

# 2025W1 STAT 301 – Statistical Modelling for Data Science

## Syllabus

---

### COURSE INFORMATION

---

Course Title	Course Code Number	Credit Value
Statistical Modelling for Data Science	STAT 301	3

### PREREQUISITES

---

[STAT 201](#) and one of [MATH 100](#), [MATH 102](#), [MATH 104](#), [MATH 110](#), [MATH 120](#), [MATH 180](#), [MATH 184](#), [SCIE 001](#).

### CONTACT

---

Course Instructor(s)	Contact Details	Office Location	Office Hours
Vivian Meng	Canvas messages ( <a href="#">preferred</a> )	TBA	TBA If you want/need a private chat, make an appointment using Canvas Message.

### OTHER INSTRUCTIONAL STAFF

---

Course TA(s)	Contact Details	Office Hours
Izzy	Private Piazza Post	<a href="#">Check Canvas</a>
Yian	Private Piazza Post	<a href="#">Check Canvas</a>
Jessica	Private Piazza Post	<a href="#">Check Canvas</a>

### COURSE STRUCTURE

---

The course is structured in weekly lectures and tutorials. **Students are expected to attend lectures and tutorials.**

### LEARNING OUTCOMES

---

By the end of the course, students are expected to be able to:

- Describe real-world examples of explanatory modelling and predictive modelling problems.
- Explain the trade-offs between model-based and non-model-based approaches and describe situations in which each might be preferred.
- Explain the difference between creating models for explanation and prediction in the context of how you choose and evaluate models and interpret the results.
- Based on the statistical question and data at hand, choose and apply a suitable method (e.g., regression, GLMs, sample size estimation, controlling for multiple testing, peeking, bandit

# 2025W1 STAT 301 – Statistical Modelling for Data Science

## Syllabus

---

algorithms, variable selection, and model diagnostics). Discuss the advantages and disadvantages of different methods that may be suitable for a given problem.

- Correctly interpret computer output when performing the statistical analyses presented in this course in the context of the statistical question being asked and the audience being reported to.
- Identify the assumptions/conditions required for each method to produce reliable results. Choose techniques to check (or at least be able to falsify) those assumptions. Discuss the consequence(s) of mapping the wrong methods to the question and/or data type.

## SCHEDULE OF TOPICS

---

This is a tentative schedule for the course. It is subject to changes.

Topic #	Topics
1	Linear regression: estimation and inference
2	Multiple linear regression (MLS)
3	MLS cont.: assumption and diagnostics
4	Logistic regression
5	Poisson regression
6	Model evaluation: goodness of fit
7	Model evaluation cont.
8	Model Regularization, Post-Inference
9	Prediction Uncertainty
10	Classifiers
11	(A/B testing if time permits)

## LEARNING ACTIVITIES

---

This course will have plenty of synchronous activities that students must work on during the lectures and tutorials. Students will work on activities in Jupyter Notebooks and iClickers.

## LEARNING MATERIALS

---

I strive not to require students to buy textbooks. That's not always possible. For this course specifically, we have:

- [Introduction to statistical learning](#) (R version) by Gareth James, Daniela Witten, Trevor Hastie, and Rob Tibshirani
- [Introduction to Data Science](#) by Rafael A. Irizarry
- [Beyond Multiple Linear Regression](#) by Paul Roback and Julie Legler
- [ModernDive: Statistical Inference via Data Science](#) by Chester Ismay and Albert Y. Kim
- [Openintro to Statistics](#) by David Diez, Mine Çetinkaya-Rundel

Additional resources may be noted during lecture.

