

# Lab 14

April 24-25, 2018

## Objectives:

1. Conduct an ANOVA test for data that has more than two categorical variables
2. Discuss follow-up tests to make pairwise comparisons of categorical variables
3. Review for quiz #4

## ANOVA analysis

We will be working with the flicker/eye color dataset from the class website. 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.

```
flick <- read.delim("http://myweb.uiowa.edu/pbreheny/data/flicker.txt")
attach(flick)
```

The first step of the ANOVA analysis is to create a linear model using the `lm()` function. The `lm()` function takes a parameter  $Y \sim X$  which is similar to other functions we have used in the past. We will use the `lm()` function to examine the relationship between eye color and flicker, where flicker is our outcome of interest (Y) and eye color is our explanatory variable (X). Once we have saved the linear model, we will use the `anova()` function to understand how much of the total variation in "flicker" is explained by our model. The equation uses the SS for group divided by the sum of SS for group and SS residuals to compute how much variability is explained by eye color.

```
model1 <- lm(Flicker~Color)
anova(model1)
```

```
## Analysis of Variance Table
##
## Response: Flicker
##           Df Sum Sq Mean Sq F value Pr(>F)
## Color      2  22.997  11.4986   4.8023 0.02325 *
## Residuals 16  38.310   2.3944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**Note:** You must calculate variability differently than what was explained in class because the output does not give you RSS0.

The output is giving you:

RSS1-RSS0= 23

RSS1=38.31

To find RSS0 you need to find  $RSS1+(RSS1-RSS0)$

From this output, we can find how much variability is explained by eye color using:  $23/(38.31 + 23)$ .

If this confuses you, find  $RSS1$  and  $RSS0$  using this code to calculate variability the same way we did in class:

```
model1 <- lm(Flicker~Color)
model0 <- lm(Flicker~1)
RSS1 <- sum(model1$residuals^2)
RSS0 <- sum(model0$residuals^2)

#formula used in class to find explained variability:
(RSS0-RSS1)/RSS0
```

```
## [1] 0.3751145
```

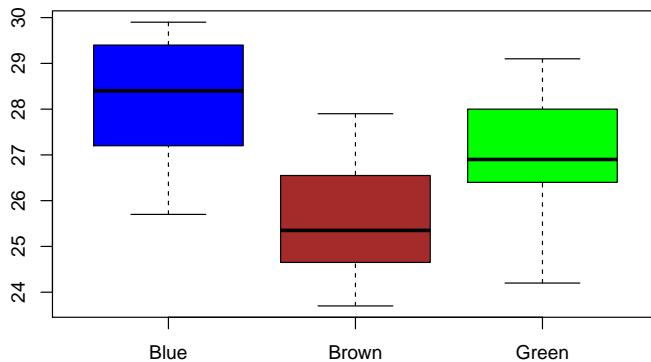
Thus, eye color explains 37.5% of the variability in critical flicker detection.

Could this variability explained by eye color be due to chance? The p-value is 0.02, so it is quite unlikely that variability explained by eye color is due to chance alone. If you are curious this p-value is based on the F statistic of 4.8023. *Overall, do you think eye color impacts flicker detection? Can we determine how eye color impacts flicker from the ANOVA test?*

## Pairwise comparison of means

The low p-value provides evidence that flicker detection differed by eye color. However, we do not know how eye colors compare to one another. Which eye color tends to detect flicker the best? Do all eye colors make a significant impact on flicker? First, let's compare flicker detection by eye color using a boxplot.

```
boxplot(Flicker~Color, col=c("blue", "brown", "green"))
```



This boxplot indicates that those with blue eyes have the best flicker detection, while those with brown eyes have the worst; green is somewhere in between.

However, we cannot say whether this difference is significant enough to conclude that blue eyes are better at detecting flicker than brown eyes (as well as other comparisons with green eyes) without performing pairwise tests.

## Multiple Comparisons

### Tukey's "Honest Significant Difference"

We learned in class why it's important to account for each separate comparison you make, in order to avoid making too many type I errors. For ANOVA, the most common method for adjusting for multiple comparisons (and the method you'll have to use on the homework) is called the "Tukey" correction (or "Tukey's Honest Significant Differences"). R makes this relatively easy to accomplish with the `TukeyHSD()` function.

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

```
TukeyHSD(fit)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Flicker ~ Color)
##
## $Color
##          diff          lwr          upr          p adj
## Brown-Blue -2.579167 -4.7354973 -0.422836 0.0183579
## Green-Blue -1.246667 -3.6643959  1.171063 0.3994319
## Green-Brown  1.332500 -0.9437168  3.608717 0.3124225
```

From the output we get an average difference and a lower and upper difference. *Interpret what these intervals tell us*

### Bonferroni Adjustment

Another way is to use Bonferroni adjustment and in R we use the `pairwise.t.test()` function.

