HARVARD T.H. CHAN | SCHOOL OF PUBLIC HEALTH

BST 210 Applied Regression Analysis

Fall 2020 – *virtual via Zoom*

Website https://canvas.harvard.edu/courses/77924

Faculty Instructor Dr. Erin Lake
Instructor of Biostatistics
Co-Director, SM Biostatistics Programs
Department of Biostatistics
Faculty Website

TAs

Derek Shyr	(Biostatistics)
Cathy Wang	(Biostatistics)
Anjali Jha	(Computational Biology and Quantitative Genetics)
Ninon Becquart	(Health Data Science)

Course Meetings All lectures, labs, office hours via Zoom in Eastern Standard Time (EST)

Lectures:	Tuesday & Thursday, 11:30am – 1:00pm EST	
Lab Sections:	Thursday	5:30pm-7pm EST
	Friday	8:00am-9:30am EST
Lab OH:	Friday	11:30am-12:30pm EST
Office hours:		
Erin:	Tuesday	1:10pm - 2:15pm EST (some Mondays)
TAs:	Wed	8:30am – 9:30am EST
	Thursday	1:30pm – 2:30pm EST
	Friday	11:30am – 12:30pm EST (see Lab OH above)

Contact Information

E-mail: eklake@hsph.harvard.edu
derek.shyr@mail.harvard.edu
cathy.wang@mail.harvard.edu
anjalijha@hsph.harvard.edu
nbequart@hsph.harvard.edu

Course Description

We are virtual for Fall semester of 2020 at Harvard. Our course teaching staff is excited to teach from this platform. While there are downsides, there are many benefits and flexibilities to be found in online pedagogy. We will exploit these in this course wherever possible, while keeping learning front and center. If there was ever a time when the world needed well trained public health scientists and biostatisticians – it's now. In fact, along with traditional weekly homework assignments, students will

participate in the group project culminating assignment, on the research topic and problem of their choosing, with guidance from an assigned TA and the course instructor.

I. ***Curriculum***:

BST 210 is an intermediate level, applied biostatistics course in both classic and modern methods for the analysis of continuous, binary, polytomous, ordinal, count and survival (time to event) response data. *This course leverages its placement within the Department of Biostatistics at Harvard, drawing on the unique confluence of core statistical methods, data science, and computational biology and genetics that can be found there, as well as examples drawn from studies and data from local renowned hospitals, and research institutions around the world.* Applied Regression Analysis (BST 210) demonstrates the ubiquity and relevance of the linear model throughout most areas of modern science.

This course provides students with the methodological development, programming (R, SAS and/or Stata), and applied skills that are necessary to critically interpret statistical issues related to regression modeling and applications in public health and related disciplines. Most models presented in this course fall under the generalized linear model (GLM) framework, which will be linked throughout. Estimation and inference will be presented within the frequentist paradigm. Parametric, semi-parametric, and non-parametric approaches for describing and modeling data through the regression perspective will be explored. Missing data mechanisms as well as methods for addressing missing data will be discussed.

Important topics covered in this course also include diagnostics, transformations, confounding and effect modification, model building and selection, model assessment and validation, goodness of fit, over-dispersion, and model interpretation. Visualization and graphical skills will be woven in, throughout. Described also are power and sample size calculations for the models delineated below.

II. ***Specific Modeling Topics***:

- Linear regression
- Multiple Linear regression
- Splines and Additive Models
- Ridge and Lasso Regression
- Logistic Regression
- Conditional Logistic Regression
- Ordinal Logistic Regression
- Multinomial (polytomous) Regression
- Poisson Regression
- Exponential, Weibull, and Gamma Regression
- Cox Proportional Hazards Regression

III. ***Computing***:

Lectures will present examples and applications primarily in R using R markdown, but will also include examples in SAS and Stata. Lab sessions will use different software packages (R, SAS, Stata) throughout the course, and examples in each language will be posted on the website. Students may utilize a preferred package, or pursue one in which they wish to develop skills; switching at times to another package for certain tasks (each package has its strengths and limitations). In the case that students need help re hardware/software/technology:

