Intro to R for Epidemiologists

Lab 5 (2/12/15)

Many of these questions go beyond the information provided in the lecture. Therefore, you may need to use R help files and the internet to search for answers. Feel free to ask questions of the instructor, the TAs, or your classmates, but try to work through as much as you can independently.

For the lab, you are expected to create an R script (.R file in the R editor) with your code corresponding to each question. Begin each question with a commented line of code indicating the question. As an example:

```
# Jenna Krall

# Question 1.
head(iris)
```

Part 1. Statistical Tests for Continuous Data

We will use the `chickwts` dataset to look at statistical tests for continous variables. This dataset contains the weight (in grams) of 6 week old chicks fed one of six different diets. For all of these questions, assume α=.05. (Source: Biometrika, vol. 35)

1. Test the hypothesis that the mean chick weight is 265 grams (ignoring feed groups). (Hint: `?t.test`)

2. Use the `tapply` statement to get the mean weight by feed.

3. Find the difference in mean weight between linseed diet and soybean diet.

4. Test whether the mean weight in the linseed diet group is the same as the mean weight in the soybean diet group.

   - Extract the confidence interval for the difference in mean weights.
   - Extract the p-value for this test

5. Test whether the mean weight in the linseed diet group is the less than the mean weight in the soybean diet group (Hint: check the alternative argument in `t.test`)

Part 2. Statistical Tests for Categorical Data

Suppose we have cross-sectionally collected data on smoking status and cancer status on 206 individuals. Our data looks like this:

	Cancer	No cancer
Smoker	94	27
Non-smoker	22	63

1. Is the proportion of smokers with lung cancer statistically significantly different from the proportion of non-smokers with lung cancer?

   - Create your table `sc_table <- matrix(c(94, 22, 27, 63), ncol = 2)`
   - Name your columns and rows

- Use the function `prop.test`

2. Use a chi-squared test to determine whether smoking and lung cancer are independent (Hint: `chisq.test`)

3. Using the epitools package, extract the odds ratio and its 95% confidence interval for the association between smoking and lung cancer (Hint: `?epitab`).

4. Using the epitools package, find the risk ratio and its 95% confidence interval for the association between smoking and lung cancer (Hint: `?epitab`. Look at the method argument).