

12.2: Post hoc Comparisons

If H_0 is rejected in a one-way ANOVA, you will frequently want to know where the differences in the means are. For example if we tested $H_0 : \mu_1 = \mu_2 = \mu_3$ and rejected H_0 in a one-way ANOVA then we will want to know if $\mu_1 \neq \mu_2$ or $\mu_2 \neq \mu_3$, etc.

To see which means are different after doing an ANOVA we could just compare all possible combinations of pairs using t -tests. But such an approach is no good because the assumed type I error rates, α , associated with the t -tests would be wrong. The α rate would be higher because in making such *multiple comparisons* you incur a greater chance of making an error.

So we need to correct our test statistic and/or the corresponding α value when we do such multiple comparisons. We will cover two such multiple comparison approaches in detail :

1. Scheffé test
2. Tukey test

and we will look at the Bonferroni approach.

Doing multiple comparisons after an ANOVA is known as post hoc testing. It is the traditional approach for comparing several means. The opening “omnibus” ANOVA lets you know if there are any differences at all. If you fail to reject the ANOVA H_0 then you are done. Only when you reject H_0 do you put in the effort of comparing means pairwise. This traditional approach, designed to minimize the necessary calculations, is not the only way to compare multiple means. The other approach is to forget about the ANOVA and then use t -tests to compare means pairwise on in combinations^[1] of means until you use up the N degrees of freedom in the dataset. Here we will stick with the traditional approach.

12.2.1 Scheffé test

The test statistic for the Scheffé test is

$$F_{s,\text{test}} = F_s = \frac{(\bar{x}_i - \bar{x}_j)^2}{s_W^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \quad (12.2.1)$$

Note that F_s is basically a t^2 quantity (recall that $F_{1,\nu} = t^2$) but with a pooled estimate s_p^2 of the common population variance σ given by the value of s_W^2 from the ANOVA. In other words F_s uses information from all of the data to estimate σ instead of from just groups i and j as a t -test would (see [Equation 10.5](#)). Note that the Scheffé test does not require equal group sizes n_i .

The critical statistic is a modification of the critical statistic from the ANOVA is

$$F'_{\text{crit}} = F' = (k-1)F_{\alpha,\nu_1,\nu_2} = (k-1)F_{\text{crit,ANOVA}} \quad (12.2.2)$$

where ν_1 and ν_2 are the ANOVA degrees of freedom. The critical statistic is the same for all pairwise comparisons regardless of the sample sizes, n_i and n_j , of the pair of groups being compared.

Example 12.2 : The ANOVA of [Example 12.1](#) found that at least one of the three means was different from the others. Use the Scheffé test to find the significant differences between the means. There has to be at least one.

Solution :

0. Data reduction.

Collect the necessary information from the omnibus ANOVA. We'll need:

$$\begin{aligned} n_1 = n_2 = n_3 &= 6 \quad \bar{x}_1 = 15.5, \bar{x}_2 = 4.0, \bar{x}_3 = 5.83 \\ s_W^2 &= 45.5 \quad F_{\text{crit,ANOVA}} = F_{(0.05, 2, 15)} = 3.68 \end{aligned}$$

1. Hypotheses.

There are 3 hypotheses pairs to test :

$$H_0 : \mu_1 = \mu_2, \quad H_0 : \mu_1 = \mu_3, \quad H_0 : \mu_2 = \mu_3 \quad (12.2.3)$$

$$H_1 : \mu_1 \neq \mu_2, \quad H_1 : \mu_1 \neq \mu_3, \quad H_1 : \mu_2 \neq \mu_3 \quad (12.2.4)$$

2. Critical statistic.

One value for all three hypothesis tests:

$$F_{\text{crit}} = (k - 1)F_{0.05, 2, 15} = (3 - 1)(3.68) = (2)(3.68) = 7.367 \quad (12.2.5)$$

3. Test statistic.

There are three of them:

μ_1 vs. μ_2 :

$$F_s = \frac{(\bar{x}_1 - \bar{x}_2)^2}{s_W^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \frac{(15.5 - 4.0)^2}{45.5 \left(\frac{1}{6} + \frac{1}{6} \right)} = 8.72 \quad (12.2.6)$$

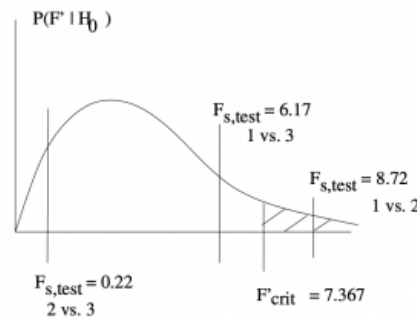
μ_1 vs. μ_3 :

$$F_s = \frac{(\bar{x}_1 - \bar{x}_3)^2}{s_W^2 \left(\frac{1}{n_1} + \frac{1}{n_3} \right)} = \frac{(15.5 - 5.83)^2}{45.5 \left(\frac{1}{6} + \frac{1}{6} \right)} = 6.17 \quad (12.2.7)$$

μ_2 vs. μ_3 :

$$F_s = \frac{(\bar{x}_2 - \bar{x}_3)^2}{s_W^2 \left(\frac{1}{n_2} + \frac{1}{n_3} \right)} = \frac{(4.0 - 5.83)^2}{45.5 \left(\frac{1}{6} + \frac{1}{6} \right)} = 0.22 \quad (12.2.8)$$

4. Decision.



For μ_1 vs. μ_2 , reject H_0 . For μ_1 vs. μ_3 , do not reject H_0 . For μ_2 vs. μ_3 , do not reject H_0 .

