

The Practice of Statistics 2

University of Toronto
Department of Statistical Sciences
STA221H1-S

Instructor:	Katherine Daignault	Class Location:	SS 2135
Email:	katherine.daignault@mail.utoronto.ca	Class Day/Time:	Mon. 3-5pm, Wed. 3-4pm
Office:	HS 376 (155 College St.)	Tutorials:	Wed. 4-5pm
Office Hours:	Tues. 3:30-5	Course webpages:	Quercus and Piazza

COURSE OVERVIEW

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.

COURSE MATERIALS

Course Content: All lecture slides and materials, as well as tutorial materials will be posted on the Quercus course page. Further, any important announcements will also be posted in Quercus. Please make sure to check it regularly so you don't miss anything.

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.

Calculators: You will need a basic (non-programmable, non-graphing) calculator for tests and exams. Phone calculators or any other devices that permit communication or access to Wifi are **NOT** allowed during tests or exams.

Statistical Software: We will be using RStudio (the free version) for performing basic statistical analysis. We will go through how to install this in class, but in general:

- First, download R (a free statistical package) from <https://www.r-project.org/>
- Then download RStudio (the free version) from <https://www.rstudio.com/>

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.

Tutorials: Weekly tutorials will begin the **week of January 13**. All tutorials are on Wednesdays from 4-5pm but located in different rooms. Check the [course timetable](#) for more information. Tutorials are an

important part of the course, as this is where you will experience working with statistical software. Each week, the TAs 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 and exams. **Please make sure that you are registered in one of the tutorial sections.**

Office Hours: Each TA will hold regular weekly office hours in HS 381. The office hour schedule will be posted on Quercus when finalized. In addition, the instructor will hold weekly office hours. It is recommended that you visit office hours whenever you have a question about the material or assignments. It is best not to leave your questions to the last minute.

GRADING SCHEME

Your final grade for the course will be determined as the **higher** of the two grading schemes below:

Assessment	Date Due/Occurring	Scheme 1	Scheme 2
Class Participation	Ongoing	5%	0%
Assignment # 1	Friday January 24 by 23:59	5%	5%
Assignment # 2	Friday February 7 by 23:59	7%	7%
Midterm	Monday March 2	25%	30%
Assignment # 3	Friday March 20 by 23:59	7%	7%
Assignment # 4	Friday April 3 by 23:59	6%	6%
Final Exam	Scheduled by FAS	45%	45%

Please note that the last day to drop the course without penalty is **March 15, 2019**.

EVALUATION BREAKDOWN

Class Participation: During lectures, I will be posing some questions to you through a software called Poll Everywhere (<https://PollEv.com/katherinedai702/register>) to see if you are following along.

- You do not need to answer the questions correctly in order for you to receive participation marks, but it would be helpful to me and you if you try to figure out the right answer, so we can both see how comfortable you are with the material. **Participation is optional** - if you choose not to participate, your final grade will come from Scheme 2 only.
- To use the software and get credit for participating, you will need to register using your official **University of Toronto email address**. (more details during the first week of class)

Assignments: There will be 4 small assignments throughout the term. They will involve working with the material presented in lectures, but will also have some parts that will require you to answer using R.

- Assignments can be done in groups of two, and only one copy needs to be submitted per group. It is important that **both students in the group** attempt all questions on the assignment, since you will need to understand how to answer these types of questions for the midterm and final exam.
- Since you are allowed to work in pairs, it will not be appropriate to post questions about the assignments on Piazza. However, generic lecture content or R questions can be posted on Piazza.
- We will be using Crowdmark to collect and grade the assignments - **you will need to upload each assignment to the Crowdmark system by 11:59PM (i.e. 23:59) on the date it is due** - instructions on how to do this will be given with each assignment, and will also be posted on Quercus.

