

16.8: Post Hoc Tests

Time to switch to a different topic. Let's suppose you've done your ANOVA, and it turns out that you obtained some significant effects. Because of the fact that the F-tests are "omnibus" tests that only really test the null hypothesis that there are no differences among groups, obtaining a significant effect doesn't tell you which groups are different to which other ones. We discussed this issue back in Chapter 14, and in that chapter our solution was to run t-tests for all possible pairs of groups, making corrections for multiple comparisons (e.g., Bonferroni, Holm) to control the Type I error rate across all comparisons. The methods that we used back in Chapter 14 have the advantage of being relatively simple, and being the kind of tools that you can use in a lot of different situations where you're testing multiple hypotheses, but they're not necessarily the best choices if you're interested in doing efficient post hoc testing in an ANOVA context. There are actually quite a lot of different methods for performing multiple comparisons in the statistics literature (Hsu 1996), and it would be beyond the scope of an introductory text like this one to discuss all of them in any detail.

That being said, there's one tool that I do want to draw your attention to, namely Tukey's "Honestly Significant Difference", or **Tukey's HSD** for short. For once, I'll spare you the formulas, and just stick to the qualitative ideas. The basic idea in Tukey's HSD is to examine all relevant pairwise comparisons between groups, and it's only really appropriate to use Tukey's HSD if it is *pairwise* differences that you're interested in.²⁴⁵ For instance, in `model.2`, where we specified a main effect for drug and a main effect of therapy, we would be interested in the following four comparisons:

- The difference in mood gain for people given Anxifree versus people given the placebo.
- The difference in mood gain for people given Joyzepam versus people given the placebo.
- The difference in mood gain for people given Anxifree versus people given Joyzepam.
- The difference in mood gain for people treated with CBT and people given no therapy.

For any one of these comparisons, we're interested in the true difference between (population) group means. Tukey's HSD constructs **simultaneous confidence intervals** for all four of these comparisons. What we mean by 95% "simultaneous" confidence interval is that there is a 95% probability that *all* of these confidence intervals contain the relevant true value. Moreover, we can use these confidence intervals to calculate an adjusted p value for any specific comparison.

The `TukeyHSD()` function in R is pretty easy to use: you simply input the model that you want to run the post hoc tests for. For example, if we were looking to run post hoc tests for `model.2`, here's the command we would use:

```
TukeyHSD( model.2 )
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = mood.gain ~ drug + therapy, data = clin.trial)
##
## $drug
##              diff          lwr          upr      p adj
## anxifree-placebo 0.2666667 -0.1216321 0.6549655 0.2062942
## joyzepam-placebo 1.0333333 0.6450345 1.4216321 0.0000186
## joyzepam-anxifree 0.7666667 0.3783679 1.1549655 0.0003934
##
## $therapy
##              diff          lwr          upr      p adj
## CBT-no.therapy 0.3222222 0.0624132 0.5820312 0.0186602
```

The output here is (I hope) pretty straightforward. The first comparison, for example, is the Anxifree versus placebo difference, and the first part of the output indicates that the observed difference in group means is .27. The next two numbers indicate that the 95% (simultaneous) confidence interval for this comparison runs from -.12 to .65. Because the confidence interval for the difference includes 0, we cannot reject the null hypothesis that the two group means are identical, and so we're not all that surprised to see that the adjusted p-value is .21. In contrast, if you look at the next line, we see that the observed difference between Joyzepam and

the placebo is 1.03, and the 95% confidence interval runs from .64 to 1.42. Because the interval excludes 0, we see that the result is significant ($p < .001$).