```
pairwise.t.test(Flicker,Color, p.adj = "bonf")
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data: Flicker and Color
##
##      Blue Brown
## Brown 0.021 -
## Green 0.606 0.451
##
## P value adjustment method: bonferroni
```

Here R automatically scales your p-value so you can compare it to 0.05 like usual. If you ran the code above with the argument `p.adj = "none"`, you would have to calculate the Bonferroni-adjusted alpha level (here it would be  $.05/3 = .017$ ) and compare results to 0.017 as a threshold for rejection.

From the p-values we can see flicker detection is significantly better for blue eyes compared to brown eyes. However, there is not sufficient evidence to conclude that flicker significantly differs between brown and green, and green and blue eyes.

## Quiz Review

### Two-sample categorical data

#### Chi-square

Be able to create a contingency table with the outcome as the columns and the groups as the rows. Know how to compute a Chi-squared test. The formula for the chi-squared statistic is the sum of  $\frac{(\text{observed}-\text{expected})^2}{\text{expected}}$ . Then find the value on a chi-squared table and **find the area to the right** (always take the complement of the probability you find on the table to compute p-value for observing something as extreme or more extreme than expected).

#### Fisher's exact test:

Know Fisher's tests the same hypothesis as a chi-squared test, but is an **exact test**, not an approximation (chi-square). Especially useful when any expected cell count is below 5, and is necessary when the expected counts are lower than 1.

#### Odds Ratio

Remember from a contingency table this is  $ad/bc$ . It is important you know how to interpret the odds. If the odds ratio  $> 1$ , for example 1.5, then we would say "The odds for (group 1) experiencing (outcome "Y") is 1.5 times the odds of (group 2) experiencing (outcome "Y")."

If the odds ratio is  $< 1$ , for example 0.6, we would say "The odds for (group 1) experiencing (outcome "Y") is 0.6 times the odds of (group 2) experiencing (outcome "Y"). OR "The odds for (group 1) experiencing (outcome "Y") are 40% lower than the odds of (group 2) experiencing (outcome "Y")." The CI for an odds ratio is  $\exp(\log(\text{OR}) \pm Z \cdot \sqrt{1/a + 1/b + 1/c + 1/d})$ , where Z for a 95% CI is 1.96.

### Two-sample continuous data

Recall, this test is used to determine whether the means of two groups from an unpaired dataset are significantly different. We can either use Welch's (computer only) or Student's test (can compute by hand). Know the difference between the two and why it would be appropriate to use one test over the other.

Review methods for handling outliers and poor skews. We can log transform the data, use a Mann-Whitney rank sum test, or use a permutation test. Know the difference between parametric and non-parametric.

### Different study designs

Understand the differences between retrospective, prospective, and cross-sectional studies.

## Example Questions

### Example 1

A study examined 793 individuals who were in bike accidents. It was found that of 147 riders that wore helmets, 17 of them had a head injury. Of the rest of the bikers who did not wear helmets, 428 did not get a head injury.

- 1) Make a contingency table for the data
- 2) Without running any tests, does there appear to be a benefit to wearing a helmet? (hint: odds ratio)
- 3) Make a 95% CI for this odds ratio
- 4) What are the expected counts for the contingency table?
- 5) Calculate chi-squared statistic and interpret

### Example 2

A study compared the miles per gallon of American cars (sample 1) to Japanese cars (sample 2). The sample size for American cars was 249 with a sample mean of 20.145 and sample standard deviation of 6.415. Japanese cars had a sample size of 79, sample mean of 30.481, and sample standard deviation of 6.108. (pooled standard deviation is 6.343)

- 1) If we assume data is normal what test do we run? What else might we consider to determine what test is most appropriate?
- 2) Conduct a t-test comparing the two group means and interpret the results
- 3) Further analysis shows that two of the American cars in the sample were getting less than 5 miles per gallon. How might this affect the test results? How might you remedy this issue?

### Example 3

The Predators (a hockey team) just reached the second round of playoffs. I was curious if the Predators experienced any benefit to playing at home this season, so I gathered the data on how many goals they scored each game and whether they were home or away (regular season only). In home games they scored an average of 3.098 goals in 41 games with a sd of 1.841. In away games (41) they scored 2.237 with a sd of 1.43. The pooled sd is 1.650.

- 1) Which test would you use to examine the association between location and goals scored?
- 2) It seems they did better at home could this be a difference explained by chance alone?

# Answers

## Example 1

1)

```
##           Y    N
## Helmet    17 130
## No Helmet 218 428
```

2) Yes, the odds ratio is  $(130*218)/(17*428) = 3.89$ . The odds of those **not wearing a helmet** experiencing a head injury is 3.89 times greater than those wearing a helmet experiencing a head injury.

Alternately, you could calculate the odds by  $ad/bc=0.26$ , but the group you will interpret the odds for is now switched. To interpret this, you can say: The odds of those who **wore a helmet** experiencing a head injury is 74% lower than the odds of those who did not wear a helmet.

3)

```
log(3.89) + c(1.96, -1.96) * sqrt(1/130+1/428+1/17+1/218)
```

```
## [1] 1.8895635 0.8272548
```

```
exp(c(0.8272548, 1.8895635))
```

```
## [1] 2.287032 6.616480
```

4) Expected probability of having a head injury (both groups):  $(17+218)/793 = 0.296$

```
##           Y    N
## Helmet    43.5 103.5
## No Helmet 191.2 454.8
```

5)

```
(17-43.5)^2/43.5 + (218-191.2)^2/191.2 + (130-103.5)^2/103.5 + (428-454.8)^2/454.8
```

```
## [1] 28.26443
```

```
pchisq(28.26443, df=1, lower.tail = FALSE)
```

```
## [1] 1.058228e-07
```

## Example 2

1) Student's

$$2) t = \frac{20.145 - 30.481}{6.34 * \sqrt{1/249 + 1/79}} = -12.62$$

Using  $(249+79 - 2) = 326$  degrees of freedom, the p-value is less than 0.05; there is a difference in MPG—American cars get fewer miles per gallon.

3) The t-test procedure can be affected by outlier, which can impact the mean and standard error.

Assuming the measurements are accurate you might consider a non-parametric (Wilcoxon Rank Sum) approach to remove the effect of these outliers.

You can also consider a log transformation to shrink the scale of the variables.

Note: A confidence interval log-transformed data will reflect the ratio group means after being exponentiated.

### Example 3

1) Probably Student's

2)  $t = \frac{3.098 - 2.237}{1.650 * \sqrt{1/41 + 1/41}} = 2.3626$ ; p-value = 0.02057