

STA302H1F - Methods of Data Analysis I

Wei(Becky) Lin

Summer 2017

COURSE DESCRIPTION

The main objectives of this course are to gain a solid understanding of the theory and application of linear regression analysis and practical skills for developing linear regression models for inference, prediction and interpreting the results. Topics to be covered include: initial examination of data, correlation, simple and multiple regression models using least squares, geometry of least squares, inference for regression parameters for normally distributed errors, confidence and prediction intervals, model diagnostics and remedial measures when the model assumptions are violated, interactions and dummy variables, ANOVA and model selection. Extensive use of **R** language to analyze data and use **Rmarkdown** for homework solution.

PRE-REQUISITE

Students should have a second year statistics course such as STA 248H1 / STA 255H1 / STA 261H1 / ECO 227Y1. Students are expected to also have the mathematics pre- and co-requisites required by students in these second-year statistics courses. You will need to know basic matrix operations. A short but good review of matrix algebra is available on the website at

https://onlinecourses.science.psu.edu/statprogram/matrix_review.

LECTURES

- Section L0101: Tuesday/Thursday 2-5pm in MC102.
- Section L5101: Monday/Wednesday 6-9pm in MC102 (no lecture on Monday, May 22 (Victoria Day))

How to find the classrooms? Please check out

http://www.osm.utoronto.ca/map/f?p=110:1:0::NO::P1_SEARCH:

INSTRUCTOR & TA OFFICE HOURS

- **Wei(Becky) Lin** (wei.lin@mail.utoronto.ca). Office: SS6011.
- Office hours (**an ideal time to discuss questions that you have**)
 - Instructor: Monday 11:00am-12:30pm in SS6011 (starts from 2nd week. For May 22, I will make it up at the same time on Tuesday, May 23.)
 - TA office hours will be scheduled and posted soon.

In general, I am not able to answer questions about the course material, assignments, and tests by e-mail for such a big class. If you really want to ask me question through mail, you might have to wait for couple days. Questions about the course material or assignments that are more appropriately discussed in tutorial or during office hours. Before you send an e-mail, make sure that you are not asking for information that is already on the course web site or the Piazza discussion board on UT portal, if you do not get a response, this may be why.

Announcements will be posted on Blackboard. Please check there regularly. If an urgent matter arises, I may contact the entire class by e-mail. In order to receive these message, please make sure you that you use your mail.utoronto.ca account so that the message won't automatically go to my Junk folder.

COURSE WEBSITE AND PIAZZA ONLINE DISCUSSION BOARD

Weekly lecture notes, assignments, practise problems, announcements and a discussion board are available on

<https://portal.utoronto.ca>.

Please note that:

The **online discussion board on Piazza**, a TA (Sarah) is assigned to answer questions your have. If you post your questions there and don't get response in 3 days, please inform me ASAP. Piazza sign up link:

piazza.com/utoronto.ca/summer2017/sta302

TEXTBOOKS

- *Applied Linear Regression Models*, 4th edition by Kutner, Nachtsheim, and Neter. (We would cover chapter 1-8, and some from chapter 9-11)
- Reference (recommended)
 - *A Modern Approach to Regression with R* by Simon J. Sheather.
 - *Applied linear regression* 4th edition by Sanford Weisberg.

EVALUATION

	Weight	Date	Time	location
Midterm	24%	June 5th (L5101), June 6th (L0101)	6-8pm (L5101), 2-4pm (L0101)	EX200 EX200
Assignment 1	3%	Thursday, May 25th	L0101/5101: due 11pm	Crowdmark
DataCamp	3%	June 23rd.	L0101/5101: due 11pm	DataCamp
Assignment 2	10%	Thursday, June 1st	L0101/5101: due 11pm	Crowdmark
Assignment 3	10%	Thursday, June 22nd	L0101/5101: due 11pm	Crowdmark
Final Exam	50%	By A&S on June 9	TBA	TBA

The midterm and exam are both **closed book and closed notes**, a non-programming calculator is allowed. The The location of the midterm will be announced later. The midterm papers are different for sections L0101/L5101 but the finals are the same. Practice problems will be posted on portal to help you prepare for the midterm and exam and are not to be handed in. Each assignment will mainly be a data analysis project for which you will use R. **You need to know the basic R syntax for the midterm and exam, and you must also know how to interpret output from R.**

If the midterm is missed for a valid reason, you must provide appropriate documentation, such as the University of Toronto Medical Certificate, University of Toronto Health Services Form, or College Registrar's Letter. You must submit this documentation within one week of the test. If documentation is not received in time, your midterm mark will be zero. If midterm is missed for a valid reason, the weight of your midterm mark will be shifted to your final. **The midterm is scheduled in the lecture time, please enrol the course only if you are able to write the midterm.**

No late assignments will be accepted without documentation of a valid reason. **Late assignment without a valid reason will get penalty of 10% per day off.**

Any requests to have marked work re-evaluated must be made within 7 days (one week) of the date the work or result was returned to the class. The request email must contain a justification for consideration and your clear section information. **All all remark requests, the whole assignment/ test will be remarked.** There is chance that you might get lower mark points after remarking.

COMPUTING

We will be using R and RStudio. The main advantages of R are the fact that R is freeware and that there is a lot of help available online. Regarding how to install R and RStudio, and learn the basic syntax of R, refer the documents created by Paul Torfs & Claudia Brauer

<https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf>

I am assuming that students have never used R before. I will give you a quick introduction to R and provide you with R source code for all of the examples in lecture, which should be sufficient for you to do your assignments. The two reference books recommended are also good to help you pick up R. Note that there are many graphics options available to produce the sophisticated plots that are in the book or online, but we will focus on the basics. There are many good reference online, here is another 100 page document which I found very helpful:

<http://www.utstat.toronto.edu/~brunner/help/R-intro.pdf>

Introduction to R by Venables, Smith and others.

For assignment, you will use Rmarkdown to write your solution. I will provide you the template. And you give your answer in the between. To learn more about R markdown, refer to

<http://rmarkdown.rstudio.com>

ACADEMIC INTEGRITY

It is academic dishonesty to present someone else's work as your own, or to allow your work to be copied for this purpose.

Here are some guidelines that apply to the computer assignments.

- In this course, it is always okay to use computer code that is presented in lecture or the textbook. Use it any way you like; you are responsible for the results.
- *The biggest danger is copying from other students in the class.* It is fine to discuss the assignments and to learn from each other, but don't copy. Never look at anyone else's printouts or show anyone yours before the quiz or exam when they might be handed in.
- Above all, do not allow anyone to see your program file before an assignment is due, and do not look at anyone else's. Never photograph someone's solution or allow yours to be photographed. To repeat: **the person who allows her/his work to be copied is equally guilty, and subject to disciplinary action by the university.**
- It is acceptable to get help with your assignments from someone outside the class, but the help must be limited to general discussion and examples that are not the same as the assignment. As soon as you get an outside person to actually start working on one of your assignments, you have committed an academic offence.
- *Don't copy, and don't let anyone copy from you.* If we catch you, you will get in big trouble.
- If this is not clear enough, the latest version of the student handout "How not to Plagiarize" is available at <http://www.writing.utoronto.ca/advice/using-sources/how-not-to-plagiarize>

You are responsible for knowing the content of the University of Toronto's Code of Behaviour on Academic Matters at <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 me. It is legitimate to discuss assignment problems with other students in the class or discussion board on portal. However, assignments must be written up completely by yourself. Do not let other students read your completed assignment solutions as this can lead to copying. Failure to comply with this is a serious academic offense.