STAT 306 - *Finding Relationships in Data*

Winter Term 1 (Sep 4 – Dec 7, 2023)

Course description

Modeling a response (output) variable as a function of several explanatory (input) variables:

- Multiple regression for a continuous response.
- Logistic regression for a binary response.
- Log-linear models for count data.
- Finding low-dimensional structure using principal components analysis.

This course emphasizes:

- Applications of statistical methods such as multiple regression, binary regression, principal component analysis.
- The use of statistical software to do the computations.
- Interpretation of statistical analysis and output of statistical software.
- Understanding of statistical modelling procedure.

There is some linear algebra (with matrix representations) to show how multiple regression is computed in software, and there is some probability (mainly expected values, variances and covariances for linear combinations) to show how standard errors are determined for parameter estimates and predictions.

Objectives

On completing the course, students should be able to demonstrate an understanding of the techniques and applications of well-known ideas in linear modelling, including:

- Model fitting
- Model selection
- Model diagnostics
- Model interpretation
- Basic ideas for generalised linear models and principal components analysis.

Pre-requisites

One of MATH 152, MATH 221, MATH 223 and one of STAT 200, STAT 241, STAT 251, STAT 300, BIOL 300, COMM 291, ECON 325, ECON 327, FRST 231, PSYC 218, PSYC 278, PSYC 366 and one of MATH 302, STAT 302.

Teaching style

This course is delivered with a flipped-classroom approach, where little time is devoted to seminar-style lectures. Instead, students learn by directly engaging with the material, for example through in-class group activities. See below for more detail.

Instructor

Xinglong Li (Email: xinglong.li@stat.ubc.ca). Please use the email only for personal matters that need to be discussed directly with the instructor. Please use **office hours** or Piazza for questions regarding assignment problems, textbook problems, labs, class notes, examples, etc.
Instructor Office Hours: Wednesdays, 1-2 pm in ESB 3174.

Classroom

Chemical and Biological Engineering Building 102
(Monday and Friday | **Class Time**: 10:30 am - 12:00 pm)

Teaching assistants

Olivia Liu (jiaping.liu@stat.ubc.ca)
Jim Liang (jim.liang@stat.ubc.ca) [Head TA]
TA office hours: To be announced. Details will be available on Canvas.

Recommended texts

There are a variety of books that cover at most of the material in this course, and it is suggested you try the UBC online library stock to find those that suit you. The course notes are:

Joe, Harry. (2020) Course Notes for STAT 306: *Finding Relationships in Data* which can be ordered from the UBC bookstore.

Teaching methods

This class uses a flipped-classroom approach, where students engage with course material before class and participate in activities during class time. Classes of approximately 1.5 hours will occur twice a week, with two online pencasts describing related materials for each week being available in advance. A pre-class activity and quiz (ungraded) is set before each class that should be completed on the morning of class days. In all sessions an in-class activity will replace at least part of the lecture component. R will be necessary for many of the in-class activities. Guided reading or other activities will be set at the end of some lectures to be completed prior to the next. There will be required lab assignments most weeks.

Programme of work

The study time should total around 8-10 hours per week. In addition to the contact hours, it is essential that learners spend approximately 5 hours per week on self-study for the course. A proposed workload for a typical week is as follows:

- Classes: 3 hours
- Lab: 1 hour
- Preparatory Activities (pre-class quizzes, pencasts, other reading): 2 hours
- WeBWorK: 2 hours
- Reading/reviewing: 1-2 hours
- Additional assignments and projects: 1-2 hours

