**Assignment 2**

(Total Points: 100)

Answer each question using the attached data and partially completed syntax. This assignment is to be completed by each student individually. Copy and paste your completed syntax into this assignment.

1. Define the following key terms in your own words (5 points each)
   1. Sampling Error
   2. Validity
   3. Reliability
   4. Degrees of Freedom
   5. Sum of Squared Error
2. Provide a summary of the 2019 Small Area Income and Poverty Estimate (SAIPE) Data, include a summary of how it’s calculated, the data sources, and the utility of this data for public health and health equity in less than 250 words (25 points)
3. Calculate the estimates below based on the Modified 2019 SAIPE Data (5 points each)
   1. The number of counties in the United States with a median household income of $75,000 or greater per year
   2. The mean household income for counties in California
   3. The mean household income for states including the District of Columbia
   4. The standard deviation for the median household income for counties in California
   5. The standard deviation for the median household income for states including the District of Columbia
4. Provide a histogram for median household income among Georgia counties, with publication quality titles for the axis and figure overall (25 points)

Partially Completed Syntax

#import the modified 2019 SAIPE Data Provide For this Assignment (Modify for location data is saved on your computer)

library(readxl)

SAIPE <- read\_excel("Desktop/Assignment\_2\_Modified\_2019\_SAIPE\_Data.xlsx",

col\_types = c("numeric", "numeric", "text",

"text", "numeric", "numeric"))

#Create a subset of data for counties in California

CaliCo <- SAIPE[ which(SAIPE$PostalCode=='CA'& SAIPE$CountyFIPS != 0), ]

#Creata a subset of data for counties in Georgia

GACo <- SAIPE[ which(SAIPE$PostalCode=='GA'& SAIPE$CountyFIPS != 0), ]

#Create a subset of data for just the state-level data (and DC)\*\*\*NEED TO COMPLETE

#Create a subset of counties with a median household income of $75k or greater\*\*\*NEED TO COMPLETE

#calculate means for household income\*\*NEED TO COMPLETE

mean(States$`Median Household Income`)

#calculate tandard deviations for household income\*\*\*NEED TO COMPLETE

sd(CaliCo$`Median Household Income`)

#calculate how many counties in the United States have a household income greater than $75,000 Per Year\*\*\*NEED TO COMPLETE

#create a histogram of Poverty All Ages Among Counties in California for 2019\*\*\*NEED TO COMPLETE