Problem Set 0: R Part

Your Name Goes Here

Details

Due Date

This assignment is due at 1:00 PM on Wednesday, Sept 19.

How to Write Up

The R part of this assignment must be completed as an R Markdown file and knit to pdf. Both the Rmd file and the pdf must be committed and pushed to GitHub.

Grading

5% of your grade on this assignment is for turning in something legible. This means it should be organized, and any Rmd files should knit to pdf without issue.

An additional 20% of your grade is for completion. A quick pass will be made to ensure that you've made a reasonable attempt at all problems.

In the range of 1 to 3 problems will be graded more carefully for correctness. In grading these problems, an emphasis will be placed on full explanations of your thought process. You don't need to write more than a few sentences for any given problem, but you should write complete sentences! Understanding and explaining the reasons behind what you are doing is at least as important as solving the problems correctly.

Solutions to all problems will be provided.

Collaboration

You are allowed to work with others on this assignment, but you must complete and submit your own write up. You should not copy large blocks of code or written text from another student.

Sources

You may refer to our text, Wikipedia, and other online sources. All sources you refer to must be cited in the space I have provided at the end of this problem set. I'd be shocked if you didn't cite our textbook.

Problem 1: Data Camp Courses

Complete the assigned chapters on DataCamp. If you have already done this for a previous course, you don't have to do it again – but it would be good to take a quick look over it and review any parts you don't remember.

Problem 2: Beta distribution

Consider the beta distribution (I don't necessarily expect you to have seen this distribution before - look it up on Wikipedia or in our text if you are unfamiliar with it; it is also described in the "Probability Distributions" handout).

(a) Symmetric Beta Distribution

i. Pick values for the parameters a and b so that the distribution is symmetric.

ii. Write down the pdf for the distribution with parameter values you selected in part i. Please do this using latex. Example latex code for a beta distribution is at the end of the probability_distributions.Rmd file linked to on the resources page of the course website.

iii. What are the theoretical mean, variance, and standard deviation of the beta distribution with parameters you chose in part i? No need to derive these, you can use standard properties of the beta distribution as given in the probability distributions handout.

iv. Using the R function `rbeta`, generate 100 random beta variates with the parameter values you chose in part i. Store them as a variable called `x` in a data frame called `beta_samples_a`.

```
beta_samples_a <- data.frame(
  x = 0 # replace the 0 to the left with a call to rbeta
)
```

v. Find the sample mean, sample variance, and sample standard deviation of your 100 random deviates. If you're not sure how to do this, check out the summarize function in the dplyr package.

vi. Create a density histogram of your 100 random deviates. On the same plot, add a plot of the pdf from part ii.

(b) Asymmetric Beta Distribution

i. Pick values for the parameters a and b so that the distribution is asymmetric.

ii. Write down the pdf for the distribution with parameter values you selected in part i. Please do this using latex. Example latex code for a beta distribution is at the end of the probability_distributions.Rmd file linked to on the resources page of the course website.

iii. What are the theoretical mean, variance, and standard deviation of the beta distribution with parameters you chose in part i? No need to derive these, you can use standard properties of the beta distribution as given in the probability distributions handout.

iv. Using the R function `rbeta`, generate 100 random beta variates with the parameter values you chose in part i. Store them as a variable called `x` in a data frame called `beta_samples_b`.

```
beta_samples_b <- data.frame(
  x = 0 # replace the 0 to the left with a call to rbeta
)
```

v. Find the sample mean, sample variance, and sample standard deviation of your 100 random deviates.

vi. Create a density histogram of your 100 random deviates. On the same plot, add a plot of the pdf from part ii.