

## Assignment 10

Instructions: Make an R Notebook and in it answer the question or questions below. When you are done, hand in on Quercus the *output* from Previewing (or Knitting) your Notebook, probably an `html` or `pdf` file. An `html` file is easier for the grader to deal with. Do *not* hand in the Notebook itself. You want to show that you can (i) write code that will answer the questions, (ii) run that code and get some sensible output, (iii) write some words that show you know what is going on and that reflect your conclusions about the data. Your goal is to convince the grader that you *understand* what you are doing: not only doing the right thing, but making it clear that you know *why* it's the right thing.

Do *not* expect to get help on this assignment. The purpose of the assignments is for you to see how much *you* have understood. You will find that you also learn something from grappling with the assignments. The time to get help is after you watch the lectures and work through the problems from PASIAS, via tutorial and the discussion board, that is *before* you start work on the assignment. The only reasons to contact the instructor while working on an assignment are to report (i) something missing like a data file that cannot possibly be read, (ii) something *beyond your control* that makes it impossible to finish the assignment in time after you have started it.

There is a time limit on this assignment (you will see Quercus counting down the time remaining).

1. A student keeps track of what time they go to bed and what time they get up in the morning. They also have an app on their phone that measures the number of hours they were asleep during that time. The data for one week are in <http://ritsokiguess.site/STAC33/sleeping.csv>, in the 24-hour clock.
  - (a) Read in and display the data. What type of things are each of your columns?
  - (b) Work out the fractional number of hours that the student was in bed each of these nights. (They may not have been asleep this whole time.) Your result needs to be a *number* since we will be doing some calculations with it shortly.
  - (c) The student is concerned with something they call “sleep efficiency”. This is the percentage of time in bed spent sleeping. Work out the student’s sleep efficiency for the seven nights in this dataframe. Which night was the student’s sleep efficiency greatest?
  - (d) Display the time spent in bed each night as a number of hours, minutes and seconds.
  - (e) Make a graph of what time the student went to bed each night. Bear in mind that you only need the times, not the dates, and that you want a graph that is informative, showing appropriately the distribution of times the student went to bed.

There is a second question on the next page.

2. Earlier, we investigated some data on predicting the height of a person from the length of their foot. The data were in <http://ritsokiguess.site/STAC32/heightfoot.csv>.

- (a) Read in and display (some of) the data.
- (b) In STAC67, you learn (or will learn) the matrix formulation of the least squares estimates of intercept and slope. This produces a vector  $\hat{\beta}$  containing estimates of both the intercept and the slope, from the formula

$$\hat{\beta} = (X^T X)^{-1} X^T y,$$

where:

- $X$  is a matrix containing a column of 1s followed by all the columns of explanatory variables
- $X^T$  denotes the (matrix) transpose of  $X$
- $M^{-1}$  denotes the inverse of the matrix  $M$
- $y$  denotes the column of response variable values.

Use the formula above to obtain the least squares estimates of intercepts and slope for this regression, using R's vector-matrix algebra. Hint: you are advised to do the calculation in steps, or else it will be very hard to read, and hard for the grader to check that it is correct.

- (c) Verify that your calculation is correct by running the regression.