

BIOS 4120: Introduction to Biostatistics
Breheny

Lab # 14

The goal of today's lab is to give a (brief) introduction to carrying out an ANOVA analysis in R. As mentioned in class, this is a fairly big topic and we are just going over the basics in this course. The second half of the lab will be review for the quiz next Thursday, April 30th.

1 Flicker/eye color data

Our sample data set for today comes from a study called "The effect of iris color on critical flicker frequency" from the *Journal of General Psychology*. The data is available on the class website under the name `flicker.txt`. An individual's critical flicker frequency is the highest frequency at which 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 a total of 19 subjects. As usual, we'll load the data into R using the code below:

```
flick <- read.delim("http://myweb.uiowa.edu/pbreheny/161/data/flicker.txt")
head(flick)
attach(flick)
```

2 ANOVA analysis

The procedure for ANOVA analysis in R is similar to that which we used for linear models, except for instead of the `lm()` function, we'll use the `aov()` function. The code below fits an ANOVA model for the Flicker data, and we can take a look at some relevant information by using the `summary()` function on the fit we created.

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

This output tells us several useful things, including the SS for the model and the SS for the residuals, but it doesn't calculate what we'd like to know which is how much of the total variation in "flicker" is explained by our model. To calculate this we must take the SS for Group and divide it by the sum of SS Group and SS Residuals. This calculation can be done using the code below (note where the values I'm using are coming from in the R output).

```
23/ (38.31 + 23) # R^2
```

So eye color explains 37.5% of the variability in critical flicker detection. Is this a bigger difference than we would expect by chance alone? Well the p -value is 0.02, so it is quite unlikely that this evidence is the result of chance alone. Recall this is based on the F statistics, in this case of 4.802. The p -value comes from the area to the right of that statistic under the F-distribution. (What are the degrees of freedom for this distribution?)

3 Pairwise comparisons of means

Recall what that low p -value tells us here; the null hypothesis that we rejected was that the mean flicker detection rate for each eye-color was the same. You may be wondering “OK, so we’ve found that there is some difference in flicker detection based off eye-color. Well which eye-color is best?” This is a very natural question, and there are several methods we use to address it. First, let’s start by visualizing the difference with a 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 we wished stop here; we have evidence that flicker detection depends on eye color and the order seems to go blue > green > brown.

However, we have to be careful here because someone may claim that we don’t have enough evidence to say that, for example, blue and green eyed individuals have different mean flicker detection abilities. To formally test this, we can carry out multiple t -tests for each two-group comparison.

4 Adjusting for multiple comparisons

Tukey’s “Honest Significant Differences”

You saw 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.

```
TukeyHSD(fit)
```

Based off the output, what can we conclude?

Bonferroni Adjustment

Another common adjustment for multiple comparisons is the Bonferroni adjustment. In R, we can do this using the code below:

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

Here R automatically scales your p -value so you can compare it to 0.05 like usual, but you could also calculate the Bonferroni-adjusted alpha level (here it would be $.05/3 = .017$). You’d still have to run the code above except with `p.adj = "none"`.

5 Quiz Review

Topic outline*

1. Two Sample Categorical Data

		Outcome		
		Y	N	
Group	1	a	b	$a+b$
	2	c	d	$c+d$
		$a+c$	$b+d$	$N = a+b+c+d$

- a. Chi-squared test:** This is an *approximate* test to determine whether there is an association between 2 categorical variables, usually presented in the form of a contingency table. Be sure to know how to do this, see example one in this lab.
- Create a table of expected counts
 - Use the formula below to calculate a chi-square statistic

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

- Check the area to the right of the statistic you calculated in step 2; determine if you reject/fail to reject the null hypothesis.
- b. Fisher's Exact test:** Tests the same hypothesis as a chi-squared test. Especially useful when any *expected* cell counts are below 5, and necessary when the expected cell counts get lower than 1.
- c. Odds Ratio:** in a standard contingency table, this can be calculated by using ad/bc .
- If odds ratio > 1 : Let's say the odds ratio is 1.5. Then we'd say "The odds for (group 1) experiencing (outcome "Y") is 1.5 times the odds of (group 2) experiencing (outcome "Y")."
 - If odds ratio < 1 : Let's say the odds ratio is 0.6. Then we'd 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")."

d. Confidence intervals for odds ratios

- Recall this formula:

$$\log(\widehat{OR}) \pm Z_{x\%} \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

- Also note for a 95% confidence interval, $Z_{95\%} = 1.96$.

2. Two Sample Continuous Data

a. Two-sample t-tests:

- i. **Welch's:** Use when the standard deviations in each group are different from one another. This version of the test is especially necessary when both the standard deviations and the sample sizes are different.
- ii. **Student's:** Slightly more powerful than Welch's test, uses the assumption that the standard deviations of the 2 groups are the same. *Know how to do this one by hand using the formulas below.*

$$SE_d = SD_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE_d}$$

1. Where:

- a. SD_p is the pooled standard deviation
- b. SE_d is the standard error of the difference in means
- c. n_1 and n_2 are the sample sizes for each group
- d. \bar{x}_1 and \bar{x}_2 are the sample means for each group

b. Methods for handling outliers and skew

- i. Mann-Whitney/Wilcoxon Rank Sum test
- ii. Log transform data
- iii. Permutation tests

3. Know the difference between retrospective, prospective, and cross sectional studies.

Note: this outline is not comprehensive

6 Example Questions:

1. The following contingency table displays the results of a study to investigate the effectiveness of bicycle safety helmets in preventing head injury (Thompson et al., 1989). The data consist of a random sample of 793 individuals who were involved in bicycle accidents during a specific one-year period.

