Stat 123 Homework Assignment 1
Due Friday May 26, 2023 by 8:00pm

Using R Markdown to complete the following assignment. If an answer does not require any R code, you can type the answer to the question outside of a chunk. Make sure that your assignment is well labelled so that it is clear where each question's answer begins. Your assignment should be submitted as a pdf (whether you knit directly to PDF, or knit to HTML or Word and then convert the file to a pdf).

1. A farmer wants to determine the proportion of pumpkin seeds planted in her field that successfully grow into pumpkins. It would take too much time to count the total number of seeds planted in the field and the total yield of pumpkins. Thus, she decides that she needs to take a sample to estimate this proportion.

   (a) State the population and the variable of interest to this farmer.
   (b) Give an example of a way the farmer could perform a convenience sample.
   (c) Give an example of a way the farmer could perform a simple random sample.
   (d) Give an example of a way the farmer could perform a stratified random sample.
   (e) What is the population parameter of interest? What would be a good statistic to use to estimate this?

2. The following question deals with the data set *lynx* which is one of the built-in data sets included with R.

   (a) Describe what information is contained in the data set. What command in R did you use to find this information?
   (b) Create a character vector called *years* which contains the years of the trappings.
   (c) Set the names of the *lynx* vector equal to *years*.
   (d) How many lynx were trapped in 1901? Use *years* as your index.
   (e) What is the average number of annual lynx trappings from 1821 to 1920, inclusive? Hint: You want to compute the average trappings for the first 100 data points.

3. The following question deals with the data set *casino* which can be found in Brightspace by clicking on Content $->$ Homework Assignments. This data set represents the winnings and losses of a group of friends who went to a local casino together.

   (a) Read in the data file and name the data frame *casino*.

   (b) Use the head() function to determine the games these friends played in the casino.

   (c) Create a character vector called $friends$ which contains the values from the first column of the data set.

   (d) Using the R command as.matrix(), create a matrix called *winnings* which contains all the columns except the first one from the *casino* data set.

   (e) Create a vector called *totals* which contains the row sums of the matrix *winnings*. What do the values in this vector represent?

   (f) Set the names of the vector *totals* equal to $friends$.

   (g) Use the R functions min(), max(), which.max() and which.min() to determine which friend won the most money and which friend lost the most money in the casino.

   (h) What was the average amount of money won or lost by the group of friends on the trip?