# 2025W1 STAT 301 – Statistical Modelling for Data Science

## Syllabus

---

### POLICY REGARDING ASSESSMENTS

---

The assessment of learning will have four main components:

- **Lectures' Worksheets (4%):** fully auto-graded with visible tests to help you identify points that need more clarification.
- **Tutorials' Worksheets (6%):** only a few exercises will have visible tests.
- **Group Project (15%):** A project that you will work with your group throughout the term. Details about the Group Project will be made available to you on Canvas.
- **Exams:** The types of questions can vary: reasoning, multiple-choice, multiple-answer, dropdown, true or false. Although most questions will be about the content, you can expect some coding questions. The coding question will not be overly complicated, and we will only check your familiarity with the main functions and packages we use in the course. We are not trying to test your software development skills here!!! Please don't spend energy trying to memorize everything. If you have done the worksheets and tutorial, this should not be a problem for you.
  - **Midterm (25%)**
  - **Final Exam (50%)** – note that the student must pass the final exam to pass the course.

**IMPORTANT!** *To pass the course it is normally necessary to pass the final examination. Otherwise, the final exam grade is normally reported. If the 50% final exam threshold is satisfied, the grade reported is weighted based on all assessment components above.*

### UNIVERSITY POLICIES

---

UBC provides resources to support student learning and to maintain healthy lifestyles but recognizes that sometimes crises arise, so there are additional resources to access, including those for survivors of sexual violence. UBC values respect for the person and ideas of all members of the academic community. Harassment and discrimination are not tolerated, nor is suppression of academic freedom. UBC provides appropriate accommodation for students with disabilities and for religious observances. UBC values academic honesty, and students are expected to acknowledge the ideas generated by others and uphold the highest academic standards in all of their actions. Details of the policies and how to access support are available on [the UBC Senate website](#).

#### Some resources

- [Academic Concession](#)
- [Academic Honesty and Standards](#)
- [Attendance](#)
- [Grading Practices](#)
- [Student Conduct and Discipline](#)
- [Viewing Marked Work](#)

# 2025W1 STAT 301 – Statistical Modelling for Data Science

## Syllabus

---

### COURSE POLICIES

---

#### Late/Absence

- Regular attendance to lectures and tutorials is expected of students. Students who are unavoidably absent from numerous classes because of illness or other reasons should inform the instructor(s) of the course as soon as possible, preferably prior to the start of the lecture/tutorial.
- There will be no make-up exams. Students who miss the midterm or an assignment and want to request an Academic Concession need to contact the Instructor as soon as possible and provide a self-declaration form. Failing to present a declaration may result in a grade of zero on the assessment.
- Late assignment submissions will receive a grade of 0, and the deadlines shown on canvas are firm with no exceptions. But note that:
  - At the end of the semester, one lecture worksheet and one tutorial worksheet will be dropped.
  - If you have extenuating circumstances and need concessions **beyond** dropping one worksheet and one tutorial, please contact the instructor. Note that concessions will only be considered after the “free drop” has occurred.

#### Missed Final

Students who miss the final exam must report to their faculty advising office within 72 hours of the missed exam, and must supply supporting documentation. Only your faculty advising office can grant deferred standing in a course. You must also notify your instructor prior to (if possible) or immediately after the exam. Your instructor will let you know when you are expected to write your deferred exam. Deferred exams will ONLY be provided to students who have applied for and received deferred standing from their faculty.

#### Regrading

If you have concerns about the way your work was graded, please open a request within one week of having the grade returned to you. After this one-week window, we may deny your request for re-evaluation. Also, please keep in mind that your grade may go up or down due to re-grading. To open a regrade request, please follow the steps below:

1. Go to Piazza and click on [New post](#).

## 2025W1 STAT 301 – Statistical Modelling for Data Science

### Syllabus

---

2. In **Post Type**, select **Question**.
3. Make the post private to instructors and TAs only. In **Post to** select **Individual Students(s)/Instructor(s)**. A text box will appear, where you must type **Instructors**.
4. In **Select Folder(s)** select the folder **regrading**
5. In **Summary** say the Assignment you want to be regraded, followed by the question and your name and student number. For example, **lab 3 -> Q3 -- Rodolfo Lourenzutti (9982313)**
6. Provide a brief reason for why the regrade is needed.
7. The TAs will see the request and will take a look at the assignment. If necessary, they will involve the instructors. Finally, once the TA is finished reassessing the assignment:
  - If the grade deserves more marks: the TA will update the mark on Canvas and comment on the question so everyone can see that the question has been addressed.
  - If your grade goes down or stays the same: the TA will answer the post on Piazza, giving the student a reason for their final decision

