DATA 603 L02: Statistical Modeling with Data

Tuesday/Thursday 5:00pm-8:00pm
October 22nd 2019-December 6th 2019

Instructor: Dr. Thuntida Ngamkham
Office: ICT517
Email: thuntida.ngamkham@ucalgary.ca

Course Description: (3-0) An introduction to the creation of complex statistical models, including exposure to multivariate model selection, prediction, the statistical design of experiments and analysis of data in R.

Course Hours: H (3-0)

Prerequisite(s): Data Science 602 and admission to the Post-baccalaureate Certificate in Fundamental Data Science and Analytics or the Post-baccalaureate Diploma in Data Science and Analytics.
Coverage of the courses includes the following topics:

- Multiple Linear Regression
- Model Diagnostics
- Model selection
- Logistic Regression
- Experimental Design
- One factor, Completely Randomized Design (CRD)
- Blocking and the Completely Randomized Block Design (CRBD)
- Two Factor
- Nested Design

Evaluation Methods:

Component(s)	**Weighting (%)**	**Date**
Assignments (4)	40%	Nov 1st , 22nd, 29th ,Dec 6th
Quizzes (2)	30%	Nov 7th , Dec 5th
Project (1)	30%	Dec 3rd

Assignments

You will be given approximately 4 assignments which are based on the material covered. Quizzes will be constructed with the expectation that problems appearing on such assignments have been completed in full and that you have a complete comprehension of the course material.

Your assignment should be your own work and must be completed in **R Markdown** and submitted as either an .html or .pdf file.

Quizzes

Two quizzes will be given during the semester. The quiz problems will be developed from material presented in lecture. These will primarily consist of programming exercises. All quizzes are conducted in an open book and open notes manner, discussion amongst peers will not be permitted in these quizzes.

Project

One project will be assigned during the semester. The project shall be completed in groups of 3-4 students. The purposes of the project are for students to be able to design and conduct an appropriate model or experimental design, as well as analyze and interpret an appropriate model or an experimental design with your own data set. Finally, the students be able to analyze the results of the selected model or the experimental design. Each student in a group is expected to make a full intellectual contribution to the project.

Syllabus and Date:

Lecture	Date	Topics
1	October 22nd	Discussing the additional course outline. Reviewing Simple Linear Regression. Introducing Multiple linear regression. Using R for multiple linear regression estimates for first order modelling. Compute confidence Intervals for regression coefficients, and test a relationship between response and predictors. Compute a coefficient of determination (R^2) and an adjusted coefficient of determination (R^2) and (adj R^2) with interpretation.
2	October 24th	Compute confidence Intervals for regression coefficients and test their coefficients. Introduce a quadratic model, an interaction effect with dummy variable.
3	October 29th	Introduce Model selections by using R: Stepwise, forward and backward selection. Using R to determine the best fit model for data.
4	October 31st	Introduce Model diagnostics for Multiple linear regression. Using R to determine the validity of the model assumptions. Apply the method to some model selected from Stepwise selection.
5	November 5th	Continue investigating on Model diagnostics and Introduce Simple Linear Logistic Regression
6	November 7th	Introduce Multiple linear logistic regression model. Using R for modelling binary response data with quantitative and qualitative predictors. Interpretations of Logistic Regression Coefficient in the Logistic Model and making predictions.
Term break Sunday-Saturday, November 10th -16th		

7	November 19th	Comparing models with Wald tests and likelihood ratio tests. Introduce Model diagnostics for Logistic linear regression. Analytical strategies. Introduction to Experimental Design. One factor: Completely Randomized Design (CRD). Terminology and concepts. Analysis of Variance. Model Assumption by using R. Non parametric method: Kruskal -Wallis Test for comparing more than two independent samples.
8	November 21st	Introduction to Blocking and the Completely Randomized Block Design (CRBD).Terminology and concepts. Analysis of Variance. Model Assumption. Non parametric method: Fried man Test for Randomized Block Design
9	November 26th	Introduction to Two Factor: a x b factorial design. Main effect and Interactions with hypothesis test by usig R. Fixed and random Effect. Two way Analysis of Variance for fixed and random effect model.
10	November 28th	Model Assumption for Two Factor: a x b factorial design. Introduction to Nested Design with the Analysis of Variance.
11	December 3rd	Analysis of Variance for Nested Design with Hypothesis Testing for both fixed and random effect and Nested Design Assumption. Discussion about the project.
12	December 6th	Project Presentation and flex class (build in to ensure that I have covered all the topics above).