UNIVERSITY OF TORONTO, DEPARTMENT OF STATISTICAL SCIENCES

STA303/1002

METHODS OF DATA ANALYSIS II

Course information

2022

HALF YEAR, HALF CREDIT

The course focuses on using and interpreting advanced statistical methods with applications in a number of different areas. The overall theme of this course is dealing with situations where the assumptions of the regression models developed in STA302 may not apply. The course is a mixture of theory and application. Assignments will involve computing with R and there is a significant focus on written and oral communication.

Class times

Class: Wednesdays
L0101: 10:00 a.m.–12:00 p.m. ET
L0201: 3:00—5:00 p.m. ET

Tutorials: *Alternating* Thursdays
12:00—1:00 p.m. ET or
5:00—6:00 p.m. ET

Delivery

- Wednesday classes will be **online** and recorded.
- The Thursday tutorial time will alternate between group activities and TA office hours. They will be **online** for January, and an optional in-person stream for the fortnightly group activities is planned after that.
- **Synchronous attendance** is *recommended* but not required to complete the course.

All **materials** will be posted on Quercus and/or the course guide, where you can also find a plain text version of this syllabus. **Course discussion board:** Piazza

Teaching team

Instructor: Liza Bolton
Please call me: Liza or Prof. Bolton/Prof. B
How do you pronounce that?

- Liza: a video...
- Bolton: like the words “bowl” + “tonne”

Pronouns: she/her
Email: sta303@utoronto.ca
Office hours: 2nd half of Wednesday classes

Head TA: Amin Banihashemi (he/him)

Please call me: Amin (A-meen [æˈmiːn or əˈmiːn] [bæniːˈhɑːʃɛmiː] decode here)
Email: sta303@utoronto.ca

Teaching assistants

Vedant Choudhary, Shuang Di, Sonia Markes, Ian Richter, Xiaochuan Shi, Lei Sun, Liam Welsh, Dongyang Yang, Kevin Zhang, Robert Zimmerman

Office hours: Alternating Thursdays, see Quercus **Email:** sta303@utoronto.ca

Land acknowledgment

The land on which our University operates is the traditional lands of the Anishinaabe, the Haudenosaunee, and the Mississaugas of the Credit. With the Dish With One Spoon treaty, these peoples agreed to share and protect this land, and all those who have come here since, both Indigenous and non-Indigenous, are invited into this treaty in a spirit of respect and peace. This land is also, more recently, subject to Treaty 13, a treaty between the Mississaugas and the British Crown.

In this course, we are coming together to discuss statistics, a field that has been part of historical and ongoing colonization, oppression, and harm of Indigenous peoples. Let us remind ourselves of our responsibilities to this land, its original peoples, and to each other and work to be ethical and culturally competent practitioners in our chosen fields.

We encourage you to consider the history of the land wherever you are.
https://www.whose.land/en/

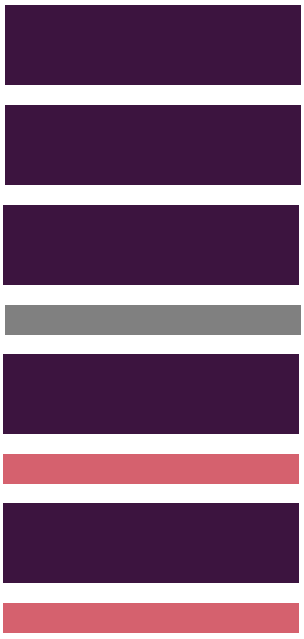
Prerequisites

STA302/1001

I.e., we will assume that you are familiar with running linear regression analyses, including checking assumptions and some of the mathematical reasoning behind the models. Material from the second-year statistical theory courses which are prerequisites to STA302 will be drawn on extensively. Knowledge of programming with R is essential.

Course format and organization

This course is composed of **five two-week modules** and **two assessment focus weeks.**



Each **module** has an "I show you, you show me" structure. It starts with you watching videos and/or doing readings and a guided code demonstration or other relevant class topics ("I show you"). This course is flipped, so the expectation is that you engage with some or all of the asynchronous content before Wednesday in the first week. Then you will do some or all of the following: a group problem-solving activity, practice quizzes, assessment activities.

During **assessment focus weeks**, no new content will be released.