#### **Autograder Policy**

Many of the questions in assignments are graded automatically by software. The grading computer has the same hardware setup as the server students work on. No assignment, when completed, should take longer than 5 minutes to run on the server. The autograder will automatically stop (time out) for each student assignment after a maximum of 5 minutes; any ungraded questions at that point will receive a score of 0. Furthermore, students are responsible for making sure their assignments are reproducible and run from beginning to end on the auto-grading computer. In short, whatever grade the autograder returns after 5 minutes (assuming the teaching team did not make an error) is the grade that will be assigned.

**Tip:** when you're done with the assignment, click "Restart and Run All" and check that the visible tests are all working and that your notebook runs in less than 5 minutes.

#### **Device/Browser Policy**

Students are responsible for using a device and browser compatible with all functionality of Canvas. Lockdown browser will be used for exams so students must address any issues with their laptop prior to the exam. We will test it with a mock quiz before the exam in the review session. Chrome or Firefox browsers are recommended; Safari has had issues with Canvas quizzes in the past.

#### **Academic Integrity**

Discussion of ideas learned in class is encourage (with other students, TAs or the instructor). This helps the leaning process. But individual work turned in by each student should be your own work. The academic enterprise is founded on honesty, civility, and integrity. As members of this enterprise, all students are expected to know, understand, and follow the codes of conduct

## 2025W1 STAT 301 – Statistical Modelling for Data Science

### Syllabus

---

regarding academic integrity. At the most basic level, this means submitting only original work done by you and acknowledging all sources of information or ideas and attributing them to others as required. This also means you should not cheat, copy, or mislead others about what is your work.

- Do not copy or paraphrase solutions from other students or from other sources.
- Do not provide your solutions to another student.

Violations of academic integrity (i.e., misconduct) lead to the breakdown of the academic enterprise, and therefore serious consequences arise and harsh sanctions are imposed. For example, incidences of plagiarism or cheating may result in a mark of zero on the assignment or exam and more serious consequences may apply if the matter is referred to the President's Advisory Committee on Student Discipline. Careful records are kept in order to monitor and prevent recurrences. A more detailed description of academic integrity, including the University's policies and procedures, may be found in the Academic Calendar.

#### Plagiarism

Students must correctly cite any code or text that has been authored by someone else or by the student themselves for other assignments. Cases of plagiarism may include, but are not limited to:

1. the reproduction (copying and pasting) of code or text with none or minimal (e.g., changing the name of the variables).
2. the translation of an algorithm or a script from a language to another.
3. the generation of code by automatic code-generation software. An "adequate acknowledgement" requires a detailed identification of the (parts of the) code or text reused and a full citation of the original source code that has been reused. The above attribution policy applies only to assignments. No code or text may be copied (with or without attribution) from any source during a quiz or exam. Answers must always be in your own words. At a minimum, copying will result in a grade of 0 for the related question. Repeated plagiarism of any form could result in larger penalties, including failure of the course.

#### COPYRIGHT

---

All materials of this course (course handouts, lecture slides, assessments, course readings, etc.) are the intellectual property of the Course Instructor or licensed to be used in this course by the copyright owner. Redistribution of these materials by any means without the permission of the copyright holder(s) constitutes a breach of copyright and may lead to academic discipline.

#### LAND ACKNOWLEDGEMENT

---

We acknowledge that the UBC Vancouver campus is situated within the traditional, ancestral and unceded territory of the Musqueam First Nation.