STA130H1S – Fall 2022

Week 10 Tutorial Handout

Today's agenda (5 min):

- Q&A/vocabulary list

- Group Presentation

- Work on final project

This Week's Vocab (15-20 min) :

- Multivariate Linear Regression (versus Simple Linear Regression)
- Review: R^2 and Indicator Variables
- Variable Selection / Hypothesis Testing
- Multicollinearity = Observed Confounding
- Statistical versus Practical Significance
- 80/20 Train-Test Split
- In Sample Scoring / Out of Sample Scoring
- Root Mean Squared Error (RMSE)
- 80/20 Train-Test Split Randomness
- Interpreting Multivariate Linear Regression Coefficients
- Baseline Groups and Indicator Variables
- Interactions
- Prediction
- Model Comparison
- `rowid_to_column()`
- Training Data
- Test Data

Discussion (15 min) : In the homework this week, you've read the story of Rosalind Franklin behind discovery of DNA structure. Take some time to think about the following questions and exchange your answers with your peers:

- Why is it difficult to pin down bias (i.e. none of the stakeholders *neither Watson and Crick, nor Max Perutz nor Maurice Wilkins (not to mention John Randall, the journal editors, and the relevant lab leads)* directly reached out to Rosalind Franklin and discuss the circumstances of the data, despite recognizing the importance of her work)?

- Are there instances where you succumbed to personal bias. How did you become aware? What would you have done differently now that you're aware of it?

- Did you have an experience where you've faced personal bias? What did you or would you do to make the other person aware of the bias you're facing?

Group Presentation (50 min) :

I will post the following material on Quercus in advance for students to access

Prior to this tutorial, students should have reviewed the videos on Plagiarism and Hedging.

Please read the following article: Ross *et al.* (2022) "Women are Credited Less in Science than are Men". This article is available on quercus (link: https://q.utoronto.ca/courses/253019/pages/m8-problem-set-8-pa per?module_item_id=3512104) and the paper online https://www.nature.com/articles/s41586-022-04966-w. (If you don't yet know how to get access to a journal article through UT library, it's a good time to ask for help from your TA or your peers.)

Once you have watched the videos and read the artcile, you will prepare a 10-min group presentation providing a summary of the above study.

Make sure you explain the following, at a minimum:

- Objective: What are the authors interested in studying? Why is it important to conduct this study?
- Methods: What type of study design is used? Who are the participants? What statistical tests are used?
- Results: What are the main findings of the study? Make sure you support any statements with facts (e.g. proportions, p-values, etc.), and present the results in the context of the study.
- Conclusions: What are the most important take away messages to you? Are there any important limitations or future directions?

Some things to keep in mind

- The introduction should state the overall purpose of the presentation
- Each main point should be "signaled." Use *signal phrases* or *signposts* to transition between your points.
- Every group member must speak during the presentation or marks will be deducted for participation.

Work on final project (20 min):

Let the students group and summarize what they've done for the final project. Ask each group to briefly talk about their research project and others to give feedback. This won't be marked but hopefully provides an oppotunity for students to develop analytical thinking required for giving constructive feedbacks as well as the making use of suggestions from their peers.