STAD29 / STA 1007 assignment 5

Due Tuesday Feb 25 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 questions 2 and 3 of the following:

1. Work through, or read through, chapter 22 of PASIAS. Problem 22.2 will prepare you for question 2, and problems 22.3, 22.5, and 22.6 will prepare you for question 3.

2. The number of hours that a battery operates might depend on the material it is made of, and the temperature at which it is operated. In an experiment, three materials (labelled A, B, and C) were tested, and three temperatures: Low (-10 ° C), Medium (20 ° C) or High (45 ° C). Twelve batteries were randomly selected from each material type and were then randomly allocated to each temperature level. The resulting life of all 36 batteries is shown in `http://ritsokiguess.site/STAD29/batteries.txt`, with the data values separated by spaces.
   (a) (2 marks) Read in and display (some of) the data.
   (b) (3 marks) Which order are the temperatures in, as far as R is concerned? Does that make sense? If not, put them in the right order by creating a column in the data frame that has the right order. (Hint: what order are they in in the data frame?)
   (c) (2 marks) Make a suitable plot of these data, given the number and types of variables you have. Put temperature on the x-axis.
   (d) (3 marks) Make an interaction plot, again putting temperature on the x-axis.
   (e) (2 marks) What do you conclude from your interaction plot? By looking at your first plot, explain briefly why your conclusion from your interaction plot makes sense.
   (f) (3 marks) Run a suitable analysis of variance, including interaction, and display the results. Was your interaction significant?
   (g) (2 marks) Explain briefly why simple effects would be a useful technique for this data set, and give an example of a comparison you would be able to make with them.
   (h) (4 marks) Find the simple effects of material at each temperature. That is, for each temperature, compare the materials at that temperature using `aov` and (if necessary) Tukey, and state your conclusions in the context of the data.

3. R has a number of built-in data sets. One of them is called `PlantGrowth`. This consists of 30 observations from an experiment to compare plant yield (measured by the dried weight of plants) under two treatment conditions and a control condition. We have two research hypotheses to consider: whether the average of the two treatments is different from the control, and whether the two treatments differ from one another.

(a) (1 mark) Display (some of) the data set.

(b) (2 marks) Make a suitable plot of the data.

(c) (2 marks) Why is this a situation where contrasts would be helpful? Explain briefly.

(d) (3 marks) Set up contrasts for the two hypotheses of interest. That is, define two vectors with mnemonic names whose values reflect what you want to compare with what. (To get the order right, think about what order the treatment groups came out on your plot.)

(e) (2 marks) Demonstrate that your two contrasts are orthogonal.

(f) (2 marks) Use your two contrasts to set up for `lm` to test them via `summary`.

(g) (2 marks) Fit an appropriate model and display its summary.

(h) (3 marks) What do you conclude? Explain briefly why that makes sense by looking at the plot you drew earlier.

Notes

[1] Otherwise, why make it?

[2] It is not enough to be able to parrot the definition; you need to *understand* the definition well enough to be able to use it.

[3] You should not look at the Tukey comparisons when the ANOVAs were non-significant, but here none of those are significant anyway.

[4] As for regression summaries, the Tukey output is for looking at rather than computing with.