BST 210: Applied Regression Analysis
Harvard T.H. Chan School of Public Health

Instructor: Dr. Erin Lake
eklake@hsph.harvard.edu

Teaching Assistants: Beau Coker (beaucoker@g.harvard.edu)
Isabella Grabski (isabellagrabski@g.harvard.edu)
Christina Howe (chhowe@g.harvard.edu)

Lectures: Fall 2019, Tuesday & Thursday 11:30am – 1:00pm, FXB G12

Lab Sections: L1: Thursday 5:30pm-7pm Kresge G2
L2: Friday 8am-9:30am Kresge 202A
L3: Friday 11:30am-1pm FXB G11

Office Hours: Wednesday 11:30am-12:30pm in Building II, Room 436: Erin Lake
Monday 8:30-9:30am in FXB G3: Christina Howe
Monday 12:30-1:30pm in Kresge 204 (except Oct 28 in Kresge 201): Isabella Grabski
Tuesday 1:00-2:00pm in Kresge LL6: Beau Coker

Course Description:

Topics include model interpretation, model building, and model assessment for linear regression with continuous outcomes, logistic regression with binary outcomes, and proportional hazards regression with survival time outcomes. Specific topics include regression diagnostics, confounding and effect modification, goodness of fit, data transformations, splines and additive models, ordinal, multinomial, and conditional logistic regression, generalized linear models, over dispersion, Poisson regression for rate outcomes, hazard functions, and missing data. The course will provide students with the skills necessary to perform regression analyses and to critically interpret statistical issues related to regression applications in the public health literature.

Prerequisites: BIO 201, ID 201, or equivalent (or majoring in BST, CBQG, HDS). (Working knowledge of topics covered in Bernard Rosner's *Fundamentals of Biostatistics*, and some programming experience.)

Evaluation: Homework (25%)
Two group projects (25%)
Midterm exam (25%)
Final exam (25%)

Notes and Textbooks:

Electronic copies of course handouts, notes, homework assignments, datasets, and other materials will be posted on the course website. Course readings are suggested from the following recommended textbooks (available online through Countway Library via the course website):

F.E. Harrell, Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis, Springer, Second Edition. (includes example programs in R)

E. Vittinghoff, D.V. Glidden, S.C. Shiboski, and C.E. McCulloch, Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models, Springer, Second Edition. (includes example programs in Stata)

Readings from additional texts will be added from time to time.
(In the course outline below, H denotes Harrell book and V denotes Vittinghoff et al. book.)

Computing:

The course lectures and laboratory sessions will use different software packages (Stata, SAS, R) throughout the course. Many examples will also be posted on the course website. For homework and group projects, students may stick with one package, but may need to switch to another package for certain tasks, as each package has its strengths and limitations. Students can however, generally focus on one statistical package that they wish to develop their skills in.

Competencies:

After successful completion of this course, students will understand and be able to implement regression analyses in public health. In particular, students will be able to:

- Choose the appropriate regression method to answer a public health question.
- Perform analyses of continuous outcomes with linear regression, categorical outcomes with logistic regression and extensions, count outcomes with Poisson regression and extensions, and time to event outcomes with proportional hazards regression.
- Implement strategies to build, select, and assess regression models.
- Summarize and interpret regression models, including assessment of confounding and effect modification and nonlinear associations.
- Develop appropriate tabular and graphical representations of regression findings.
- Interpret and critically evaluate regression methods in the public health and medical literature.

