11.5.1: Pairwise Comparison Post Hoc Tests for Critical Values of Mean Differences

As just discussed , a post hoc test is used only after we find a statistically significant (reject the null hypothesis) result and need to determine where our differences truly came from. This implies that when the null hypothesis is retained, you do not need to conduct pairwise comparisons; if there's no differences between the means, why would you look for where a mean difference is? The term “post hoc” comes from the Latin for “after the event”.

Mean Differences

The next set of post-hoc analyses compare the difference between each pair of means, then compares that to a critical value. Let's start by determining the mean differences. Table $11.5.1.1$ shows the mean test scores for the three IV levels in our job applicant scenario.

Table 11.5.1.1- Average Test Scores by Job Applicant Degree

IV Levels	Average Test Score ($\overline{X}$)
No Degree	43.7
Related Degree	84.3
Unrelated Degree	63.2

Using the average test scores for each level of the IV (Degree), we could find the difference between each pair of means:

$$\overline{X}_N - \overline{X}_R = 43.7 - 84.3 = -40.6$$

$$\overline{X}_N - \overline{X}_U = 43.7 - 63.2 = -19.5$$

$$\overline{X}_R - \overline{X}_U = 84.3 - 63.2 = 21.1$$

We could avoid negative answers if we always put the bigger number first, but Dr. MO likes to follow the order from the research hypotheses. As with all critical values, we use the absolute value of the calculated statistic anyway, so the big issue is to remember is to make sure that you subtract each mean from each other mean.

That’s it on mean differences, let’s now learn about some ways to calculate a critical value to compare these mean differences to; as always:

$$\boldsymbol{Critical < |Calculated| = \text{Reject null} = \text{means are different} = p < .05}$$

$$\boldsymbol{Critical > |Calculated| = \text{Retain null} = \text{means are similar} = p > .05}$$

Pairwise Comparisons

For this type of post-hoc analysis, you compare each of these mean differences (that you just calculated by subtracting one mean from another mean) to a critical value. What should you do if the calculated mean difference is further from zero (bigger) than the critical value? Yep, you reject the null hypothesis and say that those two means are different from each other! Sound familiar? It’s just like when using a critical z, critical t, or critical F.

As with converting the raw p-values, the critical mean difference can be computed in different ways. Tukey’s Honestly Significant Difference will be discussed here, but just know that there are other types of pairwise comparison tests that statistical software can complete with ease.

Pairwise Comparison Steps:

1. Compute a mean difference for each pair of variables.
2. Find the critical mean difference.
3. Compare each calculated mean difference to the critical mean.
4. Decide whether to retain or reject the null hypothesis for that pair of means.

Tukey's HSD (Honestly Significant Difference)

Tukey's HSD corrects for the alpha inflation caused by doing a bunch of statistical tests by correcting for the probability of a Type I error so that all the pairwise comparisons *combined* is p<.05, sorta like the Bonferroni that we just discussed. Tukey's makes it so that each individual pairwise comparison has a much smaller p-value.

Formula:

$$HSD = q * \sqrt{\frac{MSw}{n_{group}}}$$

There are a couple things to know about this formula. First, the q is found in yet another table. You probably won't need this too much because statistical software tells you which means are different from which other means, but you can find a table of q-values at this Real-Statistics.com webpage: https://www.real-statistics.com/statistics-tables/studentized-range-q-table/; make sure that you use the Alpha = 0.05 set of tables!

To find the q-value in the table, you find the column with the total number of groups in the analysis (k) on the top, then find the Degrees of Freedom for the denominator (df_W) on the side. Remember, there are different tables for the different alpha levels so don't just use the first table.

Example 11.5.1.1

What would the critical value (Tukey's HSD) be for our job applicant scenario?

Solution

Let's first find the q-value. We know that we have 3 groups (k = 3), and 30 participants. The df_W is calculated as N – k, so that's 30 – 3 = 27. According to the Alpha = 0.05 table at the table of q-values, q = 3.506.
Now we can use the q-value we just found and information from the ANOVA Summary Table to complete this formula:

$$HSD = q * \left(\sqrt{\frac{MSw}{n_{group}}}\right)$$

$$HSD = 3.506 * \left(\sqrt{\frac{111.85}{10}}\right)$$

$$HSD = 3.506 * (\sqrt{11.185})$$

$$HSD = 3.506 * (3.34)$$

$$HSD = 11.73$$

Okay, but what do we *do* with this number? Tukey's HSD is a critical value, so if any of your mean differences are bigger than this critical value of 3.82, then the null hypothesis is rejected and that set of means is statistically significantly different from each other. If the critical value (Tukey's HSD that you just calculated) is bigger than any of the differences between the means, then you retain the null hypothesis and say that they are not different enough to think that they are from different populations.

