Homework 3

Overview:

In this assignment, you'll be using R/RStudio and the GSS data set (2014) to report bivariate statistics for two dependent variables of interest. These variables must be different than the ones I present in the example below (unless you've discussed it with me). Remember, I want you to select variables that interest you, that way you could use these analyses for your final paper. Finally, the only bivariate relationship you will report for this analysis is either a **chi-square** test, a **t-test**, an **analysis of variance (ANOVA)**, or **correlation** between two variables.

What you will need:

The following can be downloaded from our course website

- **hw3_script.R**

- **GSS2014_final.csv** aka GSS Data Set (2014)

- **LABELS_script.R**

- **GSS Variables and Descriptions.pdf**

- **GSS_Codebook.pdf**

Tasks:

- Open up the Homework 3 script/code file (**hw4_script.R**). This is the file you'll be working from.

- Set your working directory

- Load the required packages/libraries

- Load the data set (**GSS2014_final.csv**).

- Load the cleaner/labels file (**LABELS_script.R**).

- Select two (2) variables that interest you. This can be 1) two categorical variables, 2) a categorical variable with two categories (e.g. sex) and an interval/continuous variable, 3) a categorical variable with more than two categories (e.g. race or religion) and an interval/continuous variable, or 4) two interval/continuous variables. To do so, you will need to look through the list of variables (**GSS Variables and Descriptions.pdf**). Once you've selected two variables, locate them in the codebook (**GSS_Codebook.pdf**) to figure out how those variables are coded.

What you will turn in:

- A brief paragraph describing:

  - A sentence or two about the data set you're using for these variables.
  - For each variable, a sentence or two describing the variable name and its description.

- A bivariate relationship between the variables, and a description/interpretation of the relationship. Remember the following, use a:

  - **Chi-Square** if you have two categorical/discrete variables
  - **T-test** if you have a categorical variable with two categories (e.g. sex) and an interval/continuous variable
  - **ANOVA** if you have a categorical variable with more than two categories (e.g. race or religion) and an interval/continuous variable
  - **Correlation** if you have two interval/continuous variables

- The text of your R code/script on a separate page.

Example Homework 3

1: Data Set Description

The data for this assignment come from the General Social Survey administered in 2014. The data set has 2,538 observations, with individuals as the unit of analysis.

2: Variable Descriptions

The two outcome variables I selected were **POLVIEWS** and **OPPSEGOV**. The variable **POLVIEWS** is a continuous measure of conservatism that ranges from 1 to 7, with higher scores on this variable indicating higher levels of conservatism. The text of the variable is as follows:

> "We hear a lot of talk these days about liberals and conservatives. I'm going to show you a seven-point scale on which the political views that people might hold are arranged from extremely liberal–point 1–to extremely conservative– point 7. Where would you place yourself on this scale?"

The variable **OPPSEGOV** is a continuous measure on the importance of civil disobedience that ranges from 1 to 7, with higher scores indicating that the respondent believes civil disobedience is more important. The text of the variable is as follows:

> "That citizens may engage in acts of civil disobedience when they oppose government actions."

3: Bivariate Description

Note: Based on your tests, report one of the following:

Chi Square: *if you have two discrete/categorical variables*

The chi square test between **RACDIF4** and **RACE** shows that there is a significant relationship between racial categories and beliefs about the causes of racial inequality: $X^2 = $ **10.902**, $p < $ **.01**. This indicates that racial groups differ on their belief about whether or not a "lack of motivation" is what leads to Black inequality. Those who identify as Other are more likely to believe that Blacks just don't have the motivation to pull themselves out of poverty.

T-test: *if you have one categorical variable with two categories (e.g. sex) and one interval/continuous variable*

The t-test between **POLVIEWS** and **SPKATH** shows that there is a non-significant mean difference in political views by whether or not people believe that Atheists should be allowed to speak: $t = $ **-2.8515**, $p < $ **.01**. This indicates that those who believe that Atheists should be allowed to speak in their community differ on their level of conservatism. Those who don't think Atheists should be allowed to speak ($M = $ 4.268882, $SD = $ 1.336061) have higher mean/average conservatism (higher values on **POLVIEWS**) than those who think Atheists have the right to speak.

ANOVA: *if you have one categorical variable with more than two categories (e.g. religion) and one interval/continuous variable*

The analysis of variance between **POLVIEWS** and **RACE** shows that there is a significant mean difference in political views for different races: $F = $ **18.46**, $p < $ **.001**. This indicates that different racial categories are more conservative than others. White people ($M = $ 4.170572, $SD = $ 1.458553) have higher

mean/average conservatism (higher values on **POLVIEWS**) than those who identify as Black or Other.

Correlation: *if you have two interval/continuous variables*

The correlation between **POLVIEWS** and **OPPSEGOV** is weak, negative, and significant: $r = $ **-.1659262,** $p <$ **.001**. This indicates that as one variable increases, the other decreases. People with higher levels of conservatism (i.e. higher values of **POLVIEWS**) believe that civil disobedience against the government is not important (i.e. lower levels of **OPPSEGOV**).

4: R Code/Script for Homework 3

```
setwd("/Users/burrelvannjr/Dropbox/methods_csuf")

library(psych)
library(Hmisc)
library(RCurl)

DATA1<-read.csv("resources/data/GSS2014_cleaned_nm.csv",header=TRUE,sep=",")
source("resources/data/LABELS_script.R")

DATA1$polviews
DATA1$oppsegov

cor.test(DATA1$polviews, DATA1$oppsegov)
```