

R Workshop Exercises

1. Load and place your data into a new folder on the Desktop called “Temp” or download mainData.csv and sample_code.R from https://github.com/korykantenga/Stata_R_Workshop and place the data into a new folder on the Desktop called “Temp”.
2. Open R Studio and create a new .R Script called “mainProgram.R” and save it in the folder with your data.
3. Import your data into R and name it “myData”. Write in and run your commands from the R Script.

```
setwd("Your_Path_Name_Goes_Here\Desktop\Temp")  
myData <- read.csv("mainData.csv")
```

4. What the research question you want to answer with the data?

5. What is the outcome variable you are interested in? What is it called in the data? What are its units?

Outcome Variable:

Variable Name:

Units:

6. Summarize (if continuous) or tabulate (if discrete) your outcome variable. Are there any missing values? Use the following to summarize and eliminate missing values in your variable.

```
summary(na.omit(myData$myVariableName))  
table(na.omit(myData$myVariableName))
```

7. What do you notice about your variable? What does it tell you? If the outcome is a continuous variable (not just 0 or 1), create a histogram and save it as myHistogram.pdf.

Some statistics of usual interest:

- Mean (It is high? low? What does a high mean say?)
- Variance/Standard Deviation (Is the variable highly dispersed or concentrated?)

8. What is your hypothesis(es) about the outcome variable? Alternatively, what do you think are predictors of the outcome variable?

9. Use a regression to test whether your predictors are important or if your hypotheses about what affects the outcome are true.

```
lm.1 <- lm(myOutcome ~ predictor1 + predictor2 + predictor3, data=myData)
summary(lm.1)
```

10. What do the p-values say about the significance of your predictor or whether there is evidence to support your hypothesis? In other words, which variables are statistically significant and at what level? 10%, 5%, 1%?

11. What can you conclude from the sign (positive vs negative) and significance (significant versus insignificant) of your regression variables and their effect on the outcome?

12. Are you missing or did you forget any important variables? Omitted variables correlated with both the outcome and the predictor variable will bias your coefficients!