Wearing Helmet	Head Injury		Total
	No	Yes	
Yes		17	147
No	428		
Total		235	793

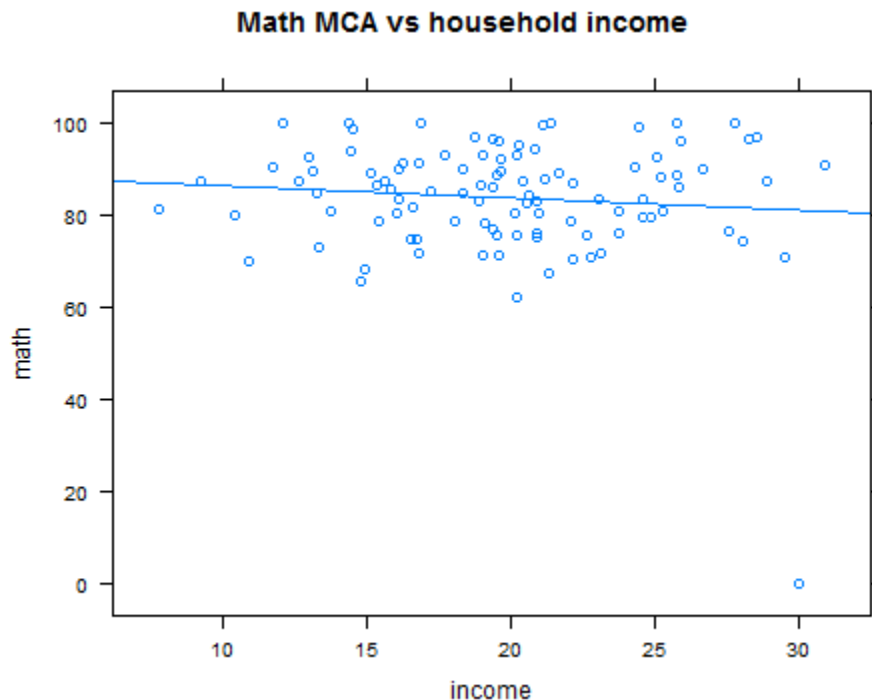
- Fill in the missing slots in the image above.
- Without running any tests for significance, does there appear to be a benefit to wearing a helmet? Answer this by calculating the odds ratio, and interpret.
- Form a 95% confidence interval for the odds ratio.
- What are the expected counts we'd find under the null hypothesis?
- Calculate and interpret the chi-squared statistic of this study.

2. A study compared the miles per gallon of American cars (sample 1) to Japanese cars (sample 2). Some summary statistics are shown below.

	American Cars	Japanese Cars
Sample Size	249	79
Sample Mean	20.145	30.481
Sample Standard deviation	6.415	6.108

The pooled standard deviation is 6.343.

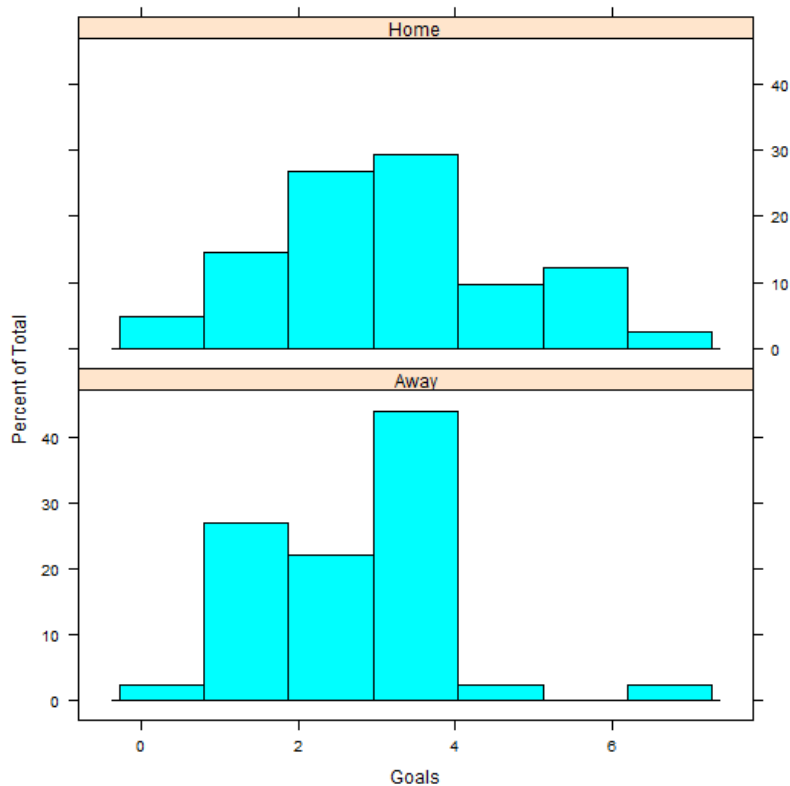
- Assuming the data are normal, what test should we use?
 - Conduct a t-test comparing the mean MPG of Japanese vs. American cars, then interpret the results.
 - Further analysis shows that two of the American cars in the sample were getting less than 5 miles per gallon. How might this affect the test results? How might you remedy this issue?
3. Consider the scatter plot below showing the relationship of MCA math grades to household income. Would you trust the Pearson correlation coefficient ($r = -0.104$)? How else might you analyze this relationship?



4. Try this on your own. The Minnesota Wild (a hockey team) just reached the second round of the playoffs. I was curious if the Wild 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). A summary of the data is below:

	Home Games	Away Games
Sample Size	127	104
Sample Mean score	3.098	2.537
Sample Standard deviation	1.841	1.433

The pooled standard deviation was 1.670. A histogram of the data is presented below.



- Which method is best in this case for testing whether there is an association between location and number of goals scored?
- It seems that the Wild did better at home games than away games this year. Could this difference be the result of chance alone?