Multiple Comparison Tests: Tukey's HSD(Honest Significant Difference)

Dr.Arnab Chakraborty *   Mrinmoy Banik †

12th October 2022

Abstract

Multiple Comparison Tests are like series of hypothesis tests which are different from performing each one the individual tests with some fixed level of significance. Instead an overall Family Wise Error Rate or False Discovery Rate is maintained. Hence sophisticated tests like Fisher's LSD and Tukey's HSD were developed. We had discussed Fisher's LSD in the previous section so in this section we will discuss Tukey's Honest Significant Difference method.

1 Overview

Named after John Tukey, in this technique we have the following setup:-

$$y_1, y_2, y_3, ..., y_n \;\; independent$$
$$y_i \overset{\text{iid}}{\sim} \mathbb{N}(\mu, \sigma^2)$$
$$\hat{\sigma}^2 \sim \frac{\sigma^2}{p}\chi^2(p)$$
$$y_i \perp \sigma^2 \;\; \forall i \in \{1, 2, 3, ..., n\}$$

And suppose we are interested in testing:-

$$H_{0ij} : \mu_i = \mu_j \;\; \forall i \neq j$$

*Department of Applied Statistical Division, Indian Statistical Institute,Kolkata Email:http://www.isical.ac.in/ arnabc/

†This portion of the book is latex-ed and edited by him,Student of B.Stat 3rd yr at ISI,BS2012

Then Tukey's HSD is one technique for accepting or rejecting each one of those $\binom{k}{2}$ many null-hypothesis.

2 Procedure

It involves the following steps:-

- The standard range is computed from the observed y_i's, i.e $max(y_i) - min(y_i)$
- The sample standard deviation is computed, i.e $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$
- This range is then standardized, i.e divided by $\hat{\sigma}$
- This standardized range , denoted by **Q** is our test statistics and has a distribution independent of $\hat{\sigma}^2$ & the common value of μ_i if they all the null hypohesis indeed were true.

$$\therefore \mathbf{Q} = \frac{max(y_i) - min(y_i)}{\hat{\sigma}}$$

2.1 Fact:-

If H_{0ij}'s are all true, hen Q has a known distribution.By known distribution we mean something that doesn't depend on unknown $\hat{\sigma}^2$ & the unknown common value of μ_i.
A simple intuitive explanation goes like this, if we add a const C to all y_i's then the statistic **Q** remains unaffected. Similarly if we multiply a constant to y_i's then both the numerator and denominator gets multiplied by the same number and hence gets cancelled out.
This known distribution is known as the **Tukey's Q Distribution** an apparently only depends on k.Once we have this distribution and its quantile funcion, we can perform the multiple hypothesis testing using the following rejection criteria:-

$$Reject \;\; H_{0ij} \quad \Longleftrightarrow \quad \frac{|y_i - y_j|}{\hat{\sigma}} > \mathbf{Q}_{(\mathbf{1}-\alpha)}$$

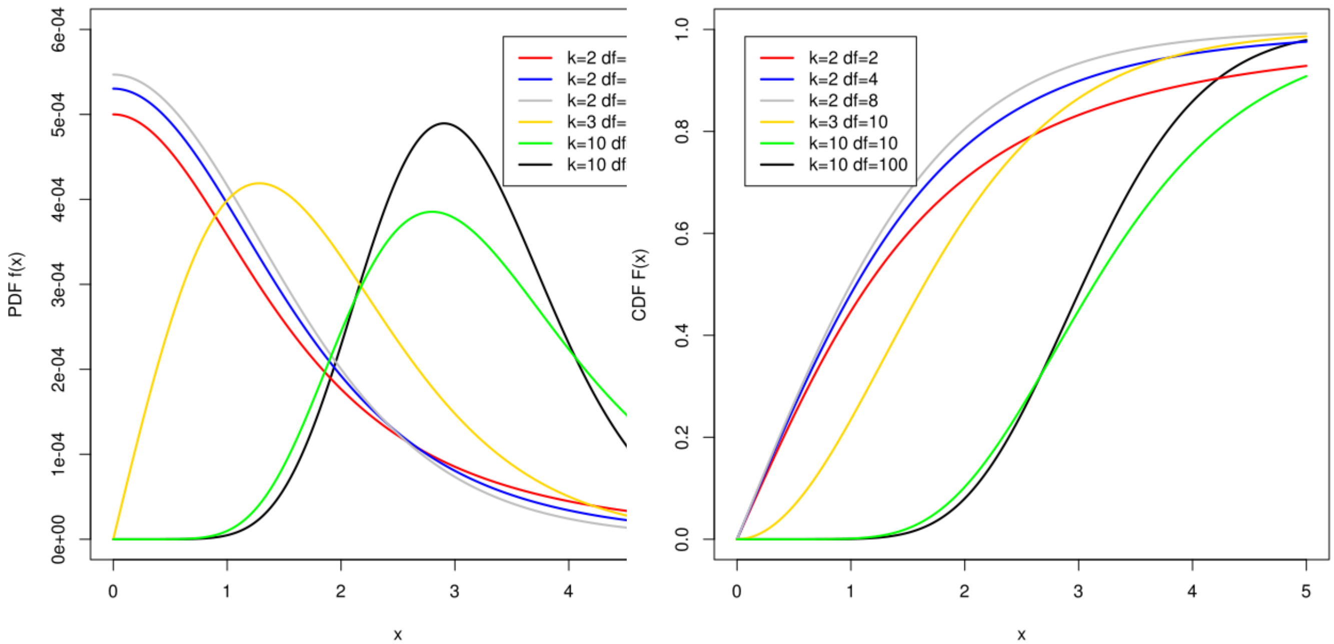
Here $\mathbf{Q}_{(\mathbf{1}-\alpha)}$ is the $(1-\alpha)^{th}$ quantile of the **Tukey's Q Distribution**.

2.2 Claim: FWER is satisfied:-

The Family Wise Error Rate(FWER) the most widely used test quality control is satisfied by this test.**FWER is the probability of rejecting atleast one H_{0ij} when all the H_{0ij}'s are true.**So if for some $i \neq j$,

$$\frac{|y_i - y_j|}{\hat{\sigma}} > \mathbf{Q}_{(\mathbf{1}-\alpha)} \implies \mathbf{Q} > \mathbf{Q}_{(\mathbf{1}-\alpha)} \because |y_i - y_j| < |max_{i=1}^{n}(y_i) - min_{j=1}^{n}(y_j)|$$

Chance of this happening is already α. So the event that atleast one H_{0ij} is rejected is a subset of the event $\mathbf{Q} > \mathbf{Q_{(1-\alpha)}}$, which has probability α. So the FWER of the test is α. The follow FIGURE 1 below shows the density plots of Tukey's Q Distribution:-



(A) PDF of Tukey's Q Distribution (B) CDF of Tukey's Q Distribution

FIGURE 1: The Tukey's Q Distribution is shown for various different values of k and degrees of freedom.(Image credited to Wikipedia)

3 Conclusion

In this section we learned about a test for multiple comparison test for pairwise hypothesis testing. **Tukey's HSD** method is a standardized range test which nicely manages the Family Wise Error Rate and unlike **Fisher's LSD** has better power characteristics, also **Tukey's test** avoids many of the counter-intuitive results that **Fisher LSD** might result when the observations are near the test boundary.

[——————————]