- *IT Helpdesk*: (617) 432-4357 helpdesk@hsph.harvard.edu
- *Hours*: Monday - Friday, 8:00 AM - 6:00 PM EST

- http://www.hsph.harvard.edu/it

IV. ***3 Prerequisites:***

1. BIO 201 or ID 201 or (BIO202 and BIO203) or (BIO206 and (BIO207 or BIO208)), or an equivalent course or series (none taken concurrently with BST 210),
   OR
   current Biostatistics, CBQG or HDS student in the Department of Biostatistics
2. Working knowledge of topics covered in Bernard Rosner's Fundamentals of Biostatistics
3. Some programming experience (R, SAS, Stata, or similar) is helpful

V. ***Who should take this course?***

- Students from both within and outside the Department of Biostatistics enroll in this course every year, across HSPH, HMS, Harvard Dental School, and some local hospitals.
- BST 210 is a core requirement or a strongly recommended course (depending on program) for all master's degree programs in the Department of Biostatistics (namely Biostatistics, Health Data Science, Computational Biology and Quantitative Genetics).
- BST 210 is a pre-requisite for fall semester course BST 222 (Statistical Inference, or taken concurrently), spring semester courses BST 226 (Applied Longitudinal Analysis) and BST 223 (Applied Survival Analysis), as well as other applied or master's level Biostatistics courses (such as Applied Bayesian Analysis, Statistical Genetics), for all students.
- Numerous doctoral and master's students from programs outside of Biostatistics, such as Epidemiology, Global Health, Public Health, Health Policy, Biomedical Informatics, Biological and Biomedical Science, Clinical Effectiveness, and Harvard Dental School, enroll in Applied Regression Analysis each year.
- BST 210 offers a unique experience to non-biostatistics and biostatistics students alike, in that it draws a diverse, talented collection of students and quantitative skill sets, around a highly technical topic. Students are provided training in methodological development of concepts while carrying out application and interpretation alongside peers of varying backgrounds. This makes for a rich and engaging learning environment.

VI. ***Learning Objectives:***

After completion of this course, students will be able to:

- Understand and implement classic and modern regression analyses in a wide array of research areas and applications, including in data science, computational biology, genetics, epidemiology, global health, biomedical science, policy and beyond.
- Choose appropriate regression methods to answer specific research questions.
- 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, visualize 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.

VII. ***Credits:*** 5 credits

VIII. **Course Website (Canvas):** https://canvas.harvard.edu/courses/77924

The course Canvas site is a critical learning portal for this course where students will:

(1) Find

- Zoom links to all course meetings (lectures, labs, office hours, other)
- Course lecture notes, homework, lab handouts, datasets, coding resources
- Links to the recommended texts (electronic via Harvard's Countway Library), and articles
- Course announcements
- All feedback on assignments
- Discussion boards, group work, course community platforms (study halls)

(2) Engage

- Attend virtual course meetings
- Submit assignments, midterm and project
- Self check-ins
- Participate in asynchronous discussion boards
- Share presentations, videos

IX. Grading, Progress and Assessment

The final grade for this course will be based on:

- Participation (5%: achievable both synchronously and asynchronously)
- Homework (35%)
- Midterm exam – take home (20%)
- Full-term project (40%)
  (40% = 10% 2 project check-ins + 20% report + 5% presentation + 5% peer project reviews)

where

- *Classroom Participation* (5%)
  This course assumes substantial student participation. Participation can be accomplished in various ways.
  - We prefer that all students attend scheduled class meetings (at a minimum, the 'lectures'). For students in specific parts of the world on drastically different time zones, this may be difficult. The teaching staff recognizes this and will work with you.
  - Various forms of asynchronous course content will be counted toward class participation.

- *Homework Assignments* (35%)
  No late assignments will be accepted—except 1, no later than 3 days late.

  - Homework assignments are due on dates indicated, in order for the teaching staff to provide timely feedback to all course participants (this request has been shared by many cohorts of students across nearly all courses at HSPH).
  - No credit will be given for late homeworks (1 exception below).
  - Each student is allowed 1 late homework, not more than 3 days late.
  - Assignments should be submitted online using Canvas via the File Upload option (supported file types are restricted to PDFs only).

