Stat 123 Homework Assignment 3

Due Wednesday March 15th by 8:00pm

Using R Markdown, please 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. Use the following commands:

   set.seed(123)

   data <- rnorm(10000, mean = 50, sd = 10)

   The commands generate a normal distribution with a mean of 50 and a standard deviation of 10.

   Create a histogram to show the range of values that covers the middle 95% of this distribution, using the title of "Normal Distribution with Mean 50 and SD 10", light blue color, and breaks equal to 30.

2. Download the data sets sampledata.csv with 100 numeric variables, name it *mydata* and save it to whatever directory you are using for this question.

   (a) Use a for loop to calculate the mean and standard deviation for each variable in *mydata*.

   (b) Use an if statement to identify the variables with a mean greater than 10.5 and store their names in a vector.

3. Use the dataset *mydata* in question 2.

   (a) Calculate the mean for each variable.

   (b) Calculate the 90% confidence interval for each variable

   (c) Print the results using a matrix with three columns named (*"Mean", "Lower_CI",* and *"Upper_CI"*).

   Hint: You need to use *mydata* and create a matrix with three columns. The number of rows equals the number of rows in *mydata*. Next, you need to apply for loops to calculate the mean and the 90% confidence interval. Finally, fill the matrix with the results.

4. Consider the built-in data set UCBAdmissions in R.

(a) If we are interested in the proportion of people that apply to Berkeley University and get accepted, what is the population of interest and what is the parameter of interest?

(b) Using the command ? UCBAdmissions, determine the variables in the dataset and describe what kind of variables they are.

(c) Create a variable in R called totalAdmissions which contains the total number of students who were admitted to the university (across all genders and departments).

(d) Create a variable in R called totalRejections which contains the total number of students who were rejected to the university (across all genders and departments).

(e) Create a variable in R called totalApplicants which contains the total number of students who applied to the university in our sample.

(f) What is the observed value of the statistic we should use to estimate the population parameter of interest?

(g) What is the estimated standard error for ^p?

(h) What is the critical value for a 92% con_dence interval for p?

(i) What is the margin of error for our estimate?

(j) Compare that to result of the approximate margin of error formula we learned earlier in the course.

(k) Determine a 92% con_dence interval for the true value of the population proportion.