Learning objectives

By the end of the course, you will be able to:

- **Wrangle** and **explore** a dataset
- Create appropriate data **visualizations**
- Describe **ethical considerations** in data analysis
- Understand the assumptions and appropriate use cases for **linear mixed models, generalized linear models, generalized linear mixed models,** and **generalized additive models**
- **Write** and **execute R code** for the model types covered
- Accurately and appropriately **interpret** the results of the model types covered and communicate these to a range of **audiences**

Textbooks

You do not have to purchase a textbook for this course. There are three texts that we will use extensively, and they are all freely available to you. Additional readings will be assigned as appropriate.

Wickham. R for Data Science. 2019.
Legler and Roback. Broadening Your Statistical Horizons. 2019.
Wood. Generalized Additive Models: An Introduction with R, 2nd Edition. 2017.
(requires you to log in with your UTORid)

Computing and minimum technical requirements

We will be using RStudio to make reproducible data analysis reports using R and R Markdown. You can use RStudio on your personal machine or through the U of T JupyterHub: jupyter.utoronto.ca.

To participate in synchronous classes and office hours you will need a **U of T Zoom account.** If you do not yet have one, go to https://utoronto.zoom.us/ to set one up. To participate fully, you will need Desktop client or mobile app: version 5.3.0 or higher or ChromeOS: version 5.0.0 (4241.1207) or higher. You can check your desktop client or mobile app version by following these instructions.

All students should consult the minimum technical requirements for participation in online learning. If you are facing financial barriers to obtaining the required technology, please contact your College Registrar's Office to obtain information regarding your potential eligibility for a need-based bursary.

Course outline

The topics listed below are subject to change.

JAN 10–21 MODULE 1: WELCOME!

- How this course works, assessments, admin, upward management
- Common tests as linear regression and linear regression recap
- Reading strategy: previewing and skimming
- [Optional] Get ahead for module 2 if unfamiliar with R Markdown, ggplot and dplyr

JAN 24–FEB 4 MODULE 2: PROFESSIONAL SKILLS FOR DATA ANALYSIS

- Data visualization
- Data cleaning, merging, and exploration
- Statistical communication
- Ethical professional practice

FEB 7–FEB 18 MODULE 3: LINEAR MIXED MODELS

- Identifying correlated data
- Fixed and random effects
- Fitting, visualizing and interpreting correlated data
- Likelihood ratio tests

FEB 21–25 READING WEEK: NO CLASS!

FEB 28–MAR 11 MODULE 4: GENERALIZED LINEAR MODELS

- Theory for GLMs
- Fitting models, checking assumptions and interpreting logistic and Poisson regression

MAR 14–18 MIXED ASSESSMENT FOCUS WEEK

MAR 21–APR 1 MODULE 5: GLMMS AND GAMS

- Generalized linear mixed models
- Generalized additive mixed models

APR 4–8 PROJECT FOCUS WEEK

Important dates!

JAN 21 Prerequisite knowledge check due
JAN 23 Last day to enrol in S courses

FEB 1 Lunar New Year
FEB 2 Prerequisite check workshop
FEB 3 Professional dev proposal due
FEB 3 Mini-portfolio due

FEB 14 Valentine's Day
FEB 17 Portfolio due

FEB 21 Family Day (U of T closed)

MAR 9 Mini-mixed assessment (12 hour window, 8 a.m.–8 p.m.)

MAR 13 Daylight savings begins
MAR 14 Last day to cancel course
MAR 16 Mixed assessment (12 hour window, 8 a.m.–8 p.m.)

MAR 31 Professional dev evidence and reflection due

APR 7 Final project due (bonus)
APR 11 Project (no bonus)

Assessments

There are two special elements of assessment/grade calculation in this course that are important to be aware of in planning your approach to it.

- *Two roads diverged in a yellow wood*[1]: You can opt for Path A or Path B to get your final mark. I will calculate your marks under both paths and then assign you the higher of the two as your final mark. (Grad students have a third option).
- A-tisket, a-tasket[2], fill up your knowledge basket[3]:
  - Personalize which of these assessments you do based on your interests and skills you want to develop.
  - You can 'max out' your basket: just keep putting grades in until you get to 5%.

Assessment	Path A	Path B	STA1002 ONLY	Due date
Mini-portfolio	5	0	0	Feb 3
Portfolio	20	25	25	Feb 17
Mini-mixed assessment	5	0	0	Mar 9
Mixed assessment	20	25	25	Mar 16
Final project	45	45	50	Apr 7 (2% pt bonus) | Apr 11 (no bonus)
Knowledge basket	5	5	0	Multiple

12-hour assessment windows, 8:00 a.m.–8:00 p.m. ET

Everything* in STA303 is due at 3:03 p.m. ET

**Mixed assessment windows are 8-8, I couldn't make the 303 work out fairly.*
Note: All times in this course are in Eastern Time (Toronto time). Please note that daylight savings will begin in Canada on March 13. If you are not based in Canada this may change the time conversion for you. Please keep this in mind.

