STAT 215: Exploring the Familywise Error Rate

Colin Reimer Dawson

Last Revised November 30, 2016

Let's do an experiment to examine what happens when we do lots of pairwise tests. Work in pairs or threes for this lab. You don't need to hand anything in; we will go over it together.

First, let's suppose that we have a dataset with 10 groups of 20 observations each, where all the means are equal. That is, we are constructing data for which we know the underlying population model, and in particular *we know that the null hypothesis is true*, and all the population means are equal.

Enter the following R code to create a synthetic dataset with this property. If you are running this in a script or a Markdown document, first type `set.seed(SOME_NUMBER)` so that you can get the same results every time you re-run or re-Knit your code.

```
## create 200 observations of a response variable drawn from a single
## common Normal population (in particular, there is one population mean)
response <- rnorm(n = 20 * 10, mean = 50, sd = 10)
## Create a grouping variable with ten values (groups), assigning 20
## observations to each group
groups <- rep(c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J"), each = 20)
## Combine these into a dataset
FakeData <- data.frame(Y = response, X = groups)
```

Take a look at the dataset to verify that it looks as expected. Now let's fit the ANOVA model.

```
the.model <- lm(Y ~ X, data = FakeData)
```

Plot the data by group, highlighting the means, and get quantitative descriptive statistics as well.

```
library("mosaic")
xyplot(Y ~ X, data = FakeData, type = c("p","a"))
favstats(Y ~ X, data = FakeData)
```

Which groups have sample means that are farthest apart?

Filter the data to include only the two groups that have the biggest difference in means, and do a t-test for those groups to see whether we would reject the null hypothesis that the means are equal. For example,

```
filter(FakeData, X %in% c("A","H")) %>% t.test(Y ~ X, data = .)
```

In fact, we can get P-values for all possible pairwise comparisons of pairs of group means as follows:

```
with(FakeData, pairwise.t.test(Y, X, p.adjust.method = 'none'))
```

Exercises

1. How many pairwise comparisons are there in total? How many of these return a significant difference (at the 0.05 level)?
2. Given that we know exactly the population model that generated the data, is it correct to reject H_0 for any pairs?
3. If you had a dataset like this handed to you and did all pairwise tests, how many times do you expect you would reject H_0 mistakenly? What is it called when this happens?
4. What is your conclusion if, instead of (or before) doing all of these pairwise comparisons, you do the overall F test?
5. Why is it a bad idea to look at the data before deciding which differences you think will matter?
6. Why might it be a good idea to do an overall F-test *before* doing any pairwise comparisons?