

Principles of Statistics 2

University of Toronto
Department of Statistical Sciences
STA221H1S Winter 2021

Instructor:	Katherine Daignault	Synchronous Classes:	M 3-5PM EST
Email:	katherine.daignault@mail.utoronto.ca		W 3-4PM EST
Course webpage:	Quercus	Office Hours:	W 1:30-2:30PM EST

COURSE OVERVIEW

How will this course operate? This course will be offered entirely online, with synchronous lectures held during scheduled class times. Synchronous classes will occur through Bb Collaborate in Quercus and will be recorded and uploaded to Quercus later each week. It is your responsibility to make sure you are available during scheduled lecture times and/or stay on top of the course material and all relevant deadlines.

**** Please ensure that you have access to reliable internet service, as evaluations will take place online and there is no guarantee that accommodations can be provided for faulty internet during an evaluation.*

Course Description: A continuation of Principles of Statistics 1 (STA220) or other similar course. We will introduce some of the major methods in data analysis. The main topics we will cover are methods for count data, simple linear and multiple linear regression, analysis of variance for single-factor and multi-factor experimental designs, and nonparametric tests. There will be a strong focus on how to apply these concepts in practice.

Learning Outcomes: By the end of this course, all students should understand the concepts above, recognize when it is appropriate to apply such methods, as well as be comfortable implementing these with statistical software.

Pre-requisites: Pre-requisites are **strictly enforced by the department, not the instructor**. If you do not have the equivalent pre-requisites, you will be un-enrolled from the course. Students should have taken one of the following: STA220H1/STA288H1/PSY201H1/GGR270H1/EEB225H1/STAB22H3/STA220H5.

COURSE MATERIALS

Course Content: We have a Quercus course page where all lecture slides, recordings and materials will be posted. Further, any important announcements will also be posted in Quercus. Please make sure to check it regularly.

Textbook: We will be following *Stats: Data and Models, 3rd edition* by DeVeaux, Velleman, Bock, Vukov, and Wong. This course will cover Chapters 22-28, with some review topics taken from Chapters 7, 8, and 10. Suggested practice problems will come from this book.

Statistical Software: We will be using RStudio for performing statistical analyses. R is a free software that can either be downloaded onto your personal computer or used in the cloud. If you choose to work with R on your personal computer, then installation will be a two step process:

1. The base R framework is available for download at <http://cran.r-project.org/> for Windows, Mac and Linux operating systems.

2. Next, RStudio is a good integrated development environment to R (makes it simpler to work in R) and can also be downloaded for free at <https://www.rstudio.com/products/rstudio/download/>.

If you don't want to download the program or run into problems with installation, you may want to consider using RStudio through the [JupyterHub](#) for University of Toronto. This will allow you to login with your official UofT credentials and use RStudio without the need for a local installation. More information about using RStudio in JupyterHub will be provided in the first class. All R code and resources for assignments will be provided in lectures and on the course page.

COURSE COMPONENTS

Lectures: During lectures, we will cover important course materials, as well as cover a number of examples illustrating the uses of these methods. Lecture slides will contain some R code and output to show how to perform these methods in practice. Each week builds on the material from previous weeks, so it is recommended that you attend lectures regularly.

Office Hours: Instructor and TAs will hold office hours through Bb Collaborate in the Quercus course page. The office hour schedule will be posted on Quercus once finalized. It is recommended that you visit office hours whenever you have a question about the material. It is more important than ever in an online class to have material clarified as quickly as possible. Don't wait until the last minute to ask your questions!

Quercus Discussion Board: We will be using the Quercus Discussion Board as an online discussion forum. **All questions about course material should be posted here** or asked during TA/instructor office hours. The instructor and TAs will monitor the board and will help answer questions but students are encouraged to answer posts and help their fellow classmates.

Tutorials: Weekly R tutorial will begin the **week of January 18**. Tutorials are an important part of the course, as this is where you will experience working with statistical software. Each week, the TA will guide you through how to perform in R the statistical procedures we discuss during lectures. This will help with assignments and for certain questions on tests. With the online nature of the course, we will run only one tutorial each week, but anyone can attend on Bb Collaborate.

COMMUNICATION

How your instructor will communicate with you: All communication will be made through Quercus announcements or during lectures. Please ensure that you check Quercus regularly so you don't miss anything important.

Where to send content questions: We will be using the Quercus Discussion board to collect student questions regarding course content, assignments, etc. All questions should be posted here. The board will be organized by week to help keep it all organized.

When to email the instructor: The instructor will only respond to emails of a private or sensitive nature. If you email the instructor with content related questions, you will be asked to post your question on Quercus so the answer may benefit all students. Should you need to email the instructor, please use your official mail.utoronto.ca email, include your full name and student number in the text, and include "STA221" in the subject line.

