Statistical Sciences
UNIVERSITY OF TORONTO

STA 302H1S / 1001HS - METHODS OF DATA ANALYSIS I
Autumn 2019
September 8 - December 20, 2019

Classes: **SECTION L0101**
Tuesdays 10:10-12:00 and Thursdays 10:10-11:00 in **OI G162**

SECTION L0201
Tuesdays 15:10-17:00 and Thursdays 16:10-17:00 in **BT 101**

SECTION L0301
Wednesdays 14:10-17:00 in **BR 200**

(except November 4-8 during Reading Week)
Campus map: http://map.utoronto.ca/utsg

Course website: Available through https://q.utoronto.ca (UofT Quercus)

Instructor: Dr. Shivon Sue-Chee
E-mail: shivon.sue.chee@utoronto.ca
Office: HS 384

Office hours: (Drop-in) Will be decided later and posted in our website
or by appointment (beginning Sept. 17)
More TA office/lab hours will be scheduled before test and assignments due dates.

Course content
The course provides a solid introduction to data analysis with a focus on the theory and application of linear regression. 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, model selection, and penalized regression.

Other topics include non-linear regression and non-parametric data smoothing techniques. Emphasis will be on methodology and interpretation of the results of data analysis. This course will also be an opportunity to develop skills in data analysis for which `R` software and `Rmarkdown` will be used.

The main learning objectives of this course are:

- to gain a solid understanding of linear regression analysis and

- to learn practical skills for developing linear regression models for inference and prediction, and to interpret the results.

Pre-requisite
Students should have a second year statistics course such as {STA238, STA248, STA255, or STA261}, a computer science such as {CSC108, CSC120, CSC121, or CSC148} and a mathematics course such as {MAT221(70%), MAT223, or MAT240} or equivalent preparation.

You will need to utilize basic matrix operations. The course textbook includes matrix algebra review materials and additional resources will be available in our course website. **Pre-requisites will be strictly enforced by the department, not the instructor**. Your computing experience will be transferred to learning and/or extending our use of `R` (and `RStudio`).

Textbook

- *SJS: A Modern Approach to Regression with R* by Simon J. Sheather (Springer)
  Chapters 1-7 contain relevant materials. Available for purchase at the UT Bookstore and as an electronic resource through the University of Toronto library website. Datasets and other resources are available at the textbook's website: http://gattonweb.uky.edu/sheather/book/.
  Practice problems from the textbook(s) will be posted on the course website. They are to help you prepare for quizzes, test and exam and are not to be handed in.

Additional Textbook References

- *KNN: Applied Linear Regression Models, 4th edition* by Kutner, Nachtsheim, and Neter (Mc-Graw Hill).
  Chapters 1- 8 contain relevant materials. A copy of this text will be on reserve at the Mathematics Library.)

- *Weisberg: Applied linear regression, 4th edition* by Sanford Weisberg (Wiley). The third edition is available as an electronic resource through the University of Toronto library website.

Evaluation

Undergraduate students will be evaluated based on the following schemes.

	Scheme 1*	Scheme 2*	Date	Time	Location
Participation*	0%	5%	Sept. 19- Dec. 4		in class/online
Assignment 1	4%	4%	**Mon., Sept. 23	due by 10pm	online
Assignment 2	6%	6%	**Tues., Oct. 8	due by 10pm	online
Term Test*	30%	25%	Tues., Oct. 22 (L0101)	10:10-11:40	TBA
			Tues., Oct. 22 (L0201)	15:10-16:40	TBA
			Wed., Oct. 23 (L0301)	14:10-15:40	TBA
Assignment 3	8%	8%	**Mon., Nov. 25	due by 10pm	online
Final Exam	52%	52%	Between Dec. 7-20	(3 hours)	see FAS

**Assignments due dates are subject to change. Graduate students will be evaluated at the graduate level based on the following scheme and according to the University Assessment and Grading Practices Policy. Graduate students should contact me regarding any alternative adjustment to the marking scheme.

