

# STA302 : Methods of Data Analysis I

University of Toronto

Department of Statistical Sciences

Course instructor : Cédric Beaulac

Email : [cedric@utstat.toronto.edu](mailto:cedric@utstat.toronto.edu)

Section : LEC0101

Lectures : Tuesday and Thursday 14:00-17:00, MC 102

Office Hours :

- Instructor : Tuesday and Thursday 13:00-13:55, Stewart Building Room 103A

- TA : Wednesday and Friday 10:00-12:00, SS623

Teaching assistant :

• Alin Morariu - [alin.morariu@mail.utoronto.ca](mailto:alin.morariu@mail.utoronto.ca)

• Sergio Betancourt - [sergio.betancourt@mail.utoronto.ca](mailto:sergio.betancourt@mail.utoronto.ca)

Course Website : [Quercus](#)

## Prerequisites

• STA238H1/ STA248H1/ STA255H1/ STA261H1/ ECO227Y1

• CSC108H1/ CSC120H1/ CSC121H1/ CSC148H1

• MAT221H1(70%)/ MAT223H1/ MAT240H1

## Course Objectives

The course covers the theory and application of linear regression analysis. The main purpose is to give students tools to confidently perform basic analysis of simple data sets. By the end of the course, student should feel comfortable importing a data set on R, checking if the linear model is a suitable tool and if so proceed with the analysis.

## **Textbook**

All of the textbooks bellow are available online on the University of Toronto library (the titles are actually clickable links!), which is a reason why I selected them. I will not strictly follow a textbook but wonder around many. The lectures slides will be inspired from a mix of these textbooks :

- Main textbook : *A Modern Approach to Regression with R*, Simon J. Sheather, Springer
- *Linear Models with R*, Julian J. Faraway, Taylor & Francis Group
- *A modern introduction to probability and statistics : understanding why and how*, F.M. Dekking et al., Springer
- *Experimental Design: Procedures for Behavioral Sciences*, Roger E. Kirk, Sage Publications

## **Communication**

As there is many registrations, I would ask students to communicate with the Sergio Betancourt (week 2, 3 and 4) or Alin Morariu (week 5, 6 and 7) instead of the instructor. They are suited to answer most of the questions, if not they will transfer the email to me. Thank you all for your understanding.

Please include the exact string "STA302H1F" in your email subject line; otherwise you email may not be read. Questions regarding course material should be ask during office hours. A TA may refuse to answer a question in an email at his discretion and ask you to come to ask it during office hours.

## Tentative Schedule

Week	Date	Topic
1	May 7th	Introduction, p-values and statistical significance
	May 9th	Hypothesis testing, t-test and ANOVA
2	May 14th	Linear regression: Least square error formulation
	May 16th	Linear regression: Maximum Likelihood formulation
3	May 21st	<b>Test #1</b>
	May 23rd	Diagnostic for the linear regression model
4	May 28th	Dummy variables and introduction to multiple linear regression
	May 30th	Interactions and multiple linear regression assumptions
5	June 4th	<b>Test #2</b>
	Jun 6th	Model selection and variable selection
6	June 11th	Ridge and Lasso regression
	Jun 13th	Statistical analysis, data science and ethics

## Lectures structure

The lectures consist of an 1h30 to an 1h45 of theory followed by an 1h of R programming. A break will separate the two sections.

## My expectations

I expect students to attend both parts of the lecture. Attending the lectures is mandatory and necessary to succeed the course. I'm also expecting a minimum of interest towards statistics and sciences.

## Evaluation

Assusment	Weight	Date
Test #1	30%	May 21st
Test #2	30%	June 4th
Final exam	40%	June 19-26

## **Assessments**

You must bring your student identification to the tests and final exam. These are closed book and closed notes. Only the concepts covered in lectures will be evaluated and practice problems will be made available.

## **Marking concerns**

Any requests to have a test re-evaluated must be made in writing within one week of the date the test was returned. Request must be sent directly to TAs (Sergio for the first test and Alin for the second). Requests must include a detailed reason for the re-evaluation.

## **Missed Tests**

If a test is missed for a valid reason, please communicate with the course instructor within 5 days of the evaluation and provide the needed documentation within two weeks.

If it is for a valid medical reason, submit a copy of the University of Toronto Verification of Student Illness or Injury form to your instructor.

If a test is missed for a valid reason then the weight of the test will be added to the final exam.

## **Academic integrity**

You are responsible for knowing the content of the University of Toronto's Code of Behaviour on Academic Matters at [www.governingcouncil.utoronto.ca/policies/behaveac.htm](http://www.governingcouncil.utoronto.ca/policies/behaveac.htm) If you have any questions about what is or is not permitted in this course, please do not hesitate to contact your instructor.

## **Accessibility needs**

Students with diverse learning styles and needs are welcome in this course. In particular, if you have a disability/health consideration that may require accommodations, please feel free to approach me and/or Accessibility Services at (416) 978 8060; [accessibility.utoronto.ca](http://accessibility.utoronto.ca)