STAT 213 (SUMMER 2021): HW2

DUE ELECTRONICALLY BY THE END OF CLASS, FRIDAY 6/11/21

Instructions and Technical Tips

Write up your solutions and save/upload the file(s) to the RStudio server in the folder `stat213/turnin/hw2/` inside your Home directory. Include `hw2` (all lowercase) in your filenames.

Using an RMarkdown document is recommended, but not required for the non-lab portion of the assignment. If using Markdown you can either put your answers to the non-lab problems in a separate document, or just add them to the end of the lab.

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!

Date: June 7, 2021.

What to turn in. You should turn in the source file (ending in `.Rmd`) and the Knitted output file for the `reading-in-data.Rmd` lab (in the `R-orientation` folder), and if you are using Markdown for the rest of the assigment, either a separate `.Rmd` and output file for that, or just add your responses to the other problems to the end of the lab document (clearly labeling what's what).

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/hw2/` directory. If this does not appear within 24 hours or so of you submitting your assignment, let me know.

Problems

1. **Lab 2**. Read through `reading-in-data.Rmd` lab (in the `R-orientation` folder) and complete the exercises. Don't put too much energy into the stuff about relative paths; do your best but if you get stuck on those parts, you can omit them without affecting your grade.
2. **Diet Plans** An article in the `Journal of the American Medical Association` reported on a study in which 160 subjects were randomly assigned to one of four popular diet plans: Atkins, Ornish, Weight Watchers, and Zone. Among the variables measured were:
   - Which diet the subject was assigned to
   - Whether or not the subject completed the 12-month study
   - The subject's weight loss after 2 months, 6 months, and 12 months (in kilograms, with a negative value indicating weight gain)
   - The degree to which the subject adhered to the assigned diet, taken as the average of 12 monthly ratings, each on a 1-10 scale (1 being complete nonadherence and 10 indicating full adherence).

   The primary goal of this study was to investigate whether weight loss tends to differ among the four diets. A secondary goal was to investigate whether the weight loss is affected by the adherence level.

   (a) Identify the explanatory and response variables for investigating the primary question.

   (b) Identify the explanatory and respone variables for investigating the secondary question.

   (c) If the researchers' analysis of the data leads them to conclude that there is a significant difference in weight loss among the four diets, can they legitimately conclude that the difference is because of the diet? Explain why or why not.

(d) If the researchers' analysis of the data leads them to conclude that there is a significant association between weight loss and adherence level, can they legitimately conclude that a cause-and-effect association exists between them? Explain why or why not.

3. **Olympic long jump.** Refer to the output below.

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -16.47       2.67   -6.18        0
## Year            0.01       0.00    9.19        0
```

```
## Analysis of Variance Table
##
## Response: Gold
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Year       1   5.69    5.69   84.47 < 2.2e-16 ***
## Residuals 26   1.75    0.07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(a) Based on the regression output, what is the intercept of the least squares regression line for predicting winning Olympic long jump length from year?

(b) Interpret the numeric value of the slope of the least squares regression line for predicting the winning Olympic long jump length from year in the context of this setting.

(c) Based on this output, what is the size of the typical error when predicting winning jump length from year?