$$\boldsymbol{Critical < |Calculated| = \text{Reject null} = \text{means are different} = p < .05}$$

$$\boldsymbol{Critical > |Calculated| = \text{Retain null} = \text{means are similar} = p > .05}$$

Let's look at that in Table 11.5.1.2

Table 11.5.1.2- Interpreting Pairwise Comparison Results

Means Compared	Mean Difference	Is Tukey's HSD of 11.73 Smaller than the Absolute Value of the Mean Difference?	Reject or Retain the Null Hypothesis?	Are the Means Similar or Different?
$\overline{X}_N - \overline{X}_R = 43.7 – 84.3$	-40.6	Yes	Reject the null hypothesis.	These two means are different from each other.
$\overline{X}_N - \overline{X}_U = 43.7 – 63.2$	-19.5	Yes	Reject the null hypothesis.	These two means are different from each other.
$\overline{X}_R - \overline{X}_U = 84.3 – 63.2$	21.1	Yes	Reject the null hypothesis.	These two means are different from each other.

But what does it mean?

To complete this analysis of the job applicant scenario, we need to go back to the research hypothesis. We had said that the applicants with No Degree will have a lower average test score than those with a Related Degree, but those with No Degree will have a similar average test score to those with an Unrelated degree. The average test score for those with a Related Degree will also have a higher average test score compared to those with an Unrelated Degree. In symbols, this looks like:

- $\overline{X}_N < \overline{X}_R$
- $\overline{X}_N = \overline{X}_U$
- $\overline{X}_R > \overline{X}_U$

Now, let's look at our mean differences. All of them were statistically significantly different, which means that our research hypothesis was partially supported.

- RH1: $\overline{X}_N < \overline{X}_R$ = Supported; those without a degree scored statistically significantly lower than those with a related degree.
- RH2: $\overline{X}_N = \overline{X}_U$ = Not supported; the research hypothesis was that the group without a degree would have average test scores that were similar to the group with an unrelated degree, but those with an unrelated degree actually scored higher than those with no degree.
- RH3: $\overline{X}_R > \overline{X}_U$ = Supported; those with a related degree did score higher than those with an unrelated degree.

And that leads us to our final write-up!

Write-Up

? Exercise 11.5.1.1

Write up a conclusion to the job applicant scenario.

Answer

Look at your conclusion. Did you include the four components that should be in all write-ups?

- The statistical test is preceded by the descriptive statistics (means).
- The description tells you what the research hypothesis being tested is.
- A "statistical sentence" showing the results is included.
- The results are interpreted in relation to the research hypothesis.

Look at Dr. MO's conclusion.

> *The research hypothesis that the applicants with No Degree will have a lower average test score than those with a Related Degree, but will have a similar average test score to those with an unrelated degree, and that the average test score for those with a Related Degree will also have a higher average test score compared to those with an Unrelated Degree*

> *was partially supported ($F(2,27) = 36.86$, $p < 0.05$. Those with a Related Degree ($M = 84.3$) did score higher than both those with an Unrelated Degree ($M = 63.2$) and those with No Degree ($M = 43.7$). However, it was hypothesized that those with an Unrelated Degree would have similar average test scores than those with No Degree, when those with an Unrelated Degree actually scored significantly higher than those with No Degree.*

Note

Did Dr. MO include the four components that should be in all write-ups?

Make sure to include the statistical sentence for the ANOVA with both Degrees of Freedom (with the smaller one from the numerator first), but you don't have to include all of that information for each pairwise comparison of the mean differences.

Summary

There are many more post hoc tests than just the ones discussed here, and they all approach the task in different ways, with some being more conservative and others being more powerful. In general, though, they will give highly similar answers. What is important here is to be able to interpret a post hoc analysis.

So, that's it! You've learned a Between Groups ANOVA and pairwise comparisons to test the null hypothesis! Let's try one full example next!

This page titled 11.5.1: Pairwise Comparison Post Hoc Tests for Critical Values of Mean Differences is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michelle Oja.

- **11.8: Post Hoc Tests** by Foster et al. is licensed CC BY-NC-SA 4.0. Original source: https://irl.umsl.edu/oer/4.