STAD29 / STA 1007 assignment 2

Due Tuesday Jan 29 at 11:59pm on Blackboard

Hand in the indicated questions. In preparation for the questions you hand in, it is worth your while to work through (or at least read through) the other questions as well.

Hand in your work on Quercus. If you did STAC32 last fall, it's the same procedure. A reminder is here: `https://www.utsc.utoronto.ca/~butler/c32/quercus1.nb.html`

You are reminded that work handed in with your name on it must be *entirely your own work*. It is as if you have signed your name under it. If it was done wholly or partly by someone else, *you have committed an academic offence*, and you can expect to be asked to explain yourself. The same applies if you allow someone else to copy your work. The grader will be watching out for assignments that look suspiciously similar to each other (or to my solutions). Besides which, if you do not do your own assignments, you *will* do badly on the exams, because the struggle to figure things out for yourself is an important part of the learning process.

You will probably need:

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Hand in question 2 below.

1. Work through, or at least read, chapter 19 of PASIAS.

2. Twenty students are all studying for a certain exam. Some of them spend more time studying, and some of them spend less. Does the amount of time spent studying affect whether a student will pass the exam? The data are in `study_time.txt`.

   (a) (2 marks) Read in and display (at least some of) the data.

   (b) (2 marks) Fit a logistic regression predicting whether a student will pass the exam, as it depends on the number of hours they studied. Your column `exam` is text, and will need to be turned into a `factor`. Display the results.

   (c) (2 marks) Is there evidence that the number of hours studying affects the probability of passing the exam? Explain briefly.

   (d) (2 marks) Is your Estimate for `hours` positive or negative? Does that make sense in the context of the data? Explain briefly.

   (e) (4 marks) Obtain predicted probabilities of passing the exam for students who study 0, 1.5, 3, and 4.5 hours. Are these consistent with what you said earlier? Explain briefly.