CEPC 0904
Professor Jack Hester
Due July 18, 2022 at 11:59pm EDT

HW 4 — Basics of statistics using R

34 points
Submission link: `https://canvas.brown.edu/courses/1088637/assignments/7889606`
Submit a knitted file in .html or .pdf format only. Do NOT submit your raw .rmd 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. FAQ about how to install R/R Studio or run the code are at the end of this document.

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.

FAQ About Installing and Using R/R Studio

The .rmd file says something about editing a header, but I can't find what you're talking about?

If you have trouble viewing the header portion, (see note in the rmd file), you can access and download the raw rmd file at `https://raw.githubusercontent.com/dotSlashJack/CEPC0904-22/master/documents/homework/HW4.rmd` or via the file shared in the slack channel and this should resolve your issue.

The header should be at the very top of the file and look like:

```
---
title: "HW4: Hypothesis Testing and Stats"
author: "[your name]"
date: "Due 7/16/2022"
output: html_document
---
```

I can't figure out how to install R on Windows. Which version do I use?

The installation file for windows can be found at `https://cran.r-project.org/bin/windows/base/`. **I recommend trying the 64 bit installation as long as you have a relatively new computer**. Click "Download R 4.2.1 for Windows" at the top.

I can't figure out how to install R on Mac, all I see are a bunch of files.

The installation file for Mac OS can be found at `https://cran.r-project.org/bin/macosx/`. Click on "R-4.2.1.pkg" near the top left.

After trying to install R, my Mac said the version isn't compatible with my Mac OS version.

Visit `https://cran.r-project.org/bin/macosx/` and scroll down until you find a version that has a MacOS version number in the paragraph describing the installer that is at or below your Mac's operating system version. Click that version number on the left to download it. You can find your Mac OS information by clicking the apple icon at the top left of your screen and then selecting "about this mac." You'll see something like "11.2.3" as the version number.

I'm getting a "package load failed" error when I run the setup code.

That means you probably need to install the packages first. You can do so by removing the # at the beginning of the install.packages lines in the setup code block. Look for "#install.packages('ggplot2')" and "#install.packages('reshape2')" (should be lines 9 and 11) and change them to "install.packages('ggplot2')" and "install.packages('reshape2')". The text color should change to black and blue from green. Then re-run the code block. You should see installation messages show up in the console at the bottom of R studio. If you want to, you can add the # back in front of those lines once installation has completed to speed things up.

I'm getting a "Rtools' error along with an "is not writeable" warning.

This usually has to do with some sort of windows permission error. If it happens when you try running install.packages, you can search for R and right click ¿ run as administrator, then type the install.packages() lines (one at a time) into R and press enter to run them. If the issue is not caused by installing packages, you can try downloading and installing Rtools directrly from `https://cran.r-project.org/bin/windows/Rtools/`. You can also try running R studio as an administrator.