STAT 213 (SUMMER 2021): HW7

DUE ELECTRONICALLY BY FRIDAY 8/13/21

Instructions and Technical Tips

Write up your solutions and save/upload the file(s) to the RStudio server in the folder `stat213/turnin/hw7/` inside your Home directory. Include `hw6` (all lowercase) in your filenames.

Using an RMarkdown document is recommended, but not required.

RMarkdown Format. If you do use RMarkdown, R code should go in code chunks, and verbal commentary (and any math equations) should go outside code chunks.

Use section headings (a line starting with one or more `#` symbols) to demarcate the start of a problem. You can nest headings by using an additional `#` symbol for each level of nesting: one for a top-level heading, two for the next level within that, etc.

Periodically "Knit" your file to verify that it is working correctly.

"Knitting" Troubleshooting. If your code runs chunk by chunk but won't Knit, try clearing your environment (broom icon in the upper right) and running chunk by chunk from the start again. The most common cause is an undefined variable. This can happen if you change your variable names some places but not others, but a very common reason is that you read in the data from a file using a menu button instead of using the `read.file()` command, which means the "reading in the data" step is not recorded in your document. If you can't Knit, I won't be able to run your code either!

If you aren't able to Knit directly to `.pdf`, it may be because you are using special characters (such as $\neq$) in your `.Rmd`. Replace these with plain text and try again.

If you have done this and are still unable to Knit directly to `.pdf`, convert your Knitted `.html` or `.docx` into a `.pdf` (possibly by "printing" the `html` from your browser to a file) and save that there. A uniform file format across students will make grading 64 problem sets much more streamlined. Thanks!

What to turn in. If using Markdown, turn in the source file (ending in `.Rmd`) and the Knitted output file (preferably as `.pdf`). Otherwise, just turn in a `.pdf`.

Date: August 10, 2021.

Verifying receipt of your work. When I run the script that collects your work after the due date, you will see a "receipt" file in the ∼`/stat213/receipts/hw7/` directory. If this does not appear within 24 hours or so of you submitting your assignment, let me know.

Problems

1. The `CAFE` dataset (in the `Stat2Data`), also described in examples throughout Chapter 10 of the text, includes information about how various U.S. senators voted on an amendment that would hamper the proposed Corporate Average Fuel Economy (CAFE) bill. The bill would have tightened regulations on fuel economy standards, and so a Yes vote on the amendment acts in opposition to tightened regulations. Examine the documentation on the dataset for more information.

   The quantitative variable `LogContr` records how much money each senator received from the auto industry, on a log scale. The `Dem` indicator is 1 if the senator caucused with Democrats, 0 otherwise.

   Fit and compare a set of logistic regression models to address the following questions. For each question, identify the model(s) you used to address the question, and interpret each coefficient.

   (a) Does the probability of a `Yes` vote increase with (log) campaign contributions?

   (b) Does the probability of a `Yes` vote differ between parties?

   (c) Is the relationship between log campaign contribution and Vote different for those that caucus with Democrats?

   (d) Does knowing whether a senator caucuses with Democrats improve predictive ability *after controlling for* campaign contributions?

2. The data in `ICU` contains information for a sample of 200 patients who were part of a larger study conducted in a hospital's intensive care unit (ICU). Since an ICU often deals with serious, life-threatening cases, a key variable to study is patient survival, which is coded in the `Survive` variable as 1 if the patient lived to be discharged, and 0 if the patient died in the ICU. Among the possible predictors of this binary survival response are the following:

`Age`	age (in years)
`AgeGroup`	`1` if under 50; `2` if 50-69; `3` if 70+
`Sex`	`1` for female; `0` for male
`Infection`	`1` if infection suspected; `0` if not
`SysBP`	systolic blood pressure (in mm of mercury (Hg))
`Pulse`	heart rate in beats-per-minute
`Emergency`	`1` if admitted from emergency dept; `0` otherwise

Consider a multiple logistic regression model with `Survive` as the response variable using the three quantitative predictors in the data (`Age`, `SysBP`, and `Pulse`).

(a) After fitting the three-predictor model, does it appear as though any of the three quantitative variables are not very helpful in this model to predict survival rates in the ICU? If so, drop one or more of the predictors and refit before going on to the next part.

(b) The first person in the dataset (ID #4) is an 87-year-old male who had a systolic blood pressure of 80 and a heart rate of 96 bpm when they checked in to the ICU. What does your final model from part (a) say about the chances that a person with these characteristics would have of surviving their visit to the ICU? Give both a point and an interval estimate on the probability scale.

(c) The patient with ID #4 did not survive to be discharged from the ICU (`Survive` = `0`). Based on your answer to (b), would you say that this result (this patient dying in the ICU) was very surprising, mildly surprising, reasonably likely, or very likely?