are based on the Simes equality, which assumes independence of comparisons, it is reasonable to suggest that dependence or correlation between the means of groups should affect the Type I error control and power (Zweifel, 2014).

In summary, the comparison of (Bonferroni, Holm, Holland, Hochberg, Hommel, Rom), Bonferroni procedure has the lowest percentage of rejections and Hommel procedure has the highest percentage of rejections whenever differences exist among the procedures. Overall, the SU procedures are little more powerful than the SD procedures. Within the SU procedures, whenever differences occurred, the Hommel procedure has slightly higher percentage of rejections than the Hochberg procedure. Within the SD procedure, whenever difference occur, Holland procedure having a slightly higher percentage of rejection than Holm procedure (Olejnik et al., 1997).

Confidence Interval: All the methods are step wise methods except Bonferroni so confidence interval cannot be obtained by any of the method so comparison is not possible with respect to Confidence Interval.

Simulation Study: This section discuss results regarding tests to be reported as adjusted p values such that, if the adjusted p value for an individual hypothesis is less than the chosen significance level \( \alpha \), then the hypothesis is rejected with FWE not more than \( \alpha \). It includes Bonferroni procedure and modification of that procedure by Holm, Holland & Copenhaver, Hommel, Hochberg and Rom.

As a concrete example, imagine that we have ten p values, and they are (in order from smallest to largest) as follows: 0.002, 0.0054, 0.007, 0.008, 0.009, 0.0094, 0.012, 0.015, 0.026, and 0.067.

We will compare probability with critical value based on Bonferroni method and modification of that procedure by Holm, Holland & Copenhaver, Hommel, Hochberg and Rom.

Table 1: Rejection criteria according to different available Tests

Table 2: Hypotheses Rejection by all these multiple comparison procedure

Table 3: Comparison of Multiple Comparison procedure

REFERENCES: