

CEPC 0904
Professor Jack Hester
Due July 19, 2021 at 11:59pm EDT

HW 4 — Basics of statistics using R

34 points

Submission link: <https://canvas.brown.edu/courses/1085057/assignments/7834144>

Submit as a .html file

Homework 4 will step you through examples of how you can use R to perform exploratory data analysis and statistical tests. **Please read all of these directions carefully as there are a lot of moving parts.**

For HW4, you will be using an R Markdown file. R Markdown files allow you to combine text and formatting with r code and then create nice looking documents (html, pdf, or word) when you're finished.

To install R and use R Studio you will need to download and install two separate packages. First, download R (I recommend version 4.1.0) from their website by going to <https://www.r-project.org/> and navigating to the download portion (via CRAN or by clicking on the version number).

Second, you will need to install R Studio. Navigate to <https://www.rstudio.com/products/rstudio/download/> and select the free download option.

I recommend you install R before R studio. Once you have downloaded and installed both, you should be able to open the .rmd homework file.

All of the questions and places to fill in answers/code are provided in the HW4.rmd file, which you can open in RStudio. You will also need to download distribution.csv and test_scores.csv and edit the file path in one of the questions to point to that file. When you're all done, make sure you knit the file (there's a button in the toolbar at the top with a yarn icon). You will be submitting the **knitted .html file**, not the .rmd file. Please update the "name" section of the header to include your name. Both the .rmd and .csv files are located on the homework page of the course website.

You will also need to make sure you run the "r setup" block at the top before beginning. **not doing so will cause errors and you will get some answers wrong because you didn't use the correct seed.** Please also read all of the .rmd document carefully and make sure you answer every question, especially since some questions have multiple parts and sub-parts. Code blocks can be run in R studio by clicking the green arrow button next to them or in the top menu bar. The knit button can be found in the menu bar as well. The file will be created in the same directory (folder) as your original .Rmd file.

For this assignment, you are encouraged to work with your group that you were assigned to in the Slack chat, but please be sure to upload your own file. You will submit the

assignment to canvas as the .html knitted file I mentioned earlier. You do not need to include the csv files or the original .rmd file.

If you have not already, you should complete the Introduction to R Datacamp module and watch the “exploratory data analysis” lecture on the course website. It is probably also helpful to watch the other week 3 statistics videos and read the stats lecture notes which can be found on the schedule page or the lecture page.