Knowledge basket

The following components are *guaranteed* knowledge basket assessment options. Additional opportunities may be offered throughout the course (Team Up!, speaker series reflections, etc.)

Assessment	%	Due date
Pre-knowledge check: completion	0.5	Jan 20
Pre-knowledge check: 80%+ or workshop	0.5	Jan 20 (80%+ score) | Feb 2 (workshop)
Professional development task proposal	1	Feb 3
Professional development task evidence & reflection	3	Mar 31
Writing & peer review (Create-Assess-Reflect) x 5	0.5 x 5	Friday-Tuesday-Friday each Module
Module check-in x 5	0.1 x 5	Last Friday of each Module

Hours expectations

While everyone has different work styles and learning needs, I want to provide some guidance around how I expect this course to look for you.

Plan to be doing 6–8 hours of work on STA303 each *week*. In a **two-week module**, this may look like:

- 2–4 hours on videos and readings
- 1–3 hours of attending synchronous class or reviewing the recording and activities
- 1–2 hours on knowledge basket assessments
- 2–6 hours on other assessments
- Remaining time attending reading announcements, office hours, checking Piazza, revision, etc.

Marking concerns / regrade requests

Any request to have an assessment remarked must be submitted to the appropriate form on the Forms page on Quercus under the following conditions:

- **Wait** 24 hours after the release of grades. Use this time to go over any sample solutions and feedback, the instructions, and relevant course materials.
- *After* the 24-hour period has finished, you will have one week to submit your regrade request. (I.e., one week + 24 hours total.)

- Your request must include a detailed and thoughtful justification referring to your answer and the relevant course material to be considered. Please note that I reserve the right to review the grading of all questions or parts when you re-submit an assessment for reconsideration (i.e., your grade could go down).
- You will receive a confirmation email upon submitting the form. Allow for two weeks for processing after the request window closes before following up.
- The specific timeline and requirements for **final project regrade requests** will be announced later.

Only answers in English (or appropriate code, mathematical symbols) can be accepted in this course. Answers submitted in a different language will receive a 0 and will not be eligible for regrading. If you have an autotranslation extension on your browser, be very careful about how this can interact with Quercus.

Please note that I reserve the right to review the grading of all questions or parts when you re-submit an assessment for reconsideration (i.e., your grade could go down).

Missed work policies

In general, late work is not accepted, without either:

- an extension approved **48 hours** before the due date, or
- a personal illness/emergency declaration no more than 3 days after the due date.

Please note that technical difficulties knitting an Rmd or getting the due time wrong *do **not** constitute personal emergencies*.

The following assessments are eligible:

- Mini-portfolio
- Portfolio
- Mini-mixed assessment
- Mixed assessment
- Professional development proposal
- Professional development evidence and reflection

Upon receipt of your request, we will contact you via email within 2 business days to confirm an accommodation, as appropriate.

EXCEPTIONS

- **Knowledge basket** assessments (other than the professional development task) are not eligible for extensions.
- There are *no* routine extensions granted for the **final project.** In exceptional circumstances, you can work with your College Registrar and me on this.

IMPORTANT NOTES

1. If too much work is missed, even for valid reasons, an **oral exam** may be required to calculate a fair mark, at the discretion of the instructor. Please ensure you and/or your College Registrar get in touch with me as early as possible if this may be the case for you.
2. If you have accommodation letters from an accessibility advisor, make sure you read the instructions in the 'Accommodations and accessibility' section.
3. Unless discussed with your instructor first and an agreement is come to, if you submit an assessment, it will be assumed that you deemed yourself fit enough to do so and your grade will stand as calculated. No accommodation will be made based on claims of medical, physical, or emotional distress after the fact.

Communication policy

AKA HOW TO GET YOUR QUESTIONS ANSWERED

Following our course communication policy helps ensure you receive answers and supports in a timely fashion, and also shows respect for the teaching team's time and effort. We reserve the right to ignore any correspondence that does not conform to this policy.

Course logistics? e.g.,
- What is the deadline for the final project?
- Where do I submit the assignment?

