STAT 3012/3912 Applied Linear Models – Semester 1, 2018

Assignment 2

Due: Fri 8/6/2018 1pm
Submit via Turnitin on canvas.

Notes and Instructions:

- This assignment carries weight 7.5% towards your final mark for STAT3012 and 5% for STAT3912.
- You will be assessed on your ability to articulate a sound report in the format set output in `report_template.Rmd`. All the R scripts to replicate your finding should be visible. Your assumptions should be clearly stated. You should submit the pdf generated from after you "knit" `report_template.Rmd`.
- In statistical analysis, domain knowledge and soft skills are just as important as the hard skills like mathematics. You will probably not have the domain knowledge required to understand the data in context and you will need to be able to get the gist of the context of the data. You will have opportunities to extract relevant domain knowledge from Emi. During this time
  - Emi will pretend to be a plant breeder with some basic understanding of statistics.
  - Emi as a plant breeder is quite busy with field work and is not good at replying back to emails. She comes time to time to the main campus (to be specific, she will be there at your lecture times) and that will be your opportunity to ask her questions related to this assignment (note: this question time will be held at the end of the lecture after the main teaching material is taught if time permits). No questions related to assignment will be answered outside of these times. No questions related to assignment should be asked at consultation hour aside from administrative matters.
  - If there are questions by multiple students, the order will be determined by their seating in the lecture. The priority decreases by the distance of the seat away from the lecturer, i.e. the person seated closest to the lecturer will get to voice their question first.
- Do NOT include your name in the assignment.
- In submitting your assignment in Turnitin, you acknowledge that the work is in your own. Plagiarism is not tolerated.

Help with statistical analysis

Emi Tanaka <emi.tanaka@sydney.edu.au> Fri, May 18, 2018 at 10:00 AM
To: "you@uni.sydney.edu.au" <you@uni.sydney.edu.au>

Dear statistician,

Thanks for agreeing to come onboard this project. As I mentioned to you, I made a hybrid between wheat and barley using a special procedure. I'm calling this hybrid Warley[1]. Since then I have been able to increase the diversity to 10 different genetic lines and multiply enough seeds for testing.

I believe Warley has potential to become a major staple food. It has substantial yield compared to current commercial wheat varieties. It's possible that it can replace wheat as the staple food! I conducted the trials at 3 locations testing 10 Warley testing lines. I have yield measured in tonnes per hectare in the attached file. I want you to analyse this data to find out the best Warley lines.

I also need to start planting for my next experiment so could you generate an experimental design for the next one? The layout for the next experiment is the same as this one, i.e. locations are Narrbari, Northstar and Horsham and I have a land with plots laid out in a rectangular array of 5x4, 5x8 and 10x10 available. I want to test the top two lines from this data in the next experiment and also I will have 3 additional new lines. I should have enough seed to sow 40 plots for each lines.

Thanks statistician!

Kind regards,

Emi

Dr. Emi Tanaka
Warley Breeder

1 attachment
yield.csv

[1]This is all made up.