



Statistical Sciences
UNIVERSITY OF TORONTO

STA 303H1S / 1002HS - DATA ANALYSIS II

Winter 2019

January 8- April 4, 2019

Lectures: **SECTION L0101**
Tuesdays 10:10-12:00 and Thursdays 10:10-11:00 in **BA 1160**

SECTION L0201
Tuesdays 15:10-17:00 in **KP 108** and Thursdays 12:10-13:00 in **MP 202**
(except February 18-22 during Reading Week)
Campus map: <http://osm.utoronto.ca/maps.html>

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

Instructor: Dr. Shivon Sue-Chee
E-mail: shivon.sue.chee@utoronto.ca
Office: EP 104

Office hours: Tuesdays 12:10-13:00 in **EP 104** and Thursdays 11:00-12:00 in **BA 1160**
or by appointment (beginning Jan. 15)
More TA office/lab hours will be scheduled before test and assignments due dates.

Course content

The overall theme of this course is the use of linear models in situations where the assumptions of the multiple regression model developed in STA 302/1001 may not apply. The topics fall into two main categories:

- I. Categorical and count variables- including analysis of variance, logistic regression, Poisson regression, and log-linear models for contingency tables, and
- II. Correlated observations- including time series and repeated measures analysis

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, rather than the underlying theory. This course will also be an opportunity to continue 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 generalized linear models and
- to expand practical skills in statistical modelling and interpretation of results.

Pre-requisite

Students should have STA302/1001 or equivalent preparation. Students are also expected to have the mathematics pre- and co-requisites required by students in all courses leading up to STA302. This course is slightly less theoretical than STA302, but please do not attempt the course without the required mathematical background. **Pre-requisites will be strictly enforced by the department, not the instructor.** I am assuming that you have used R (and RStudio) previously, at the level used in STA 302/1001.

Textbook References

There is no required textbook for this course. The following is a comprehensive list of relevant textbooks.

- *Agresti: Categorical Data Analysis, 3rd edition* by Alan Agresti (Wiley)
Includes material on contingency tables, Poisson regression, log-linear models, and logistic regression. Chapters 2, 4, 5, and 6 contain material relevant to this course. (Will be on reserve at the Mathematics Library)
- *KNN: Applied Linear Regression Models, 4th edition* by Kutner, Nachtsheim, and Neter (Mc-Graw Hill).
Chapters 8, 11, 13 and 14 contain relevant material. (Will be on reserve at the Mathematics Library)
- *SJS: A Modern Approach to Regression with R* by Simon J. Sheather (Springer)
Chapter 8 (logistic regression), Chapter 9 (time series) and Chapter 10 (repeated measures). Available as an electronic resource through the University of Toronto library website.
- *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.
- *TSS: The Statistical Sleuth, 3rd edition* by Ramsey and Schafer (Brooks/Cole)

Course website

The course website is available through Quercus via

<https://q.utoronto.ca>

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

The discussion board will be open to all and you are encouraged to use it for course-related questions. The discussion board is designed to get help on course content fast and efficiently from classmates. The discussion will be moderated on a weekly basis by the instructor. Additional moderation will be subject to TA availability. Please use the forum in accordance to its purpose. Inappropriate posts will not be tolerated and will be 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.**

Evaluation

Undergraduate students will be evaluated based on the following schemes:

	Scheme 1*	Scheme 2*	Date	Time	Location
Quizzes*	0%	8%	Jan. 22- Apr. 4		in class/online
Assignment 1	4.5%	4.5%	Fri., Jan. 25	due by 10pm	online
Assignment 2	7.5%	7.5%	Fri., Feb. 15	due by 10pm	online
Term Test*	33%	25%	Thurs., Feb. 28 (L0101)	10:10-11:40	TBA
			Thurs., Feb. 28 (L0201)	11:10-12:40	TBA
Assignment 3	10%	10%	Thurs., Mar. 21	due by 10pm	online
Final Exam	45%	45%	Between Apr. 6-30	(3 hours)	see FAS (Feb. 28)

Graduate students will be evaluated at the graduate level based on the following scheme and according to the University Assessment and Grading Practices Policy.

	Scheme 1*	Scheme 2*	Date	Time	Location
Quizzes*	0%	10%	Jan. 22- Apr. 4		in class/online
Assignment 1	5%	5%	Fri., Jan. 25	due by 10pm	online
Assignment 2	10%	10%	Fri., Feb.15	due by 10pm	online
Term Test*	30%	20%	Thurs., Feb 28(L0101)	10:10-11:40	TBA
			Thurs., Feb 28 (L0201)	11:10-12:40	TBA
Assignment 3	15%	15%	Thurs., Mar. 21	due by 10pm	online
Final Exam	40%	40%	Between Apr. 6-30	(3 hours)	see FAS (Feb. 28)

*Students will have the option to participate in quizzes/surveys, in class or via Quercus. Roughly, participation quizzes will be held on a weekly basis and will count during the period January 22 to April 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.

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. Due to the flexibility in due dates, there are no accommodations for late or missed assignments.

Practice problems will be posted on the web. They are to help you prepare for the quizzes, test and exam and are not to be handed in.

The test will be written in locations to be announced (TBA). You will not need to know *R* syntax on the test and exam, but you will need 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, the weight of the test will be shifted 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 the course number, your official name and student number for identification purposes. Regrading requests will be processed within two weeks of the date the work was returned to the class.

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/>.

For assignments, it is preferred that you submit a PDF document, compiled using RStudio and RMarkdown or R and L^AT_EX. To learn more about RMarkdown, refer to

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

L^AT_EX is available as a free document preparation software at <https://www.latex-project.org/get/>.

Support for downloading and learning R (and RStudio) will be provided by the teaching team (Instructor and TAs). In lectures, examples with R syntax will be provided, which should be sufficient for you to do your assignments.

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

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!

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