

ENGSCI 314, First Semester, 2018

Third Statistics Assignment, Due: 1pm Tuesday, 29TH MAY

Instructions concerning this assignment:

I am providing you an R Markdown document called **ENG314A3.rmd** (available on Canvas) which will have some answers already filled in. You will need to fill in and complete the rest of the document. The data files you will be using for the assignment are described in the questions and are available from Canvas. Make sure you put these datasets in the same place you put the R markdown document because it is going to look for them there. The first change you need to make to the markdown document is put your name and ID number at the top.

- The assignment will be worth 5% towards your final grade.
- **Hand the assignment in to me in my office - room 314A, building 303.** If I am not in, you can hand it in to the Stats office (room 305) and ask them to pass it on to me.
- The total marks for this assignment will be **45** (this includes 5 marks for presentation and communication).
- There are **5 Presentation and Communication marks** for this assignment as follows:
 - **Name and ID number** at top of R Markdown document.
 - **Space saving and printing assignment 2-up.** Not printing out unnecessary output (listing ENTIRE data sets or showing erroneous R output). Assignment work printed out in "2-up" layout. 2-up layout prints 2 pages side-by-side reduced to one page.
 - **Readability.** This is for your general communication ability in the assignment. This includes sentences clearly conveying the correct idea; sentences making sense; comments not being excessively long or short; conclusions following logically from previous statements.
 - **Use of Natural Language in Executive Summaries.** In executive summaries, this is for discussing the analysis in context, not using variable names, using units when known and rounding sensibly.
 - **Keeping to the Point in Executive Summaries.** In executive summaries this is for not going into far more detail than required.

Notes: Questions **1 and 2** are open questions. The approach to answering them is:

- Comment on the questions of interest or the goal of the analysis.
- Look at the plot of the data and comment on it.
- Fit a model to the data
 - Check the model assumptions.
 - Change model and repeat checks as needed. You may have to do this more than once.
- Generate inference output required from final model.
- Write a Method and Assumption Checks section.
 - This will detail the steps you took and why you took them in building the model.
 - It will include brief descriptions of the model assumption checks.
 - It will include a mathematical statement for the final model you fitted.
- Write an Executive Summary.

Question 1. [23 Marks]

A software company is developing a new computer game. The manager wants to see what effect several different variables have on the length of time players survive in the game. This information could then be used to develop combinations of settings so that consecutive levels of the computer game get more difficult. A large group of play testers is available and each is randomly allocated to play a version of the game with varying combinations of the variables. The resulting data is stored in the file "game14.txt", which contains the variables:

Time	The survival time for the player (in minutes).
Speed	The speed setting at which the software was running (set on a numeric scale from 1 to 9, higher values = higher speed).
Strategy	The complexity of the strategy used by the computer opponent in the game (either "Low", "Med" or "High" complexity).
StrategyC	As above, but recoded with single characters (either "L", "M" or "H").
Background	The background setting for the game (either "Factory", "Rural" or "Urban").
BackgroundC	As above, but recoded with single characters (either "F", "R" or "U").

Notes:

- Limit any comments about the pairs plot to 2 or 3 brief sentences.
- For simplicity sake, DO NOT include any interaction terms in your model (they will prove to be non-significant and we haven't dealt with multiple interactions).
- In addition to the usual analysis, create a plot showing only the variables in the final model and also showing the final fitted model. (The coded factor(s) will be useful here.)
- For the final fitted model, check out all possible pairwise comparisons for any significant factors.

Question 2. [17 Marks]

A researcher conducted an experiment to see if food cooked in iron pots has a higher iron content than food cooked in other types of pots. Meat dishes and vegetable dishes were cooked in three different types of pots (iron, clay or aluminium). The iron content of the food was then measured. The data are stored in the file "Iron.txt", which contains the variables:

Iron	The iron content of the food (in milligrams per 100 grams)
Food	The type of food cooked: Meat or Vegetable
Pot	The type of pot used to cook the food: Aluminium, Clay or Iron

Is the iron content of food cooked in iron pots greater than that cooked in other types of pots? Do any differences depend on the type of food cooked?

Notes:

- When you apply Occam's Razor – that is, simplify the model by successively removing the least significant term, make sure you only remove terms one at a time.