

STAT 158 Take home mid-term exam Due Friday March 15th by 2pm

Instructions: for each question, copy and paste any relevant R code and output into a document. The code and output for each question should go together. Make sure you answer any direct questions that are asked.

Do not obtain assistance from anyone else for this exam. Don't work with classmates or talk to tutors. If you need clarification on a question, you can email me at ffn@colostate.edu.

Please turn this exam in *on canvas* - do not email it to me.

Exam questions:

1. During a previous semester of introductory Statistics, I asked students to write down their height, whether they were male or female, their favorite color, and asked them to "pick a number, any number". The results are in the data set `Stat_class_data.csv` on Canvas.
 - a. Read the data into R. State the names of the columns in this dataset, and report what "class" R assigned to each one.
 - b. How many females are in this data set? How many males?
 - c. Get R to tell you all the values that favorite color takes on. What are the three most popular favorite colors?
 - d. Have R give you summary statistics for height. What is the value for median height?
 - e. Make a plot of "number". Describe what this plot is showing you.
 - f. Get R to calculate how many values for "number" are greater than 5000. What proportion of students is this?
 - g. Rows 77 and 127 contain extreme outliers for the "number" variable. Use R to create a new variable set that excludes these rows but contains the values of "number" for all the other rows. Report the mean of number with and without these two extreme outliers included.
2. Create a vector called "x", containing a sequence of numbers that go from -3 to 3, by increments of 0.1. Then create another vector that computes:

$$y = \frac{1}{\sqrt{2\pi}} \cdot e^{-\left(\frac{x}{2}\right)^2}$$

This is the formula for the normal distribution. Make a plot of x vs. y to verify that you did the calculation correctly. Note: to use π in an R formula, just type "pi" (without the quotation marks)

3. For this question, you'll randomly generate data and then plot it.
 - a. Create a vector containing 1000 values randomly sampled from the integers 1 through 100. Make a histogram of this vector.
 - b. Create another vector containing 1000 values randomly sampled from a normal distribution with mean=10 and standard deviation=2. Make a histogram of this vector.
 - c. Create another vector that combines the values generated in parts a. and b. together into a single vector. Make a histogram of this vector.
 - d. Create another vector that adds the values of the vectors from parts a. and b. to each other. Make a histogram of this vector.
 - e. Do the histograms in c. and d. look about the same, or do they look different? Briefly explain why you think they look about the same or why they look different.