**Course Review: Fall 2018**

Part I: Create an R library

Work with a partner to develop a library of all the commands we have learned this semester and what each one is useful for. The library might have the following sections:

1. One continuous variable
   1. Descriptive statistics
   2. Inferential statistics
2. One categorical variable
   1. Descriptive statistics
3. Two continuous variables
   1. Descriptive statistics
   2. Inferential statistics
4. Two categorical variables
   1. Descriptive statistics
   2. Inferential statistics
5. One continuous and one categorical
   1. Descriptive statistics
   2. Inferential statistics

Part II: Use the library to practice

**The take-home part of the Final Exam will look a lot like this:**

You receive a data set saved in SPSS format from the University of Kentucky in Lexington; it is saved online in an SPSS file at <https://tinyurl.com/ydyns6e6>. The data is from a survey of UKY students who smoke and what they think of a new smokefree campus policy. The codebook that came with the data says the questions asked were as follows:

* age--What is your age in years?
* age\_initiation--How old were you when you smoked your first cigarette?
* gender--What is your gender?
* race--What is your race?
* risky\_behavior--participates in risky activities (higher = riskier)
* quit--If you smoke, do you plan to quit?
* parent--Do you have a parent who smokes?
* sibling\_smoke--Do you have a sibling who smokes?
* friends\_smoke--Among your friends, what percentage do you think smoke? type a percentage between 1 and 100 (do not include percent sign)
* familiar\_policy--How familiar are you with the rules and / regulations regarding the upcoming tobacco-free policy at the University of Kentucky?
* policy\_success--the policy will be successful with reducing secondhand smoke exposure and encouraging people to quit?
* motivation--thinks the campus policy will increase my motivation to quit?

You want to understand two things: characteristics associated with risky\_behavior and characteristics associated with quitting smoking:

1. Replace blank and other clearly missing data points with NA
2. Identify and compute the appropriate descriptive statistics for the variables in the data set. Make a table or write a paragraph using the descriptive statistics to describe the sample.
3. Conduct the appropriate bivariate statistical tests for the relationship between quit and each of the other variables (try at least one categorical and one continuous)
4. Conduct the appropriate bivariate statistical tests for the relationship between risky\_behavior and demographics you are interested in (try at least one categorical and one continuous)
5. Interpret bivariate results
6. Build a simple linear regression model using friends\_smoke to explain risky\_behavior
7. Interpret the model coefficients and their significance, overall model significance, and model fit
8. Build a logistic regression model including any of the variables significantly associated with quit
9. Compute and interpret the OR/CI, model fit, model significance for the model in #8

Part III: Examples of multiple choice questions

**The in-class part of the Final Exam will look a lot like this:**

1. Which of the following best describes an independent samples t-test?
   1. testing the difference between means for two related groups
   2. testing the difference between means for two unrelated groups
   3. testing the difference between means for three or more groups
   4. comparing a claimed or hypothesized mean to a sample mean

1. The equation that describes a line is written y = mx + b. When written this way, which part of the equation represents the slope of the line?
   1. m
   2. y
   3. b
   4. x