

STA1008H: Applied Statistics**Instructor:** Jun Young Park, PhD**E-mail:** junjy.park@utoronto.ca**Lectures:** Tuesday 10 AM-noon, Friday 10-11 AM**Office hours:** The official office hours for this course are Fridays between 11 AM and noon. Additional office hours are available on Tuesdays between 1 and 2 PM (Sept 14, 21, 28). To meet university's policy on safety, the office hours will be online.**Course deliveries:** Lectures will be delivered online via Zoom for the first two weeks (until September 23, 2021). After then, the course will go in person. The classroom will be

- Tuesdays: O.I.S.E (252 Bloor St W.), room 8200
- Fridays: Bahen Centre Information Tech (40 St. George St), room 2139

We will mostly use Friday to cover R software related to course materials. Please check Quercus regularly for updates and surveys.

Course description: This course is designed to provide key statistical insights to researchers using statistics for their research. After reviewing key concepts in statistics, we'll move on to understand the properties of linear models and its extensions to more advanced methods, including different types of analysis of variance (ANOVA) and random effects. Computer-intensive methods, including Monte Carlo, permutation, and bootstrapping, will be introduced to understand the context better. As time allows, we will cover up to two additional topics based on students' demand.

Software: This course uses R and students are expected to learn basic R programming throughout this course. The most recent version is R-4.0.2, which can be downloaded in this link: <http://cran.utstat.utoronto.ca/>. It is also highly recommended to install RStudio, an integrated development environment (IDE) for R. It is free for personal use and it is available for download at <https://rstudio.com/products/rstudio/download/>.

Evaluation: Homework (50%), 1 midterm exam (take-home, 25%), 1 final exam (take-home, cumulative, 25%)

- There will be **five** homework sets, requiring a solid understanding of theories and computing. Each homework is due sharply at the appointed date, and any late submission will be penalized. Students are encouraged to discuss homework problems with colleagues, but *every student must write up solutions independently*. Do not copy others' work.
- It is **not** allowed to discuss exam problems with anyone except for the instructor.

Prerequisites: It is assumed that students are familiar with introductory statistics covering the followings:

- Exploratory data analysis: boxplots, histograms, scatterplot
- Basic probability: expectations, variance, conditional probability
- Normal distribution: z test, z table
- Sampling distribution

- Hypothesis testing (p -value, confidence interval, Type 1/2 errors)

Topics

- Review of introductory statistics and introduction to simulation studies and permutation tests
- Correlation and simple linear regression
- One-way ANOVA design
- Multiple linear regression and two-way ANOVA with interactions
- Model checking and variable selection.
- Introduction to correlated data and analysis random effects model
- Bootstrapping
- Logistic regression

Textbooks: There is no single textbook that addresses all topics covered in this course. The course slides will be uploaded to the Quercus. I will select topics from the textbooks below, but please note that the levels of the textbooks are highly variable.

1. *Applied Longitudinal Analysis* by Fitzmaurice et al., 2nd edition
2. *Applied Linear Regression Models* by Kutner, Nachtsheim, and Neter, 4th edition
3. *Applied Multiple Regression/Correlation for the Behavioral Sciences*, Cohen et al, 3rd edition
4. *Answering Questions with Data* by Crump, available at crumplab.github.io/statistics/
5. *Data Analysis and Graphics Using R: An Example-Based Approach* by Maindonald and Braun, 3rd edition
6. *Linear Mixed Models: A Practical Guide Using Statistical Software* by West et al., 2nd edition
7. *Linear Models with R*, Faraway, 2nd edition
8. * *Mathematical Statistics with Resampling and R* by Chihara and Hesterberg, 2nd edition
9. *Statistical Methods for Psychology* by Howell, 8th edition
10. * *The Statistical Sleuth: A Course in Methods of Data Analysis* by Ramsey and Schafer, 3rd edition

*: textbooks where most homework problems will be adopted.

Challenges: For any other challenges during your study, please visit Health and Wellness Centre or the Graduate Wellness Services.