	Scheme 1*	Scheme 2*	Date	Time	Location
Participation*	0%	10%	Sept. 19- Dec. 4		in class/online
Assignment 1	5%	5%	**Mon., Sept. 23	due by 10pm	online
Assignment 2	10%	10%	**Tues., Oct. 8	due by 10pm	online
Term Test*	30%	20%	Tues., Oct. 22 8(L0101)	10:10-11:40	TBA
			Tues., Oct. 22 (L0201)	15:10-16:40	TBA
			Wed., Oct. 23 (L0301)	14:10-15:40	TBA
Assignment 3	15%	15%	**Mon., Nov. 25	due by 10pm	online
Final Exam	40%	40%	Between Dec. 6-20	(3 hours)	see FAS

*Students will have the option to participate in quizzes/surveys/ worksheets in class. Roughly, participation quizzes will be held on a weekly basis and will count during the period September 19 to December 4. They are to foster student engagement and provide formative feedback on understanding of the course materials. Please note that participation is optional. Final grades will be the better performance of Scheme 1 and Scheme 2.

Crowdmark is a management system that efficiently facilitates large-scale assessment. It will be used in this class to take test attendance and for grading and returning course work electronically.

The assignments will each be a short data analysis project for which you will use R. Assignments are to be submitted online into Crowdmark by 10pm on the due dates. Late assignments will be accepted but subject to a 20% penalty per day late. Late submissions will not be allowed beyond 48 hours of the due date. Since assignments are short and the class is given more than a week to complete them, there are no additional accommodations for late or missed assignments.

The test will be written in locations to be announced (TBA). You will need to know basic `R` syntax on the test and exam, however emphasis will be on how to interpret output from `R`.

Both the midterm test and final exam will be closed-book and closed-notes; relevant formulas and tables will be provided. You will need a basic, non-programmable, scientific calculator. Calculators on phones and other devices equipped with remote access will not be permitted during the term test or final exam.

Missed Test Policy
If the test is missed for a valid reason, you must submit proper (original, complete and appropriate) documentation **in person within one week of the test to the instructor**. If documentation is not received in time, your test mark will be zero. If the documentation is validated, an online makeup test will be arranged within 2 weeks of the original test date. The makeup test will be worth 6% and the remaining weight will be shift to the final exam.

Re-grading Policy
Any requests to have marked work re-evaluated must be made in writing (that is, by email or brought to my office) within one week of the date the work was returned to the class. The request must contain a justification for consideration. Be sure to include your official name and student number for identification purposes. Electronic requests must be sent to regradesta302@gmail.com. Regrading requests should be processed by the teaching team within two weeks of the request date.

Missed Final Exam Policy
If the final exam is missed, students will be evaluated according to Scheme 1.

Computing
We will use `R` and `RStudio` for statistical computing. The main advantage of `R` is that it is a freeware and there is a lot of available help resources online. R is available for download at http://cran.r-project.org/ for Windows, Mac and Linux operating systems. `RStudio` is a good integrated development environment to `R` and can also be downloaded for free at https://www.rstudio.com/products/rstudio/download/.
Support for downloading and learning `R` (and `RStudio`) will be provided by the teaching team (Instructor and TAs). Additional resources will be given in our website. In lectures, examples with `R` syntax will be provided, which should be sufficient for you to do your assignments.

For assignments, it is preferred that you submit a PDF document, compiled using `R` and Word or, `RStudio`

and `RMarkdown` To learn more about `RMarkdown`, refer to

<div align="center">

http://rmarkdown.rstudio.com/index.html

</div>

Course website
The course website is available through Quercus via

<div align="center">

https://q.utoronto.ca

</div>

and will be used to post lecture notes, practice problems, quizzes, assignments, announcements and grades.

