

158 Multiple Testing Lab

Student

4/8/2020

Global analysis:
given all different cold treatments
is there a difference
Local analysis: which treatment should i
take

We saw Bonferroni's inequality last time, which only used the fact that probabilities are non-negative in establishing the multiple testing fix. Thus, it works anytime you are dealing with probabilities that are non-negative (that's always!). But if there is other information about the tests that you are doing, and how they might relate to each other, you can exploit that information and get better bounds (you want to use all the α you are allotted, since this is more power for your testing). We want to get $P(\text{type I error somewhere})$ as close to .05 as possible.

Tukey's idea is based on knowing exactly what we are looking at. We are looking at how far apart a collection of \bar{Y}_i are from each other, looking at pairwise differences. This is Tukey's idea.

Think about a collection of \bar{y}_i generated under the null, and see how far apart the furthest two are from each other (the biggest difference is known as the range):

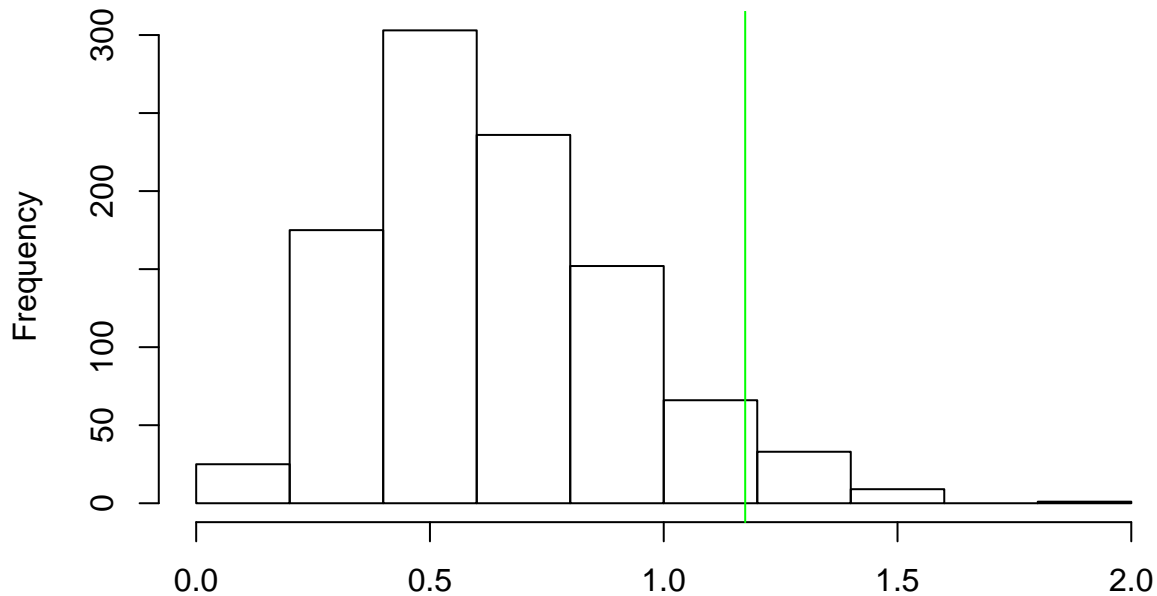
```
n <- 10
y1 <- mean(rnorm(n)); y2 <- mean(rnorm(n)); y3 <- mean(rnorm(n)); y4 <- mean(rnorm(n))
biggest.diff <- max(c(abs(y1-y2),abs(y1-y3),abs(y1-y4),abs(y2-y3),abs(y2-y4),abs(y3-y4)))
biggest.diff
```

```
## [1] 0.708227
```

Now do that over and over again, to get a sense of how far two means can reasonably be from each other when the null is true.

```
n <- 10
biggest.diff <- c()
for (i in 1:1000) {
  y1 <- mean(rnorm(n)); y2 <- mean(rnorm(n)); y3 <- mean(rnorm(n)); y4 <- mean(rnorm(n))
  biggest.diff[i] <- max(c(abs(y1-y2),abs(y1-y3),abs(y1-y4),abs(y2-y3),abs(y2-y4),abs(y3-y4)))
}
hist(biggest.diff)
abline(v=quantile(biggest.diff, .95), col='green')
```

Histogram of biggest.diff



biggest.diff

Only 5

percent of the time did the biggest range exceed the green line. Thus, if I get new data, and two sample means are further apart than that, I conclude that they are estimating different values of μ_i . Under the null, I'll only conclude this wrongly 5 percent of the time.

