

STAT 301 - Statistical Modelling for Data Science

(2026, Summer Term 2)

Lecture:

Tutorial:

Thursdays 12:30pm–1:30pm, ESB1012.

Class webpage:

UBC Canvas STAT301.

Instructor:

Lang Wu, Professor, Department of Statistics, UBC.

Personal webpage: <https://www.stat.ubc.ca/~lang>

Email: lang@stat.ubc.ca

Office: *Earth Science Building (ESB) 3126*, 2207 Main Mall.

Office hours:

See Canvas class page. Instructor and TA office hours start in the second week of class. Most office hours will be online via Zoom, but some may be in person (you will be notified by email in advance). You are also welcome to talk to the instructor/TAs *in-person* right after class/tutorial. *You are strongly encouraged to post your questions on **Piazza discussion board** via Canvas class page.*

Teaching assistants:

See Canvas.

Course policy:

Regular attendance to lectures and tutorials is expected of students. **You must bring a laptop to each lecture class for in-class activities.** Late assignment submissions (in-class activities, lecture and tutorial worksheets, project) will not be accepted, and the deadlines shown on canvas are firm with no exceptions.

Email policy:

Due to large volumes of emails the instructor receives daily, the instructor may be unable to respond individual emails on time. *Please use email for confidential and personal matters only.* Please use the online **Piazza discussion board** on Canvas to discuss course materials.

Textbook:

There is no required textbook for this course. The lecture notes posted on Canvas will define the course materials. Some **references** are listed at the end, which can be freely accessed online via the UBC library.

Prerequisite:

DSCI 100, STAT 201, and one of MATH 100, MATH 102, MATH 104, MATH 110, MATH 120, MATH 180, MATH 184, SCIE 001. Access to computer is required in lecture class (if you don't have a laptop, you may be able to load one from UBC library).

Lecture:

The focus of the course is *understanding* of statistical models and methods (e.g., choosing the right models/methods for a given dataset), not memorization and tedious computation, as well as case studies and real data analysis. All lectures are in-person. Almost all lectures will involve *iClicker questions* and/or *in-class activities*. These are designed for you to test your understanding of the course materials and for you to practice during class.

Tutorial:

Tutorials start in the *second week* of class, in-person.

Assignments:

The assessment of learning will have five components:

- *Lectures' Worksheets*: fully auto-graded with visible tests to help you identify points that need more clarification. Weekly.
- *Tutorials' Worksheets*: only a few exercises will have visible tests. Weekly.
- *Group project*: 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.
- *iClicker*: We will be using iClicker in most lecture classes. You will be graded on *both participation and performance* (50% each). Weekly. The two lowest iClicker marks will be dropped.
- *In-class activity*: There are designed for you to practice in class. You must submit your work by the end of each class. Weekly.

Both lecture and tutorial worksheets are due each

Sunday by 11:59pm.

You should complete these assignments individually and independently. *Please read the "Instructions for all worksheets and tutorials" on Canvas "Assignments" section for more details.* The lower two lecture worksheet scores and the lowest two tutorial worksheet scores will be dropped.

Since the summer classes are intensive, there may be two lecture worksheets and two tutorial worksheets each week. Please see Canvas Assignment section for more details. Worksheets and tutorial problem sets will be done using Jupyter Notebook to an external site and R. Students will access the worksheets and tutorials in Jupyter Notebooks through Canvas. *If you have any questions regarding lectures or tutorials' worksheets, please post them on Piazza.* **No late submissions will be accepted.**

Group project:

There will be one group project, due

Sunday, August 2, by 11:59pm

(upload on Canvas). Please see Canvas project description for details.

Piazza discussion board:

We will use the Piazza online platform (access via Canvas) to allow students, TAs, and the instructor to discuss course materials. *In particular, if you have any questions regarding the assignments and course materials, please post them on Piazza. Do not email TAs or the instructor.* The TAs and the instructor will be spending time moderating and posting to Piazza, if the questions cannot be answered by students. *Students are strongly encouraged to answer questions posted on Piazza* (answers by students will be checked by the TAs or instructors and may be revised if necessary).

Class announcements:

Class announcements will be sent via emails. Please check your emails regularly.

R software:

We will use software R in class. R code will not be tested in the exams directly, but it is important to understand course materials illustrated in R (these will be tested in the exams).

Exams:

There will be a midterm exam and a final exam – both are written exams in person. The midterm exam will be on

Thursday, July 23, in class.

Details of the exams can be found on Canvas. Final exam date/time will be announced by the university. **No makeup exams.** If you miss a midterm with good reasons (e.g., sick), the midterm weight will be transferred to your final exam. If you miss the final exam with good reasons, you have to formally apply for a deferred final exam through your Faculty (e.g., Faculty of Science). Please see Canvas for more details about exams.

Evaluation:

Your final grade will be computed based on the following weighting scheme: iClicker and in-class activity 10% (5% each), lecture worksheets: 5%, tutorial worksheets: 5%, group project: 10%, midterm: 20%, final exam: 50%. The weights may be adjusted if necessary.

Lecture topics:

1. Linear regression models (estimation, inference, assumptions, and diagnostics),
2. Logistic regression models,
3. Poisson regression models,
4. Model evaluation and goodness of fit,
5. Model regularization and post-inference,

6. Prediction uncertainty,
7. Classifiers,
8. A/B testing (if time permits).

References:

The following references are closely related to course materials and can be found on UBC Library:

- 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