

MATH 345/545: Homework Ch. 1

UMass Boston – Summer 2021

Please show your work in all exercises: equations, R code, and graphs.

One way to combine all these things is to type your homework in R Markdown, using RStudio. There are several examples of R Markdown (aka .Rmd) files in **Blackboard > Group HW > Homework Files and Examples (Folder)** which also includes the pdf's produced by "Knitting" each .Rmd file (e.g., by using the Knit command button in RStudio). The R Markdown cheat sheet is a useful reference (available in **Blackboard > Course Materials > R Resources (Folder)**). Alternatively, you can use Word, LaTeX or other text editors to combine text, figures, and code and save the result in a single pdf document. If you want to combine that with handwritten material, you may need to use a pdf editor (*Preview* is an option on the Mac – *PDF-Xchange* is a free Windows app that produces a watermark on the output, which is not a problem). Whatever works!

Data files for exercises are posted in **Blackboard > Group HW > Homework Files and Examples (Folder)**. To load a dataset into R using RStudio, first download the file from Blackboard and then use the RStudio *Files* tab in the lower right panel: initially this will list the files in your starting ("Home") folder (such as your *Documents* on Windows or */Users/<your_user_id>* on MacOS); from there you can navigate to wherever you downloaded the file, then click on the filename and choose the option to *Import Dataset*. (If you need to install readr that's fine.)

Exercise 1

Textbook exercise 1.22 (sum of deviations from sample mean is zero).

1.22. Prove that the sum of the deviations of a set of measurements about their mean is equal to zero; that is,

$$\sum_{i=1}^n (y_i - \bar{y}) = 0.$$

Exercise 2

Textbook exercise 1.23 (television commercials). Note: check out the `pnorm` function in R.

1.23. The mean duration of television commercials is 75 seconds with standard deviation 20 seconds. Assume that the durations are approximately normally distributed to answer the following.

- What percentage of commercials last longer than 95 seconds?
- What percentage of the commercials last between 35 and 115 seconds?
- Would you expect commercial to last longer than 2 minutes? Why or why not?

Exercise 3

Steam is a software platform used to distribute and manage multiplayer online games. A day was selected at random and the peak number of simultaneous users for 18 games is given in the following table (this data is also available in the `players.csv` file):

930	858	827	849	744	849	763	753	781
948	678	742	769	754	782	862	894	861

- Find the sample mean and the sample median.
- Suppose the last observation had been 2861 instead of 861. Find the sample mean and sample median for this revised data set. Explain how this change in the data affects the mean and median found in part (a).

Exercise 4

A typical road bridge is constructed to last 50 years. The mean age of all bridges in the United States is approximately 43 years. The number of structurally deficient bridges in each state and the District of Columbia as of 2009 is given in the dataset `bridges.csv`.

- How many data points are there in this dataset?
- Plot the histogram of the data (and include it in your homework submission). Comment on its shape (is it flat? Unimodal? Multimodal?) and on its skewness (left skewed? Right skewed? Symmetric?). If you find skewness, provide a plausible explanation for it.
- Display a box plot of the data (also include it in your submission). Does it show potential outliers?
- Report the five-number summary of the data. How does the mean compare to the median?
- Find the sample variance s^2 and the sample standard deviation s of the number of structurally deficient bridges.
- Suppose each state repaired 10% of all structurally deficient bridges. Find the sample variance and the sample standard deviation for the resulting dataset. *Note: if for a state, the new number of structurally deficient bridges is not integer, this is not a problem.*
- How do your answers to parts (e) and (f) compare?