This value depends on the number of levels of my variable (and thus the number of comparisons), the standard deviation of the data σ and the sample sizes. To deal with the standard deviations, we “studentize” or turn them into t scores. Thus, this histogram turns into what is known as the studentized range distribution. The other two quantities become parameters of the distribution (k and $df(SSE)$).

As you have seen, we can do all this with TukeyHSD in R. As proof of concept, I'll compare the size of an interval using Bonferroni and Tukey.

```
data(iris)
fit <- aov(Sepal.Width~Species, data=iris)
TukeyHSD(fit)

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Sepal.Width ~ Species, data = iris)
##
## $Species
##              diff          lwr          upr      p adj
## versicolor-setosa -0.658 -0.81885528 -0.4971447 0.0000000
## virginica-setosa   -0.454 -0.61485528 -0.2931447 0.0000000
## virginica-versicolor 0.204  0.04314472  0.3648553 0.0087802
```

For Versicolor vs Setosa, the CI for $\mu_{Ve} - \mu_S$ is (-.819, -.497).

Now, if I had used Bonferroni, I'd have to use $\frac{.05}{3}$, and I would get

```
means <- mean(~Sepal.Width|Species, data=iris)
c(means[2]-means[1] + qt(.025/3, 147) * sqrt(anova(fit)$Mean[2]*(1/50 + 1/50)), means[2]-means[1] + qt(
```

```
## versicolor versicolor
## -0.8225234 -0.4934766
```

It's bigger! k is only 3 here, so the savings isn't that great.

How to Compute Arbitrary CIs

CIs are great. They not only tell you whether you'd reject a null hypothesis (is 0 contained in the interval?), but they tell you how much different two means are (for instance). Do them instead of report p-values (or at least in addition to p-values).

We might have cause to ask more interesting questions than $\mu_i - \mu_{i'}$.

A result we used earlier in the semester for calculating the standard deviations of $\hat{\beta}_i$ will be useful again. It is, for X_i independent,

$$\text{Var}(\sum a_i X_i) = \sum a_i^2 \sigma_i^2 \quad \text{where} \quad \text{Var}(X_i) = \sigma_i^2$$

So for the problem above, looking at $\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}$, we have $a_1 = 1$ and $a_2 = -1$, so

$$\text{Var}(\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}) = (1)^2 \frac{\sigma^2}{n_1} + (-1)^2 \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

using the homoskedasticity assumption. Thus, the standard deviation is the square root of that, and the estimated standard deviation (standard error if you will) is

$$s_{\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}} = \sqrt{MSE \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

we assume variance is σ^2 everywhere
if not true, we can transform y

which you might pick out in my code above:

```
sqrt(anova(fit)$Mean[2]*(1/50 + 1/50))
```

```
## [1] 0.06793755
```

Questions about these "pairwise comparisons" or pairwise differences are comparisons of two means. This should sound like a t-test is in order, and in fact this is what we would do with Bonferroni. Defining $D = \mu_i - \mu_{i'}$ with $\hat{D} = \bar{y}_i - \bar{y}_{i'}$, we have as our CI

$$\hat{D} \pm t^*(.025/m; \sum n_i - k) \sqrt{MSE \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

where m is the total number of things you are going to infer.

So what else might we be interested in inferring?

There are different types of things we might be interested in. Consider the vehicles that Pomona owns, for various purposes (athletics, field trips, etc). Suppose that 20 percent are hybrids, 50 percent are non-hybrid sedans, and the rest are passenger vans. We have data about their fuel efficiency (which is clearly variable, depending on the type of driving, the driver, the wind, etc).

Pairwise differences are obvious: do hybrids get better fuel efficiency than the sedans in our fleet, etc.

Do small cars get different fuel efficiency than vans: $C = \frac{\mu_1 + \mu_2}{2} - \mu_3$

What is the average fuel efficiency of my fleet: $L = .2\mu_1 + .5\mu_2 + .3\mu_3$

What is the average fuel efficiency of a hybrid: $L = \mu_1$

In general, we are interested in what we will call linear combinations.

