

# STAD29 / STA 1007 assignment 6

Due Tuesday March 10 at 11:59pm on Quercus

Packages for this one:

```
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 problems 2 and 4.

1. Work through the rest of Chapter 22 of PASIAS.
2. Back in STAC32, we had some children who were learning to read. We now have some more, but the experimenters were concerned that the total income of each child's family might also affect the child's reading score. (You might imagine that a larger family income, other things being equal, would be associated with a higher (better) reading score.) There are, this time, four reading methods, labelled `method1` through `method4`. The data for this study are in [http://ritsokiguess.site/STAD29/reading\\_again.csv](http://ritsokiguess.site/STAD29/reading_again.csv).
  - (a) (2 marks) Read in the data and display (some of) the data frame.
  - (b) (3 marks) Make a suitable plot of the data. Add regression lines for each method (*without* the grey envelopes). Bear in mind that we are trying to predict reading score from everything else.
  - (c) (2 marks) Describe any effects of income and reading method on reading score that you see on the graph.
  - (d) (2 marks) Run an analysis of variance of reading score as it depends on reading method. Display the results.
  - (e) (3 marks) Compare, using a suitable graph or numerical summary, the reading scores for the different reading methods. What is your main conclusion?
  - (f) (3 marks) Repeat the previous part, but this time comparing the family income by reading method (and not the reading scores). Again, comment briefly.
  - (g) (3 marks) Run a suitable analysis of covariance, and use `drop1` with `test="F"` to test the significance of the two explanatory variables. What do you conclude?
  - (h) (3 marks) Compare the P-values for `method` from the analysis of covariance in the previous part, and the analysis of variance you did earlier. Which one do you think is more trustworthy? Explain briefly.
3. Work through Chapter 23 of PASIAS.

4. Researchers are comparing different ways to give technical information about diet. Specifically, 33 subjects are randomly assigned to one of three groups. The first group receives technical dietary information from a website. Group 2 receives the same information from a nurse practitioner, while group 3 receives the information from a video made by the same nurse practitioner. Each subject then made three ratings: difficulty, usefulness, and importance of the information in the presentation. The researcher looks at the three different ratings of the presentation to determine if there is a difference between the modes of presentation. In particular, the researcher is interested in whether the website is superior because that is the most cost-effective way of delivering the information. In the dataset, the ratings are presented in the variables `useful`, `difficulty` and `importance`. The variable `group` indicates the group to which a subject was assigned.

The data are in <http://ritsokiguess.site/STAD29/dietary.csv>.

- (a) (2 marks) Read in and display (some of) the data. Make some kind of comment about whether you have what you were expecting.
- (b) (2 marks) Explain briefly why MANOVA is something we would consider to analyze these data.
- (c) (4 marks) Make side-by-side boxplots of each of the three explanatory variables for each of the three groups. For full credit, do this in *one* `ggplot`. Hint: use the idea from C32 for plotting residuals against each of the explanatory variables in a regression, all in one plot.
- (d) (2 marks) What does your plot suggest to you about the research objectives mentioned in the opening paragraph? Explain briefly.
- (e) (4 marks) Run a suitable MANOVA on these data. Think about what your response variables are, what you have to do with them, what your explanatory variable(s) is/are, and how to run the analysis. There are several steps. What do you conclude in the end?