Online discussion forum
This term you will have the option to use Piazza for class discussion. This is a collective discussion forum for sections- L0101, L0201 and L0301. To sign up for the discussion forum go to the link:

<div align="center">

http://piazza.com/utoronto.ca/fall2019/sta3021001

</div>

If you decide not to use Piazza it will not disadvantage you in any way, and will not affect official University outcomes (e.g., grades and learning opportunities). If you choose not to opt-into Piazza then you can ask questions or discuss course material with the instructor or TAs during office hours.

Be sure to read Piazza's Privacy Policy and Terms of Use carefully. Take time to understand and be comfortable with what they say. They provide for substantial sharing and disclosure of your personal information held by Piazza, which affects your privacy. If you decide to participate in Piazza, only provide content that you are comfortable sharing under the terms of the Privacy Policy and Terms of Use.

The Piazza system is highly catered to getting you course material help fast and efficiently from classmates, the TAs, and the instructor. Rather than emailing questions, I encourage you to post your questions on Piazza. Moderation of the forum is subject to TA availability. Further details will be announced later. Please use the forum in accordance to its purpose. Inappropriate posts will not be tolerated and will be dealt with accordingly.

Accessibility Needs
The University of Toronto is committed to accessibility. If you require accommodations for a disability, or have any accessibility concerns about the course, the classroom, or course materials, please contact Accessibility Services as soon as possible at accessibility.services@utoronto.ca or https://www.studentlife.utoronto.ca/as.

Intellectual Property
Course materials provided on Quercus, such as lecture slides, assignments, tests and solutions are the intellectual property of your instructor and are for the use of students currently enrolled in this course only. **Providing course materials to any person or company outside of the course is unauthorized use.**

Videotaping of lectures will not be permitted under any circumstances. Permission to audio record lectures must be sought in advance. Failure to comply can result in legal action against all parties involved.

Academic Integrity
You are responsible for knowing the content of the University of Toronto's Code of Behaviour on Academic Matters at http://www.artsci.utoronto.ca/osai/students. 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 a few guidelines that apply to assignments:

- It is legitimate to discuss assignment problems with other students in the class and learn from each other.

- However, instructions should be strictly followed and assignments must be written up completely by yourself.

- Do not let other students read your completed assignment solutions as this can lead to copying.

- 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 assignments. As soon as you get an outside person to actually start working on one of your assignments, you have committed an academic offence!

- In general, you are free to discuss homework with each other, though all the work you turn in must be your own; you must not copy mathematical derivations, computer output and input, or written descriptions from anyone or anywhere else, without reporting the source within your work. (This includes copying from solutions provided to previous semesters' of the course.) Unacknowledged copying or unauthorized collaboration will lead to severe disciplinary action, beginning with an automatic grade of zero for all involved and escalating from there. Please read the UT Policy on Cheating and Plagiarism, and don't plagiarize.

Failure to comply with these guidelines is a **serious** academic offence; offences will be investigated and dealt with accordingly.

Communication

E-mail is appropriate for emergencies or private matters. Use your *utoronto.ca account. You will not get a response if you email from other email addresses. Write a proper email including the course number, section number, a subject in the subject line. The email should contain the addressee, your official name and UTORid for identification purposes. I will generally answer e-mail within two business days.

Announcements will be posted on Quercus. Please check there regularly. If an urgent matter arises, I may contact the entire class by e-mail. In order to receive these messages, ensure that your account is connected to your *utoronto.ca email and email notifications are enabled.

The TAs and instructor are here to help you. Ask questions and let me know promptly if there are any concerns.

Need additional help?

Here are some options:

- Visit the instructor and/or teaching assistants during office hours.

- Post questions on the discussion forum.

- Email the instructor in cases of emergencies or personal matters.

Your responsibilities

It is your responsibility to be an active learner by staying connected via our course website, being involved in classroom sessions and diligently completing assessments. Class times will be designed to actively engage you in the course material. In order for these sessions to be effective, coming prepared, by learning about the week's concepts through the textbook, is essential.