Course content? e.g.,
- Why do we sometimes use glm() and sometimes use glmer()?
- My code won't run for question #1

Info to share with class? e.g.,
- I have a link/resource/opportunity to share with my classmates

PIAZZA FORUM

- Link in Quercus's course navigation menu.
- Posts can be anonymous for your classmates, but instructors and TAs will be able to see your name.
- Before posting a question, double-check the syllabus AND search to see if someone else has already asked a similar question (you can edit the question to add yours or post a follow-up at the bottom).
- Try to answer your classmates' questions—this is a great way to reinforce your own understanding while also helping your classmates! Don't worry if you aren't 100% sure of the answer—answers will be reviewed/endorsed/completed by the teaching team!

Missed an assessment due to illness or personal emergency?

Want to request a regrade of an assessment?

FORMS

- Use the appropriate form linked on the Forms page on Quercus.
- If you cannot meet a deadline because you are ill, please also refer to the **Missed Work** section in this syllabus. A doctor's note is not required, but if you have one you can upload it as supporting documentation.
- If you wish to request a regrade, please also refer to the **Marking Concerns** section in this syllabus. Be prepared to provide a detailed justification and possible supplementary materials. "I worked hard on this so I should get a better mark" is not an appropriate justification (yes, I do receive emails like that).

Personal/sensitive circumstances? (i.e., something which is not appropriate to share with the whole class)

COURSE EMAIL: sta303@utoronto.ca

- Send emails from your utoronto.ca email address to ensure they don't automatically go to a Junk folder.
- Include your full name and UTORid.
- This account will be monitored by the head TA and course instructor; if you want to reach Prof. Bolton *only* please include [Prof. Bolton] in the **subject line**; do not email her directly about course matters.
- Allow at least 24 hours for a response during the week (Monday to Friday, ET) and do not expect responses on the weekend. Do not send a follow-up email until at least two business days (Toronto time) later.

NEVER send Quercus mail to the STA303 teaching team
NEVER use the 'Add Comment' option on Quercus. They will not read.

Please!

Accommodations and accessibility

If you have an accommodation letter from your accessibility advisor that is relevant to this course, please do the following:

- Email your letter to sta303@utoronto.ca with "Accommodation letter" as part of the email subject, CC your advisor and let us know anything else you wish us to know/any questions you have. Please do this as soon as possible after you enroll in the course/receive this syllabus.
- Confirm any accommodations for each specific assessment **one week** before the assessment. (I.e., if you receive extra time for timed assessments, confirm this one week prior to the mixed assessments, even if we have already discussed this at the beginning of the semester.)

ACCESSIBILITY SERVICES

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

RELIGIOUS ACCOMMODATION

At the University of Toronto, we are part of a diverse community of students, staff, and faculty from a wide range of cultural and religious traditions. For this course, I have sought to avoid scheduling compulsory activities in ways that will clash with religious holy days (not captured by statutory holidays). If you anticipate missing a course activity due to a religious observance, please let me know as early in the course as possible. With sufficient notice—ideally at least three weeks—we can work together to make alternate arrangements.

RECOGNIZED STUDY GROUPS

I would highly recommend you get involved with an RSG. RSGs are small study groups of 3 to 6 students from the same course who meet weekly to learn course content in a collaborative environment.

MEET TO COMPLETE

Meet to Complete is an online "study with me" space where you can study alongside other students. To join Meet to Complete, enroll in the Meet to Complete course on Quercus. Learning, even online, doesn't need to be lonely!

Feeling distressed?

You may find yourself feeling overwhelmed, depressed, or anxious. Lots of people feel the same way. There is help available from mental health professionals 24 hours a day via online and phone-based services listed on this page in the course guide, as well as a range of other helpful U of T and community resources.

Accessibility Services (see above) also provides supports for mental health concerns.

Intellectual property statement

Course material that has been created by your instructor (i.e., lecture slides, questions/solutions, and any other course material and resources made available to you) is the intellectual property of your instructor (or the credited holder of the copyright) and is made available to you for your personal use in this course. Sharing, posting, selling or using this material outside of your personal use in this course is not permitted under any circumstances and is considered an infringement of intellectual property rights. If you would like to record any course activities in this course, you MUST ask permission from your instructor in advance. According to intellectual property laws, not asking permission constitutes stealing.

Academic integrity

PLAGIARISM

You may be at risk of plagiarizing if you do not understand the rules and your responsibilities. You must not present the work of others as your own. This includes, but is certainly not limited to, copying text and including it in your writing without a citation and quotation marks.

There are many resources to help you learn more:

- https://www.academicintegrity.utoronto.ca/perils-and-pitfalls/
- https://www.academicintegrity.utoronto.ca/smart-strategies/
- This video will be assigned later in the course.