Tentative Course Outline on Next Page:

(subject to change while course is in progress)

9/3*	Intro/Review	**Week 1:** Course introduction	Predictive multivariable modelling, simple linear regression, continuous and categorical predictors, connections with correlation and t-tests (H 1; V 1, 3.3)	
9/5	Lecture 1			Lab Wk1
9/10	Lecture 1	**Week 2:** Multiple linear regression	Model formulation and interpretation, assessment of confounding and effect modification, connections with analysis of variance (H 2.1-2.3; V 4.1-4.6)	
9/12	Lecture 2			Lab Wk2
9/17	Lecture 3	**Week 3:** Assessing model fit	Checking model assumptions of linearity, normality, constant variance; outlying, high leverage, and influential points (H 2.7; V 4.7)	HW1
9/19	Lecture 3			Lab Wk3
9/24	Lecture 4	**Week 4:** Relaxing linearity assumptions for continuous predictors	Nonlinear terms, data transformations, indicator variables, splines, additive models (H 2.4)	HW2
9/26	Lecture 5			Lab Wk4
10/1	Lecture 6	**Week 5:** Multivariable model selection	Bias-variance trade-off, variable selection methods, stepwise procedures, information criteria, overfitting, collinearity, data reduction, purposeful selection of covariates, model validation using resampling (H 4, 5, 7; V 10)	HW3
10/3	Lecture 7			Lab Wk5
10/8	Lecture 8	**Week 6:** Logistic Regression	Odds ratios, simple and multiple logistic regression, model formulation and interpretation, assessment of confounding and effect modification (H 10.1-10.3; V 5.1-5.3)	HW4- assigned (Group Project)
10/10	Lecture 9			Lab Wk6
10/15*	Lecture 10	**Week 7:** Assessing Model Fit	Model building, regression diagnostics, goodness of fit, model validation (H 10.4-10.11; V 5.4)	HW4- due
10/17	Lecture 11			Lab Wk7
10/22*	Lecture 12	**Week 8:** Extensions of Logistic Regression	Proportional odds model for ordinal responses, multinomial model for nominal responses (H 13.1-13.3)	HW5
10/24	Lecture 13			Lab Wk8 & review
10/29	Review Lecture	**Week 9:** Review and Midterm exam		**Midterm Exam**
10/31	**MIDTERM**			Lab Wk9
11/5	Lecture 14	**Week 10:** Conditional logistic regression and generalized linear models	Matched responses, conditional likelihood, conditional logistic regression, modelling the mean and variance in generalized linear models, over dispersion, robust variance estimation (V 8.1-8.3)	HW6
11/7	Lecture 15			Lab Wk10
11/12*	Lecture 16	**Week 11:** Poisson regression and extensions	Poisson models for counts and rates, negative binomial and zero-inflated Poisson models (V 8.1-8.3)	HW7
11/14	Lecture 17			Lab Wk11
11/19	Lecture 18	**Week 12:** Introduction to	Censored survival outcomes, survival and hazard functions, Kaplan-Meier and Altschuler-Nelson	HW8 – assigned

11/21	Lecture 19	survival analysis	estimators, parametric survival models, log rank tests (H 16.1-16.5, 17.1-17.2; V 3.5, 6.1)	(Group Project) Lab Wk12-13
11/26 11/28	Lecture 20 **Holiday**	**Week 13:** Proportional hazards modeling	Model formulation and interpretation, partial likelihood, assessment of confounding and effect modification, model building, tied responses (H 19.1-19.3; V 6.1, 6.2)	HW8 – due No Lab
12/3 12/5	Lecture 21 Lecture 22	**Week 14:** More on proportional hazards, sample size and power	Time-varying covariates, the stratified proportional hazards model, assessment of the proportional hazards assumption. (H 19.4-19.11; V 6.3-6.6)	HW9 Lab Wk14
12/10 12/12	Lecture 23 Lecture 24	**Week 15:** Missing Data	Missing Data and putting the topics together	Lab Wk15 & review
12/17 12/19	Review Lecture 25 **FINAL**	**Week 16:** Review and Final exam	Putting the topics together, and final exam	**Final Exam**

10/15* Indigenous Peoples' Day (school holiday) falls on Monday of this week.

10/22* End of Fall Term 1 (F1) falls on Friday 10/25 of this week.

11/12* Veteran's Day (school holiday) falls on Monday of this week.