A note on email and discussion board etiquette: Please make sure that you communicate politely and respectfully with all members of the teaching team and your fellow classmates. Written communications can sometimes take a tone other than what was intended (e.g. can come off as dismissive, rude or insulting), so make sure you re-read or read out loud your email/post before sending it to make sure it

has the tone you intended. For more tips on respectful communication, see [professional communication tips](#).

GRADING SCHEME

Students will be graded according to the below grading scheme:

Assessment	Date Due/Occurring	Scheme 1
Assignments (× 4)	Friday Jan. 29 by 23:59EST	10%
	Friday Feb. 12 by 23:59EST	10%
	Friday March 5 by 23:59EST	10%
	Friday April 9 by 23:59EST	10%
Term Test	Monday March 8 from 3-5PM EST	20%
Data Analysis Project	Wed. March 24 by 23:59EST	20%
Final Assessment	Scheduled by FAS between April 13-23	20%

Please note that the last day to drop the course without penalty is March 15, 2021.

MINIMUM PASSING REQUIREMENT

In order for the instructor to be able to reasonably assess the ability of each student with the course material, a minimum amount of work must be submitted to provide enough evidence of proficiency. To this end, students must submit the following assessments in order to be considered for passing the course: **term test, the project, and the final assessment**. If a student fails to submit the minimum passing requirement, they may not be eligible to pass the course as not enough work has been completed for a meaningful grade to be awarded.

EVALUATION BREAKDOWN

Assignments: There will be 4 assignments that must be completed **individually**. These are opportunities to practice and receive feedback on applied problems, as well as working with statistical software.

- Since assignments are to be done individually, it is not appropriate to post questions about how to do or approach assignment questions on the discussion board. However, you may ask for clarification, or general content or R questions.
- Assignments will be submitted through Crowdmark, meaning you will need to upload PDF, PNG or JPEG versions of your assignment answers. Instructions will be provided on the Quercus assignment page when posted.
- Punctuality is key to keep the course moving for everyone. There will be a 10% penalty for each day that the assignment is late (e.g. if you submit 15 minutes, 2 hours or 19 hours late, these all receive a 10% late penalty). No assignments will be accepted 48 hours past the due date.
- You will be given 1.5-2 weeks to complete each assignment which should be plenty of time to complete the assessment. In general, no extensions will be granted unless under extreme and unusual circumstances that will be assessed on a case by case basis. If you think you fall into this extreme circumstance category, then you must email the instructor no later than 24 hours BEFORE the due date.

Term Test: The term test will be conducted online. The test will be two parts: the first part will be multiple choice and will be 50 minutes in duration on Quercus, while the second part will be short answer questions and will be 40 minutes in duration with an additional 20 minutes to upload solutions to Crowdmark. More details on submissions will be communicated closer to the test date. The term test will take place on **Monday March 8 from 3-5PM EST in scheduled class time**. The term test will cover material from Weeks 1-7 inclusive.

Data Analysis Project: The final project will be due on **March 24, 2021** and will consist of a data analysis involving regression on a dataset which must be completed individually. Students will be required to demonstrate their understanding of the methods taught in lecture by developing a reasonable regression model that addresses the research question using the techniques taught in class. The students will be responsible for choosing how to build their regression model but they will need to justify their steps and decisions. The submission of this project will take the form of a pre-recorded video presentation of their results with appropriate visual aids.

Final Assessment: There will be a final assessment in this course taking the form of an online timed test (similar to the term test). The final assessment will occur during the final assessment period between **April 13-23** and will be scheduled by the Faculty of Arts and Science. The final assessment will be cumulative and will be 3 hours in duration. More details will be communicated at a later date.

REGRADE POLICY

Regrade requests will be accepted for the term test and the data analysis project only. These will be accepted no later than 1 week from the day that the graded work has been returned. There will be a Quercus survey set up at that time where regrade submissions can be made. All regrade requests will require a written justification of exactly where a grading error occurred and an explanation of how the work meets the rubric requirements. No regrade requests will be accepted by email.

MISSED ASSESSMENT POLICY

Missed Assignments: There are no accommodations for missed assignments or mini project. Extensions may be granted in extreme situations at the discretion of the instructor if received no later than 24 hours prior to the deadline. The mini project must be submitted as part of the minimum work requirement.