$L = \sum c_i \mu_i$ estimated by $\hat{L} = \sum c_i \bar{y}_i$ weighted sum of various means

If $\sum c_i = 0$, we call it a contrast (hence the C notation above).

If it is such that all the c_i are zero except one which is 1 and the other is -1, we call it a pairwise difference.

Tukey gives a way to nicely control the family error rate for all pairwise differences.

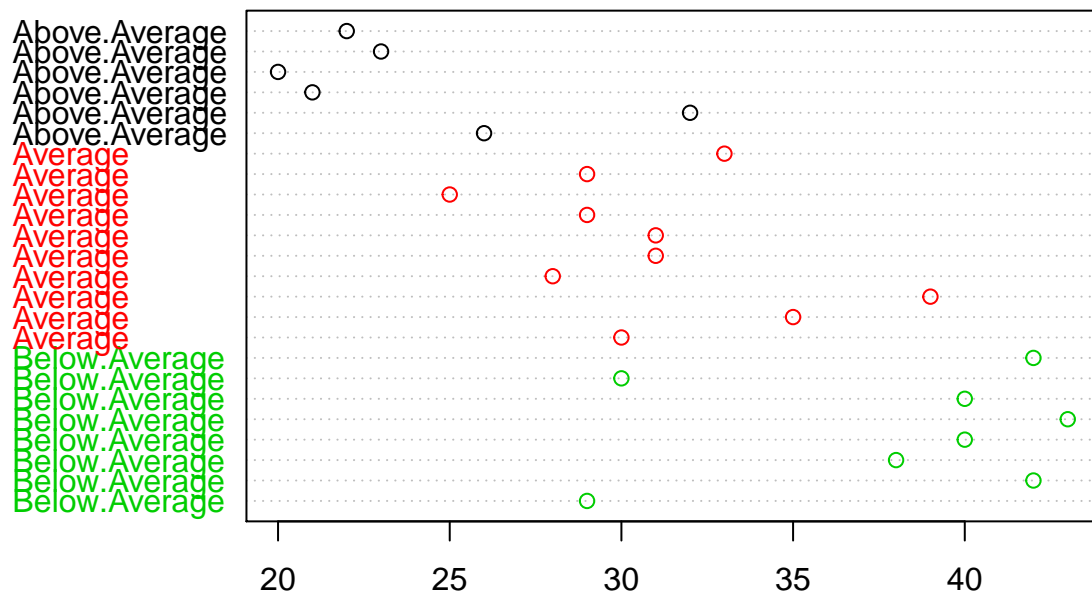
There is a technique called Scheffe's method, that works for all contrasts (all infinitely many of them!). It suffices to know that this exists. We won't get into it.

If you want linear combinations, you have to use Bonferroni.

Questions (Homework):

1. Are these classes of linear combinations nested? Put them in order from smallest to largest.
2. Use the formula for variances of linear combinations above to find (a) the variance of a general linear combination, and (b) the contrast considered in the example above (that isn't the pairwise difference).
3. Suppose that I was interested in the specific 3 linear combinations above, only. What is the value of α that I should use to control the family wise error rate at .05?
4. Find the t-multiplier (using qt(prob, df)) I would use if I had 100 total cars. Note: the df is how much data I had to estimate the unknown σ^2 .
5. If all I was interested in was pairwise differences, which do you think would give me smaller intervals, Tukey's method or Scheffe's?
6. Here's data that looks at the length of rehab required for corrective knee surgery as a function of physical fitness level at time of surgery

```
pt <- c(29, 42, 38, 40, 43, 40, 30, 42, 30, 35, 39, 28, 31, 31,
       29, 25, 29, 33, 26, 32, 21, 20, 23, 22)
fitness <- c(rep("Below.Average", 8), rep("Average", 10), rep("Above.Average", 6))
data.p4 <- data.frame(pt, fitness)
dotchart( data.p4$pt, data.p4$fitness, col=as.numeric(data.p4$fitness)) #an easy plot I don't really li
```



summary of an analysis of this data.

Write a brief