Introduction to Biostatistics (171:161)
Breheny

Lab #14

The goal of today's lab is to give a (brief) introduction to carrying out an ANOVA analysis in SAS/R. As mentioned in class, this is a fairly big topic and we are just going over the basics in this course.

1 Flicker/eye color data

Our sample data set for today comes from a study on "The effect of iris color on critical flicker frequency" from the *Journal of General Psychology*. It's on the web site under the name `flicker`. An individual's critical flicker frequency is the highest frequency at which the flicker in a flickering light source can be detected. At frequencies above the critical frequency, the light source appears to be continuous even though it is actually flickering. This study recorded critical flicker frequency and iris color for 19 subjects.

2 ANOVA analysis

The procedure in SAS for carrying out ANOVA tests is called, unsurprisingly, `PROC ANOVA`. The central element of the procedure is the `MODEL` statement, in which you specify the manner in which you think the outcome depends on other variables. In `R`, one can carry out an ANOVA analysis with the `aov` command. The syntax is as follows:

SAS:

```
PROC ANOVA DATA=flicker;
  CLASS Color;
  MODEL Flicker = Color;
RUN;
```

`R`:

```
fit <- aov(Flicker~Color)
summary(fit)
```

(Note: In SAS, we need a `CLASS` statement, just as in `PROC TTEST`.)

Unfortunately, `R` does not calculate the explained variance by default. From the summary, we need to calculate the fraction of `Sum Sq` that is explained by the model (this is the top line; the bottom line is the variance that is unexplained; together, they add up to give the total sum of squares, RSS):

```
> 23/(23+38.31)
[1] 0.375142
```

So eye color explains 37.5% of the variability in critical flicker detection. This is a larger percentage than we would expect by chance alone: the p-value is 0.02. Once again, this is based on a test statistic (F) of 4.802; if we were doing this by hand, we would have to go to the appropriate F curve to determine the area to the right of 4.802, which would give us our p-value. To get the p-value of the F statistic with a computer, we will use the following code in R:

```
> pf(4.802,2,16,lower.tail = FALSE)
[1] 0.0232
```

(Note: In R, there are separate commands for fitting the model and carrying out tests concerning the reduction in unexplained variability. This may seem annoying now, but as we shall see, there are times when a person wants to do other things to a model besides F-tests, and in those cases it is convenient to have fitting and testing separated.)

3 F Statistic Calculation

To calculate the F-statistic in R like we discussed in class, we will use the data from the previous dataset and the following code:

```
null <- lm(Flicker~1) #Reduced model
complex <- lm(Flicker~as.factor(Color)) #complex model
SSR1 <- sum(null$residuals^2)
SSR2 <- sum(complex$residuals^2)
d0 <- 1 #Number of parameters of the reduced model
d1 <- 3 #Number of parameters of complex model
n <- 19
sigmahat <- SSR2/(n-d1)

((SSR1 - SSR2)/(d1-d0))/sigmahat
[1] 4.802346
```

This yields the same F value as provided by the ANOVA function before. SSR1 and SSR2 denote the residual sums of squares from the null model and the complex model, respectively.

4 Pairwise comparisons of means

Recall that the null hypothesis in an ANOVA test is that the means of the groups are all the same. Our test in the previous section would seem to indicate that this is not the case. This, by itself, is not particularly useful information: what are those sample means? Which eye colors had better flicker detection, and which were worse? Are the individual comparisons significant?

Note that in SAS, `PROC ANOVA` provides a boxplot by default; recall that to make one in R, we use `boxplot`:

```
boxplot(Flicker~Color)
```

The boxplot indicates that blue-eyed individuals have the best flicker detection, while brown-eyed individuals have the worst. We could, if wished, stop at this point: we have evidence that flicker detection depends on eye color and the order seems to go blue > green > brown.

However, one may question whether we really have enough evidence to say that blue- and green-eyed individuals have different average flicker detection abilities. To address that question, we can carry out t-tests for each two-group comparison. Of course, we will have to adjust for multiple comparisons if we do this, since there are 3 possible two-color comparisons.

The most common method for adjusting for multiple comparisons *when conducting pairwise comparisons in an ANOVA model*[1] is an approach developed by John Tukey (my favorite statistician, by the way). Both SAS and `R` provide automatic methods for accomplishing this (both the comparisons and the p-value adjustments). In SAS, we add a `MEANS` statement to `PROC ANOVA`; in `R`, we can use the function `TukeyHSD` (the "HSD" is for "Honest Significant Differences", as opposed to the unadjusted t-tests, which would offer a dishonest picture of significance):

SAS:

```
PROC ANOVA DATA=flicker;
  CLASS Color;
  MODEL Flicker = Color;
  MEANS Color / Tukey;
RUN;
```

`R`:

```
TukeyHSD(fit)
```

Both procedures do the same thing; note that in `R`, we are using our fit that we obtained earlier – this is one of the advantages of separating the fitting and testing.

5 Bonferroni Adjustment

Another common adjustment made to multiple comparisons is using the Bonferroni Adjustment. In `R` the p-values are automatically adjusted so we can compare the p-values to 0.05. In SAS, it's easier to do a multiple comparison and compare the p-values to the new alpha level (0.05/Number of comparisons)

SAS:

```
PROC GLM DATA=flicker;
  CLASS Color;
  MODEL Flicker = Color;
  LSMEANS Color / PDIFF ADJUST = T;
RUN;
```

`R`:

```
pairwise.t.test(Flicker,Color,p.adj="none")
pairwise.t.test(Flicker,Color,p.adj="bonf")
```

(Note: The new alpha level we wish to compare the p-values to in the SAS output is $0.05/3 = 0.017$)

The implication is clear: we have reasonably strong evidence that blue-eyed individuals have better flicker detection than brown-eyed individuals, but we have very little evidence that blue-eyed individuals have better flicker detection than green-eyed individuals, or that green-eyed individuals have better flicker detection than brown-eyed individuals.

[1] Tukey's procedure only applies to this very specific case: pairwise comparisons of means in an ANOVA model. It is not a general method for adjusting p-values like the Bonferroni correction or false discovery rate.