Homework #2

Homework Assignment #2

Directions:

- This assignment will ask you to do write out formulas and complete calculations by "hand." Do not use R to calculate these. Use a calculator or Excel to conduct these calculations. Show your work in any of the following ways 1) turn in the excel file; or 2) "type" out the calculations in this word document; or 3) submit a scanned copy of your handwritten work.
  - Whichever format you select just be sure to reference the question number for each of your answers
- If the questions below ask you to execute R commands, then copy all R syntax (indicated via `this is R code`) into the R script (but you must delete the ‘ at the front and end of the syntax within the R script in order for it to run)
  - The R script should have a "comment line" indicating what number question the R syntax refers to for this assignment
- *Submit your answers file along with your R script to the D2L Dropbox*

Before you begin:

Download the CA Data [If you did not complete this during class]:

- Create a new data folder called "ca"
  - hed612 >>> data >>> ca
- Download the California Dataset from D2L (under Datasets)
  - Place the "caschool-v2" dataset into the "ca" folder you created in the previous step

Create a new R Script for this homework assignment

- Open the RProject you created last week (should be in your main hed612 folder)
- Once the RStudio window opens, within the R project session, open a new R Script
  - files >>> New File... >>> R Script
- Save the file as `HW2_lastname.R` within lecture2 subfolder

About the data The CA Schools dataset contains data on test performance, school characteristics and student demographic backgrounds. The data used here are from all 420 K-6 and K-8 districts in California with data available for 1998 and 1999. Each observation indicates one school district.

Questions related to caschools-v2.dta

1. Within the R script created above: load the haven and tidyverse libraries; check that your directory is set to the R project; open the California Schools data
2. Copy the following syntax to your R script and "run" it to subset the California Schools data to the first 5 observations in the dataset

```
caschool <- caschool[1:5,]
```

3. Write (or "type") the formula for calculating sample mean AND calculate the sample mean for the 5 observations for the variable "computer"
   - You can view the 5 observations in R via the global environment OR copy, paste, run the following in your R script:

     ```
     caschool['computer']
     ```

4. The variable computers is a measure of total number of computers each district owned. In your words, interpret the sample mean you calculated in question #5.

5. Write (or "type") the formula for calculating sample standard deviation AND in words describe what the standard deviation measures.
6. Calculate the sample standard deviation for the 5 observations of the variable "computer"
    - You can view the 5 observations in R via the global environment OR copy, paste, run the following in your R script:

      ```
      caschool['computer']
      ```
7. Check your calculated work by copying, pasting, and running the following commands in your R script.

```
caschool %>% summarise(mean(computer, na.rm=TRUE))

caschool %>% summarise(sd(computer, na.rm=TRUE))
```