QBS Biostatistics 2 (121): Regression

Spring Quarter 2019

Instructors: Todd MacKenzie and Tor Tosteson
Teaching Assistants: Yuka Moroishi and Xin Ran
Class Meetings: Tuesdays & Thursdays 2:30 – 4 pm, WTRB 571 E/W

Course Description

We cover the theory and applications of statistical modelling, also known as regression, as practiced in the quantitative biomedical sciences. We present linear models, generalized linear models including logistic and Poisson regression, and models for times-to-event (survival analysis). The course emphasizes the dual goals of modelling (i) prediction and (ii) causal inference, and includes applications of penalized (regularized) regression and propensity scores. In addition we describe methods for missing data. We use the statistical software R. There are two 90 minute meetings per week. The typical layout of the 90 minutes will be (i) review of previous material and homework, readings (10 minutes), (ii) didactic presentation with questions (30 minutes), students doing labs in class with instructor and TAs present to answer questions (25 minutes), (iii) didactic presentation continues (20 minutes), (iv) review and questions or preview of homework (5 minutes). The final meeting (lasting up to 2 hours) will feature presentations of class team projects by students.

Course Learning Outcomes

In this course, students will:

- Understand and apply regression for various types of dependent variables
- Understand confounding and causal inference tools to deal with it
- Understand prediction using linear models and its generalizations

Prerequisites: QBS 120 or permission of instructor

Teaching Methods & Philosophy

Students are encouraged to read materials (slides and textbook) before class and to come prepared with questions, and to expect to implement the statistical methods during the lab component of each class.

Class Climate & Inclusivity

A teacher always learns from the students. All questions are a chance for both the questioner, other students and the teacher to increase their understanding.

Office Hours: Email the instructor at any time to arrange a time to meet. Yuka Moroishi is available Fridays from 10:50am to 12:50pm. Xin Ran is available Mondays from 10am to Noon.

Student Evaluation

Assignments, 1 each for Weeks 1 to 8 (40%): There will be weekly assignments posted by Thursday at midnight. Problems will include applications of modelling and other and statistical analyses as well as mathematical derivations. Students are expected to submit the assignments using an R Markdown document. Assignments are due on Mondays by midnight, and briefly discussed at the beginning of the Tuesday session.

Team Project Due Week 10 (20%): The team of students will analyze data using a regression method and interpret their findings. The dataset chosen by the student needs approval by the instructor by the 7th week of class. Teams will consist of 3 to 5 students. The team will submit a report (e.g. slides) and make a 7 minute presentation with an additional 3 minutes to answer questions.

Class Participation (10%): Students are encouraged to ask and answer questions in class, and contribute to the learning experience of others.

Final Exam (30%): The final exam will be a take-home and equivalent in length to two of the weekly assignments.

Where class fits in terms of Data Science, Type and Applications?

Data Science		
Analytics	Algorithm	Statistical Inference
30	10	60
Theory vs Application		
Theory	Application	
25	75	

Reading Material (readings from both textbooks each week)

- *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models: 2nd Edition* by Vittinghoff, E., Glidden, D.V., Shiboski, S.C, McCulloch, C.E. 2012

- *An Introduction to Generalized Linear Models* by Dobson, A.J., Barnett A.G., CRC Press, (2018), 4th edition.

Course Schedule

Week	Dates	Topic	Key Words	Readings
1	March 26,28	Linear Models 1: one and two sample methods, multivariable regression, confounding	Least Squares, Coefficients, ANCOVA, Adjusted, Predictions, Residuals	V 3.1-3.3, 4.1-4.4 D 6.1 – 6.5
2	April 2,4	Linear Models 2: non-linearity, model selection	Interactions and Effect Modification, Nonlinearity and Smoothing, Transformations	V 4.5-4.7, 10.1-10.4 D: 6.7 – 6.8

3	April 9,11	Logistic regression and other methods for binary endpoints	Logistic regression, odds ratio, maximum likelihood	V 5.1-5.4, V8.1 D: 7.1 – 7.9
4	April 16,18	Generalized Linear Models and Modelling Counts	Link functions, Family, Poisson Regression, Deviance	V 8.1-8.5 D: 9.1 – 9.8
5 Tor	Apr 23, 25	Hierarchical and Longitudinal Data I: Mixed Effects Models	Random Effects, Random Intercept, Random Slope, Shrinkage, Conditional	V 7.5-7.8 D 11.5 - 11.7
6	April 30,May 2	Hierarchical and Longitudinal Data 1I: GEE	Within and Between Variation, Correlation Structure, Marginal (Population Averaged)	V 7.1-7.4 D 11.1 – 11.4
7	May 21, 23	1. Propensity Scores 2. Missing Data	1. Matching, Binning, Inverse weighting, Adjusted survival curves 2. Ignorability, Multiple Imputation	1. V 9.1-9.4 2. V 11.1-11.6
8	May 7, 9	Survival Analysis 1: Cox's Proportional Hazards Model	Censoring, Kaplan-Meier, Competing Risks, Hazard Ratio, Semi-parametric	V 6.1-6.2 D: 10.1 – 10.5
9	May 14, 16	Survival Analysis 2: Extensions to Cox's Model	Time-Dependent Hazard Ratios, Time-Dependent Covariates, Stratified Baseline Hazards	V 6.3-6.4 D: 10.6 – 10.7
10	May 28	Student Team Project Presentations		
11	May 30-June 3	Final Exam Take Home		