Week	Class Dates	Assessments	WeBWorK	Labs
1	Sep 8	No labs. In-class activities/clicker questions starting from the first class.	None	
2	Sep 11/15	Labs begin.	WW 1	Lab 1
3	Sep 18/22	Last date to withdraw with no W shown on transcript is Sep 18, 2023.	WW 2	Lab 2
4	Sep 25/29		WW 3	Lab 3
5	Oct 6		WW 4	
6	Oct 13	**Assignment 1 due** Oct 15	WW 5	
7	Oct 16/20		WW 6	Lab 4
8	Oct 23/27	Last date to withdraw with W shown on transcript is Oct 27, 2023.		Lab 5
9	Oct 30/Nov 3	**Midterm Test**, Nov 3, 2023	WW 7	Lab 6
10	Nov 6/10	**Group Project Proposal due** Nov 6	WW 8	Lab 7
11	Nov 17		WW 9	
12	Nov 20/24	**Assignment 2 due** Nov 26	WW 10	Lab 8
13	Nov 27/Dec 1		WW 11	
14	Dec 4	**Group Project due** Dec 7		
	Dec 10 -	**Exam Period**		

iClicker cloud

We will be using iClicker Cloud in lectures. iClicker Cloud is a response system that allows you to use your own computer or mobile device to respond to questions posed by instructors during class. You need to set up an iClicker Cloud account and add STAT 306 as a course to this account. To do so, please follow https://lthub.ubc.ca/guides/iclicker-cloud-student-guide for details. Please ensure that your iClicker Cloud student ID identifies you by your name as it appears in Canvas.

Piazza Discussion Board

You can use "Piazza Discussion Board" to post your questions. This is where you can discuss ideas, strategies, and resources for solving the problems with your classmates. Please DO NOT POST ANSWERS to the questions in the WeBWorK assignments/written assignments and Labs. Instead, share your thoughts and approaches to solving the problems. Asking others how to solve a problem without first trying to solve it yourself will not be beneficial for your learning. TAs will not give the solution for assignment questions before the due date. But they will surely give hints as needed and let you know the correct directions. If you need more clarification, it is always better to contact TAs or the instructor during our office hours. Don't expect TAs will answer all your questions posted in Piazza page. To access the course Piazza page, please go to **"Piazza"** in the left menu in the Canvas course page.

WeBWorK

Please see the WeBWorK assignment dates in the Canvas course page. Please go to **"WeBWorK"** in the left menu in the Canvas course page to access WeBWorK.

Labs

Groups will be assigned for the labs and other group activities by the beginning of the third week of class. Lab assignments start the second week of class. Lab materials will be provided in the lab sessions. You are expected to work with your group in the lab to solve the lab assignments and hand in the finished lab activity at the end of the lab.

Group Project

There will be a group project in which students will work in pre-assigned groups on a data set of their selection. Groups will be assigned for the beginning of the third week of class. Further details will be available on Canvas. The final project is a written report submitted during the last week of term. You will be graded on writing. There is an interim stage proposal for review that is due at the beginning of Week 10.

Rough Lecture Topic Outline

There follows a provisional guide to the lectures. It is possible that the material covered in the classes will differ slightly from the description below. For each lecture, there is a corresponding pre-class activity, an in-class activity, and there will be associated questions in the labs and the WeBWorK.

1. Introduction and Motivation. Exploring Relationships Between Two Variables.
2. Least Squares Estimation for the Simple Linear Model.
3. Residuals. Properties of the Model.
4. Confidence Intervals for the Slope and an Expected Response.
5. Prediction Intervals.
6. Distribution Theory; Why the T Distribution?
7. Matrix Formulation of Linear Models.
8. Properties of Least Squares Estimators in Matrix Form.
9. Properties of Residuals and the Residual SS.
10. Dummy Variables in Linear Models.
11. More on Categorical Variables in Linear Models.
12. Quadratic Models and Curve Fitting.
13. Partial Correlation / Review
14. Model Selection, Including Mallows' Cp Statistic.
15. Mid-Term Test.
16. Leverage, Influence, Outliers, and the “Hat” Matrix.
17. Transformations.
18. Continuous Interactions / A Case Study.
19. Introducing Logistic Regression.
20. Further Logistic Regression.
21. Model Selection in Logistic Regression.
22. Introducing Poisson Regression and Principal Components Regression.
23. Further Poisson Regression and Principal Components Analysis.