- In general, extensions for assignments will not be granted unless a valid reason is provided. These will be assessed on a case by case basis.
- There will be a 10% penalty for each day that an assignment is late (e.g. if you submit your assignment on Saturday instead of Friday, you will lose 10% automatically; if you submit on Sunday, you lose 20%, ...).

Midterm exam: The midterm will be held **during the regular scheduled lecture time** during Week 8. It will be approximately 1.5 hours long and will cover all lecture materials from Week 1 up to and including Week 6.

Final exam: The final exam will be a 3 hour cumulative exam and will occur during the exam period (April 6-25). The date of the final will be communicated through Quercus as soon as it is scheduled.

MIDTERM/FINAL EXAM INFORMATION

Aids: Both midterm and final exam are closed-book, however you will be allowed a one single-sided 8.5×11 inch handwritten aid sheet for the midterm, and one double-sided 8.5×11 inch handwritten aid sheet for the final exam. These are useful studying tools, so it is recommended that students spend some time on creating and modifying these throughout the term.

Grading/Regrading: Assignments and tests will be graded using Crowdmark. This will allow students to obtain feedback quicker and without the need to hand back paper copies. Regrading requests will only be considered for the midterm and must be made in writing within one week of the grade being released. All regrade requests must provide a justification in order to be considered.

Missed Midterm: There are no make-up tests. If the midterm is missed for a valid reason, the weight of the assessment will be moved to the final exam. Such situations include:

- a valid medical reason: the student must submit a University of Toronto [Verification of Student Illness or Injury form](#) to the instructor in person within one week of the missed test. The form will only be accepted if it is the original form, completed as per the instructions, and indicates the degree of incapacitation on academic functioning. Forms indicating a negligible or mild degree of incapacitation will not be considered a valid medical reason.
- other valid reason (e.g. death in the family): the student must obtain prior approval to miss the midterm from the instructor, with supporting documentation as applicable.
- If no valid reason is provided, or prior approval of absence is not obtained, the midterm will receive a grade of 0.

COMMUNICATION

Please do not email the instructor with questions related to the content of the course. It is much easier to have these types of questions answered by coming to either the instructor or TA office hours. Alternatively, you can post them on Piazza and see if your fellow students can help answer your question. Email is appropriate if you have a personal or emergency matter to discuss. If you have to email the instructor, please include “STA221” in your subject line, your student number in the body of the email, and send it from your Utoronto email.

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/>).

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 (Assignment 1 due - Friday Jan. 26 by 11:59PM)	Chapter 7/8, 23.1
4	Simple linear regression Part 2: inference about slope and intercept, predictions, ANOVA for regression, full R example	Chapter 7/8, 23.2-6
5	Multiple linear regression Part 1: multiple regression models, interpreting regression parameters, assumptions, inference on single coefficient (Assignment 2 due - Friday Feb. 7 by 11:59PM)	Chapter 26.1-4
6	Multiple linear regression Part 2: inference on all (or set of) coefficients, adjusted r-squared, diagnostics for regression models	Chapter 26.4-5, 27.2
READING BREAK		
7	Multiple linear regression Part 3: Using indicators and interactions, variable selection, collinearity and regression uses	Chapter 27.1, 27.3
8	Multiple linear regression Part 4: Full example of multiple linear regression in R (Midterm in class - Monday Mar. 2)	
9	Experimental design and testing of means: fundamentals of design, factors and blocking, single factor comparison of means, F statistic and tables	Chapter 10.3-5, 24.1
10	Single factor ANOVA: the model, assumptions, Bonferroni multiple comparisons, ANOVA on observational data (Assignment 3 due - Friday Mar. 20 by 11:59PM)	Chapter 24.2-5
11	Multi-factor ANOVA: two-factor model, assumptions and inner workings, ANOVA involving interactions	Chapter 25
12	Nonparametric tests: Wilcoxon rank sum, Kruskal-Wallis, rank test for paired data, Friedman test, rank correlation (Assignment 4 due - Friday Apr. 3 by 11:59PM)	Chapter 28