5. Interpretation.

The results of the Scheffé test at $\alpha = 0.05$ conclude that only the mean numbers of interchange employees between toll roads 1 and 2 are significantly different.

□

12.2.2 Tukey Test

The test statistic for the Tukey test is

Rendered by QuickLaTeX.com

where, again, s_W^2 is from the omnibus ANOVA, \bar{x}_i is the mean of group i and we must have equal sample sizes for all groups: $n = n_i$ for all i . There is a Tukey test statistic for unequal n , and it is used by SPSS, but we won't cover that here.

The critical statistic, q_{crit} , comes from a table of critical values from a new distribution called the q distribution. The critical values are tabulated in the [Tukey Test Critical Values Table](#). To use this table, you need two numbers going in :

1. k = number of groups
2. $\nu = \nu_2$ = degrees of freedom for s_W^2

Reject H_0 when $q_{\text{test}} > q_{\text{crit}}$. In this case we don't have a picture of the q distribution handy (although it is basically the absolute value of t), so we just

use the $\text{image } q_{\{\text{mbox}\{\text{crit}\}\}}$ title="Rendered by QuickLaTeX.com" height="16" width="66" style="vertical-align: -6px;"> rule similar to how we use the p -value.

Example 12.3 : Repeat Example 12.2 using the Tukey test instead of the Scheffé test.

Solution : 0. Data Reduction.

We use the same data from the omnibus ANOVA :

$$n_1 = n_2 = n_3 = 6 \quad \bar{x}_1 = 15.5, \bar{x}_2 = 4.0, \bar{x}_3 = 5.83$$

$$s^2_{\text{W}} = 45.5 \quad F_{\text{crit, ANOVA}} = F_{(0.05, 2, 15)} = 3.68$$

1. Hypotheses.

The 3 hypotheses pairs to test are the same :

$$H_0 : \mu_1 = \mu_2, \quad H_0 : \mu_1 = \mu_3, \quad H_0 : \mu_2 = \mu_3 \quad (12.2.9)$$

$$H_1 : \mu_1 \neq \mu_2, \quad H_1 : \mu_1 \neq \mu_3, \quad H_1 : \mu_2 \neq \mu_3 \quad (12.2.10)$$

2. Critical statistic.

Use the **Tukey Test Critical Values Table**. Go into the table with

- Number of groups = $k = 3$.
- $\nu = \nu_2 = N - k = nk - k = (6)(3) - 3 = 18 - 3 = 15$.

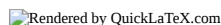
and $\alpha = 0.05$ to find

$$q_{\text{crit}} = 3.67 \quad (12.2.11)$$

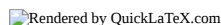
3. Test statistic.

Again, there are three of them :

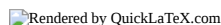
μ_1 vs. μ_2 :



μ_1 vs. μ_3 :



μ_2 vs. μ_3 :



4. Decision.

Reject H_0 when $\text{image } q_{\{\text{mbox}\{\text{crit}\}\}}$ title="Rendered by QuickLaTeX.com" height="16" width="66" style="vertical-align: -6px;">. This only happens for one hypothesis pair : For μ_1 vs. μ_2 , reject H_0 . For μ_1 vs. μ_3 , do not reject H_0 . For μ_2 vs. μ_3 , do not reject H_0 .

5. Interpretation.

The results of the Tukey test at $\alpha = 0.05$ conclude that only the mean numbers of interchange employees between toll roads 1 and 2 are significantly different. (Same result as the Scheffé test. Usually this happens but when it doesn't, you need to use some kind of non-mathematical judgement.)

12.2.3 Bonferroni correction

A more conservative (less power) approach to multiple comparisons (post hoc testing) is to use Bonferroni's method. The fundamental idea of the Bonferroni correction is to add the probabilities of making individual type I errors to get an overall type I error rate.

Implementing the idea is simple. Do a bunch of t -tests and multiply the p -value by a correction factor C . There are a number of ways to choose C (you will have to dig to find out which method SPSS uses). The easiest (and most conservative) is to set C equal to the number of pairwise comparisons done. So if you have k groups then C is given by the binomial coefficient:

$$C = \binom{k}{2}. \quad (12.2.12)$$

Another way is to look at the total degrees of freedom, ν_{pairs} , associated with the pairwise t -tests and compare it to the total degrees of freedom in the data, $\nu = N$ (or one could argue $\nu = N - 1$), to come up with

$$C = \frac{\nu_{\text{pairs}} + \nu}{\nu}. \quad (12.2.13)$$

Since there is some ambiguity as to what we should use for C , we will not do Bonferroni post hoc testing by hand. However, be able to recognize Bonferroni results in SPSS, treating the value of C as an SPSS blackbox parameter.

-
1. Combinations of means may be compared using "contrasts". For example $\mu_1 + \mu_2$ might be compared with $2\mu_3$. Contrasts are not covered in Psy 234. [↩](#)
-

This page titled [12.2: Post hoc Comparisons](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Gordon E. Sarty](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.