YOU are responsible for knowing the content of the University of Toronto's Code of Behaviour on Academic Matters. The University of Toronto treats cases of academic misconduct very seriously. Academic integrity is a fundamental value of learning and scholarship at the U of T. Participating honestly, respectfully, responsibly, and fairly in this academic community ensures that your U of T degree is valued and respected as a true signifier of your individual academic achievement.

Other potential offences include, but are not limited to:

- Looking at someone else's answers.
- Letting someone else look at your answers.
- Misrepresenting your identity.
- Falsifying or altering any documentation required by the University.
- Falsifying institutional documents or grades.

All suspected cases of academic dishonesty will be investigated following the procedures outlined in the Code of Behaviour on Academic Matters. If you have any questions about what is or is not permitted in this course, please do not hesitate to contact me.

Academic integrity

SPECIFIC ADVICE ON UNTIMED ASSESSMENTS

As a general rule, for untimed assessments, I encourage you to discuss course material with each other and ask others for advice. However, it is ***not permitted*** to share R code or written answers for anything that is to be handed in. For example, "For question 2 what R function did you use?" is a fair question when discussing course material with others in the class; "Please show me your R code for question 2" is not an appropriate question.

If writing or code is discovered to match another student's submission or outside source, this will be reported as an academic offense. **When asked to hand in code and the output it creates, the code you submit must have been used to generate the document.** If it does not (i.e., the submitted code does not match the submitted output), this is also considered an academic offense.

RULES FOR TIMED ASSESSMENTS (E.G., MIXED ASSESSMENTS)

While all timed assessments in this course are open-book, they are not "open-person". You MUST NOT discuss any details of the assessment with anyone else during the assessment window, regardless of your completion status. This includes, but is not limited to, current classmates, friends, and tutors. For example, even asking someone "which slide did you look at to answer question 3" is not appropriate for timed assessments.

NOTE: BE CAREFUL ABOUT PRIVATE TUTORING COMPANIES

You may have been contacted by private tutoring companies trying to sell their services to you for statistics courses. Please be extremely careful with these services as some forms of tutoring can pose an academic offence risk. A good tutor helps you understand the subject area and supports your learning. A good tutor does not give you answers. **There are no shortcuts to learning. Learning takes time and effort.**

Be cautious about giving money to companies whose motivation is profit. They may tell you they have 'insider information'. They don't. They may even offer you the opportunity to commit academic offenses. Please do not put your University of Toronto education at risk by participating in these kinds of unacceptable behaviours. If you have any questions or concerns about what is okay and what is not in your course, please ask!

Course design principles

Here are some of the principles around which I have designed this course. I hope they might provide some useful insight into why some things are the way they are, and help you think about how to navigate this course and make the best of it.

The teaching team really wants you to have a great time in this course, learn lots of delicious statistics, and become confident, competent, and useful statistical thinkers. Please approach us and this course with an open mind and help us make it a good experience for you by providing thoughtful, constructive feedback.

HUMANS LEARN BETTER 'LITTLE AND OFTEN' *BUT* EVERYONE IS BURNOUT FROM TWO YEARS OF A PANDEMIC

I know some students absolutely hate weekly tasks and how common they have become with online learning, and that all the little tasks and deadlines can become overwhelming. I also know that cramming is the absolute worst way to learn and actually retain that learning.

This is why in this course:

- I have taken things out or made them baskets options, based on student feedback
- There is a *knowledge basket* that lets you approach frequent low-stakes practice in the way that suits you best and aims to facilitate and reward spaced repetition, the way science says is the best way to learn (this free course is fantastic for learning more about how to learn).
- Has been converted to have a two-week module structure to give you more flexibility.
- Has two assessments pathways with different numbers of assessments.

WRITING IS GOOD FOR STATISTICIANS

Writing not only helps you explain yourself to others, it can also be a fulfilling act of creative personal expression and a way to clarify your own understanding of a concept. Writing is an important part of this course because it is an important skill for your future careers/next steps in education. Lots of support and information here: https://writing.utoronto.ca/.

COURSE CONTENT IS ACCESSIBLE

My intention is to make this course accessible as possible with captions for all video and audio and Quercus/course guide design that is easy for folks using screen readers to navigate. If there is something I could do differently in this area that would make your life easier and you're comfortable to tell me, please do! One thing I know isn't great but, regretfully, don't have the resources to change, is the 'quality assurance' of the autogenerated captions for videos. Please reach out on Piazza if you're ever unsure about something they say.