**Homework 2**

**Instructions:**

Place all work into a **typed one-to-two page** document. Upload **both the document and the CleanNLS.csv dataset created in Problem 2e** onto the Canvas homework page by the due date. Documents must be in one of the following formats: `.doc`, `.docx`, or `.pdf`. Data for completing these problems may be found under the "Files/Homework/Homework 2" page on Canvas.

The file NLS2023.csv contains data on a set of individuals. For each individual, the following variables are measured: Urban (1 if the individual lives in an urban area, 0 otherwise), Siblings (number of siblings), White (1 if the individual is white, 0 otherwise), Christian (1 if the individual is Christian, 0 otherwise), FamilySize, Height (in inches), Weight (in pounds), and Income.

1. Data Inspection:

   (a) **(1pt)** How many observations are there in the dataset? How many variables are there in the dataset?

   (b) **(1.5pt)** For each variable, how many missing entries are there? Also, give the line or lines of R code used to find this answer.

   (c) **(0.5pt)** Using the `sort()` function, find the largest number of siblings for an individual in the dataset. Also give the line or lines of R code to find this answer.

   (d) **(1pt)** Using the `order()` function, order the data according to the FamilySize variable from largest to smallest. What is the largest family size in the dataset? Also, give the line or lines of R code used to find this answer.

2. Data Cleaning:

   (a) **(1pt)** Remove all individuals from the dataset with 0 income. Give the line or lines of R code for this step.

   (b) **(1pt)** Afterward, replace all missing incomes with the average of the non-missing incomes. Give the line or lines of R code for this step.

   (c) **(1pt)** Apply the log transformation (either base $e$ or base 10) to the incomes and add this as a column to the NLS2023 data frame. Give the line or lines of R code to perform these steps.

   (d) **(1pt)** Apply the square root transformation to heights and add this as a column to the NLS data frame. Give the line or lines of R code to perform these steps.

   (e) **(2pt)** After performing steps (a), (b), (c), and (d), export this cleaned dataset as a .csv file called CleanNLS.csv. Give the line or lines of R code to perform this step. Also, upload this file onto Canvas.