Missed Term Test:

- If a student missed the term test for a valid medical reason, and have **both** filled out the absence declaration form on ACORN **and** emailed the instructor with details and confirmation of completion of the absence declaration **within one week** of the test, then they have the opportunity to write a make-up test at a date and time scheduled by the instructor.
- If the test is missed for any other reason, prior approval must be obtained from the instructor in order to be eligible to write the make-up test. If approval to miss the term test has not been obtained before the test takes place, a grade of zero will be given.
- To meet the minimum work requirement for this course, you must write the make-up if you missed the term test for a valid and documented medical reason, otherwise a meaningful grade cannot be calculated for you to pass the course.
- NOTE: since the term test is online and all students receive the test link, if you write the test or look at the test questions, you forfeit your eligibility to write the make-up test. Therefore you must make your decision about whether you are well enough to write the test before the test has begun.

Missed Data Analysis Project: The project is a required component of the course and must be submitted to meet the minimum work requirement. The project will be released in the week after the reading break which is ample time to complete the assignment. Therefore, no accommodations will be provided for a missed project. Further, extensions will not be granted unless under extreme circumstances that have been communicated to the instructor no later than 2 days prior to the deadline.

Missed Final Assessment: The final assessment must be completed in order to meet the minimum work requirement to pass the course. In the event of illness, you must declare your absence using the absence declaration form on ACORN. Further details will be provided during the final assessment period.

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.** This includes providing materials to predatory tutoring companies.

ACADEMIC INTEGRITY

The University treats cases of plagiarism and cheating very seriously. It is the students' responsibility for knowing the content of the University of Toronto's [Code of Behaviour on Academic Matters](#). All suspected cases of academic dishonesty will be investigated following procedures outlined in the above document. If you have questions or concerns about what constitutes appropriate academic behaviour or appropriate research and citation methods, you are expected to seek out additional information on academic integrity from your instructor or from other institutional resources (see <http://academicintegrity.utoronto.ca/>). Here are a few guidelines regarding academic integrity:

- You may consult class notes/lecture slides during tests and projects, however sharing or discussing questions or answers with other students is an academic offence.
- Students must complete all assessments individually. Working together is not allowed.
- Paying anyone else to complete your assessments for you is academic misconduct.
- Sharing your answers/work/code with others is academic misconduct.
- Looking up solutions to test problems online or in textbooks and copying what you find is an academic offence.
- All work that you submit must be your own! You must not copy computer output and input, or written answers from anyone or anywhere else. 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 UofT Policy on Cheating and Plagiarism, and don't plagiarize.

ACCESSIBILITY NEEDS

The University of Toronto offers academic accommodations for students with disabilities. If you require accommodations, or have any accessibility concerns about the course, the classroom, or course materials, please contact Accessibility Services as soon as possible: accessibility.services@utoronto.ca or <http://accessibility.utoronto.ca>.

CLASS SCHEDULE - TENTATIVE

Below is a tentative schedule and list of topics to be covered in class. The content corresponds to Chapters 22-28 in the textbook, with occasional review from earlier chapters. The instructor reserves the right to modify this schedule as needed due to time constraints.

Week	Content	Textbook
1	Introduction: working with R and RStudio, review of relevant STA220 material	
2	Comparing counts: tests for goodness of fit, homogeneity and independence	Chapter 22
3	Simple linear regression Part 1: linear model, least squares approach, assumptions, residuals, correlation/coefficient of determination	Chapter 7/8, 23.1
Jan. 29	Assignment 1 due by 11:59PM EST on Crowdmark	
4	Simple linear regression Part 2: inference about slope and intercept, predictions, ANOVA for regression	Chapter 7/8, 23.2-6
5	Multiple linear regression Part 1: multiple regression models, interpreting regression parameters, assumptions	Chapter 26.1-3
Feb. 12	Assignment 2 due by 11:59PM EST on Crowdmark	
READING BREAK		
6	Multiple linear regression Part 2: inference on single coefficient, inference on all (or set of) coefficients, adjusted r-squared	Chapter 26.4-5
7	Multiple linear regression Part 3: diagnostics for regression models, variable selection, collinearity	Chapter 27.2-3
Mar. 5	Assignment 3 due by 11:59PM EST on Crowdmark	
8	Multiple linear regression Part 4: using indicators and interactions	Chapter 27.1, 24.1
Mar. 8	Term Test in class from 3-5PM EST	
9	ANOVA Part 1: Review on experiments, single factor comparison of means, F statistic and tables, the model, assumptions,	Chapter 10.3-5, 24.1-3
10	ANOVA Part 2: Testing pairs, Bonferroni multiple comparisons, two-factor model, assumptions and inner workings,	Chapter 24.4-5, 25.1-2
Mar. 24	Data Analysis Project due by 11:59PM EST	
11	ANOVA Part 3: ANOVA involving interactions Nonparametric tests: Wilcoxon rank sum, Kruskal-Wallis,	Chapter 25.3, 28.1-2
12	Nonparametric tests: rank test for paired data, rank correlation test	Chapter 28.3-5
Apr. 9	Assignment 4 due by 11:59PM EST on Crowdmark	