So far, so good. What about the situation where your model includes interaction terms? For instance, in `model.3` we allowed for the possibility that there is an interaction between drug and therapy. If that's the case, the number of pairwise comparisons that we need to consider starts to increase. As before, we need to consider the three comparisons that are relevant to the main effect of `drug` and the one comparison that is relevant to the main effect of `therapy`. But, if we want to consider the possibility of a significant interaction (and try to find the group differences that underpin that significant interaction), we need to include comparisons such as the following:

- The difference in mood gain for people given Anxifree and treated with CBT, versus people given the placebo and treated with CBT
- The difference in mood gain for people given Anxifree and given no therapy, versus people given the placebo and given no therapy.
- etc

There are quite a lot of these comparisons that you need to consider. So, when we run the `TukeyHSD()` command for `model.3` we see that it has made a *lot* of pairwise comparisons (19 in total). Here's the output:

```
TukeyHSD( model.3 )
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = mood.gain ~ drug * therapy, data = clin.trial)
##
## $drug
##              diff          lwr          upr          p adj
## anxifree-placebo 0.2666667 -0.09273475 0.6260681 0.1597148
## joyzepam-placebo 1.0333333 0.67393191 1.3927348 0.0000160
## joyzepam-anxifree 0.7666667 0.40726525 1.1260681 0.0002740
##
## $therapy
##              diff          lwr          upr          p adj
## CBT-no.therapy 0.3222222 0.08256504 0.5618794 0.012617
##
## $`drug:therapy`
##              diff          lwr
## anxifree:no.therapy-placebo:no.therapy 0.10000000 -0.539927728
## joyzepam:no.therapy-placebo:no.therapy 1.16666667 0.526738939
## placebo:CBT-placebo:no.therapy 0.30000000 -0.339927728
## anxifree:CBT-placebo:no.therapy 0.73333333 0.093405606
## joyzepam:CBT-placebo:no.therapy 1.20000000 0.560072272
## joyzepam:no.therapy-anxifree:no.therapy 1.06666667 0.426738939
## placebo:CBT-anxifree:no.therapy 0.20000000 -0.439927728
## anxifree:CBT-anxifree:no.therapy 0.63333333 -0.006594394
## joyzepam:CBT-anxifree:no.therapy 1.10000000 0.460072272
## placebo:CBT-joyzepam:no.therapy -0.86666667 -1.506594394
## anxifree:CBT-joyzepam:no.therapy -0.43333333 -1.073261061
## joyzepam:CBT-joyzepam:no.therapy 0.03333333 -0.606594394
## anxifree:CBT-placebo:CBT 0.43333333 -0.206594394
## joyzepam:CBT-placebo:CBT 0.90000000 0.260072272
## joyzepam:CBT-anxifree:CBT 0.46666667 -0.173261061
##              upr          p adj
## anxifree:no.therapy-placebo:no.therapy 0.7399277 0.9940083
## joyzepam:no.therapy-placebo:no.therapy 1.8065944 0.0005667
## placebo:CBT-placebo:no.therapy 0.9399277 0.6280049
## anxifree:CBT-placebo:no.therapy 1.3732611 0.0218746
## joyzepam:CBT-placebo:no.therapy 1.8399277 0.0004380
## joyzepam:no.therapy-anxifree:no.therapy 1.7065944 0.0012553
## placebo:CBT-anxifree:no.therapy 0.8399277 0.8917157
## anxifree:CBT-anxifree:no.therapy 1.2732611 0.0529812
## joyzepam:CBT-anxifree:no.therapy 1.7399277 0.0009595
## placebo:CBT-joyzepam:no.therapy -0.2267389 0.0067639
## anxifree:CBT-joyzepam:no.therapy 0.2065944 0.2750590
## joyzepam:CBT-joyzepam:no.therapy 0.6732611 0.9999703
## anxifree:CBT-placebo:CBT 1.0732611 0.2750590
## joyzepam:CBT-placebo:CBT 1.5399277 0.0050693
## joyzepam:CBT-anxifree:CBT 1.1065944 0.2139229
```

It looks pretty similar to before, but with a lot more comparisons made.

This page titled [16.8: Post Hoc Tests](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Danielle Navarro](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.