

13.6: Post-hoc Analysis – Tukey's Honestly Significant Difference (HSD) Test⁸⁵

When the Null Hypothesis is rejected in one factor ANOVA, the conclusion is that not all means are the same. This however leads to an obvious question: which particular means are different? Seeking further information after the results of a test is called post-hoc analysis.

The problem of multiple tests

One attempt to answer this question is to conduct multiple pairwise independent same t-tests and determine which ones are significant. We would compare μ_1 to μ_2 , μ_1 to μ_3 , μ_2 to μ_3 , μ_1 to μ_4 , etc. There is a major flaw in this methodology in that each test would have a significance level of α , so making Type I error would be significantly more than the desired α . Furthermore, these pairwise tests would NOT be mutually independent. There were several statisticians who designed tests that effectively dealt with this problem of determining an "honest" significance level of a set of tests; we will cover the one developed by John Tukey, the Honestly Significant Difference (HSD) test.⁸⁶ To use this test, we need the critical value from the **Studentized Range Distribution** (q), which is used to find when difference of pairs of sample means are significant.

The Tukey HSD test

📌 The Tukey HSD test

Tests: $H_o : \mu_i = \mu_j$ $H_a : \mu_i \neq \mu_j$ where the subscripts i and j represent two different populations

Overall significance level of α : This means that **all pairwise tests** can be run at the same time with an overall significance level of α

Test Statistic: $\mathrm{HSD} = q \sqrt{\frac{\mathrm{MSE}}{n_c}}$

q = critical value from Studentized Range table

MSE = Mean Square Error from ANOVA table

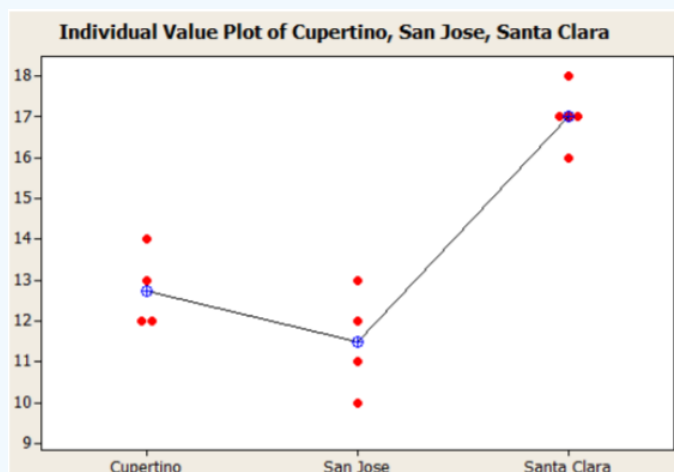
n_c = number of replicates per treatment. An adjustment is made for unbalanced designs.

Decision: Reject H_o if $|\bar{X}_i - \bar{X}_j| > \mathrm{HSD}_{\text{critical value}}$

Computer software, such as Minitab, will calculate the critical values and test statistics for these series of tests. We will not perform the manual calculations in this text.

✓ Example: Party Pizza

Let us return to the Tofu pizza example where we rejected the Null Hypothesis and supported the claim that there was a difference in means among the three restaurants.



In reviewing the graph of the sample means, it appears that Santa Clara has a much higher number of sales than Cupertino and San Jose. There will be three pairwise post-hoc tests to run.

Solution

Design

$$H_o : \mu_1 = \mu_2 \quad H_a : \mu_1 \neq \mu_2 \quad H_o : \mu_1 = \mu_3 \quad H_a : \mu_1 \neq \mu_3 \quad H_o : \mu_2 = \mu_3 \quad H_a : \mu_2 \neq \mu_3$$

These three tests will be conducted with an overall significance level of $\alpha = 5\%$.

The model will be the Tukey HSD test.

Here are the differences of the sample means for each pair ranked from lowest to highest:

Test 1: Cupertino to San Jose: $|\bar{X}_1 - \bar{X}_2| = |12.75 - 11.50| = 1.25$

Test 2: Cupertino to Santa Clara: $|\bar{X}_1 - \bar{X}_3| = |12.75 - 17.00| = 4.25$

Test 3: San Jose to Santa Clara: $|\bar{X}_2 - \bar{X}_3| = |11.50 - 17.00| = 5.50$

The HSD critical values (using statistical software) for this particular test:

$$\text{HSD}_{\text{crit}} \text{ at } 5\% \text{ significance level} = 1.85 \quad \text{HSD}_{\text{crit}} \text{ at } 1\% \text{ significance level} = 2.51$$

For each test, reject H_o if the difference of means is greater than HSD_{crit}

Test 2 and Test 3 show significantly different means at both the 1% and 5% level.

The Minitab approach for the decision rule will be to reject H_o for each pair that does not share a common group. Here are the results for the test conducted at the 5% level of significance:

Data/Results

Refer to the Minitab output. Santa Clara is in group A while Cupertino and San Jose are in Group B.

```

Grouping Information Using Tukey Method

      N    Mean  Grouping
Santa Clara  5  17.0000  A
Cupertino   4  12.7500  B
San Jose    4  11.5000  B

Means that do not share a letter are significantly different.

```

Conclusion

Santa Clara has a significantly higher mean number of tofu pizzas sold compared to both San Jose and Cupertino. There is no significant difference in mean sales between San Jose and Cupertino.