- *Midterm Exam* (20%)

  - The midterm exam will be take-home, assigned on Tuesday and due on Saturday. No collaboration is allowed.

- *Group Project culminating assignment* (full-term) (40%)

  - The full-term project is truly the culminating 'where the rubber meets the road' experience for BST 210 students, which will be introduced on the first day of class, and will be due by last week of the semester.
  - This project **cannot coincide or contain any part of a project for ANY other class*.*
  - Students are invited to work in location/affinity groups as a research team, emulating real-world scientific collaboration (teaching staff will design groups based on the class makeup once the semester is underway).
  - Group members each will be accountable for evaluating their own, and 3 other group projects.
  - Groups will meet together to develop their topic.
  - Groups will meet with other group(s) in a consulting forum, for feedback.
  - Groups will submit include a 'Specific Aims' type proposal in the Project Check-Ins.
  - Groups will submit a structured written report (outline of which will be delineated by teaching staff), a video presentation for class members to view during the final week of class, and an evaluation of 4 group projects (their own + 3 other peer groups' projects).

Course Readings

*All texts/readings are available electronically, free of cost

Electronic copies of course handouts, notes, homework assignments, datasets, and other materials will be posted on the course website. Readings from additional texts and articles will be added from time to time. *Note: In the course outline, H denotes Harrell book, V denotes Vittinghoff et al. book, KNNL denotes Kutner, Nachtsheim, Neter and Li book, and CH denotes Chatterjee and Hadi book.*

Readings are suggested from the following recommended textbooks (available electronically 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)

- M. Kutner, C. J. Nachtsheim, J. Neter, W. Li, Applied Linear Statistical Models, McGraw-Hill Irwin, Fifth Edition.

- S. Chatterjee and A.S. Hadi, Regression Analysis by Example, Wiley, Fifth Edition.

Course Schedule

Key Dates:*

- 9/1 First class falls on Tuesday of this week.
- 9/7 Labor Day (school holiday) falls on Monday of this week.
- 10/12 Indigenous Peoples' Day (school holiday) falls on Monday of this week.
- 10/20 End of Fall Term 1 (F1) falls on Friday 10/23 of this week.
- 11/11 Veteran's Day (school holiday) falls on Wednesday of this week.
- 11/26 Thanksgiving falls on Thursday of this week.
- 12/17 Last class falls on Thursday if this week.

Harvard University Policies

***Policy on Collaboration**
Because many students learn best when working in a group setting, we encourage collaborative learning in this course. Some students may find that working in groups is especially helpful in preparing to tackle the assignments and trying to master the statistical software. You may always discuss the assignments with other students (and with the instructor and TAs, of course). The only restriction on collaboration is that students submit individual work for homework assignments, not group work. That is, all written work must be your own, in your own words, reflecting your understanding of the assignment. Solutions prepared "in committee" or by copying, paraphrasing, or summarizing someone else's work are not acceptable. All computer output you submit must come from work that you have done yourself; handing in output from someone else's computer session is also not acceptable.

***Inclusivity Statement**
Diversity and inclusiveness are fundamental to public health education and practice. Students are encouraged to have an open mind and respect differences of all kinds. The teaching staff shares a responsibility with you for creating a learning climate that is hospitable to all perspectives and cultures. Please contact the course instructor if you have any concerns or suggestions.

***Bias Related Incident Reporting**
The Harvard Chan School believes all members of our community should be able to study and work in an environment where they feel safe and respected. As a mechanism to promote an inclusive community, we have created an anonymous bias-related incident reporting system. If you have experienced bias, please submit a report here so that the administration can track and address concerns as they arise and to better support members of the Harvard Chan community.

***Title IX**
Please see the course website for information from the university.

***Academic Integrity**
Each student in this course is expected to abide by the Harvard University and the Harvard T.H. Chan School of Public Health School's standards of Academic Integrity. All work submitted to meet course requirements is expected to be a student's own work. In the preparation of work submitted to meet course requirements, students should always take great care to distinguish their own ideas and knowledge from information derived from sources. Students must assume that collaboration in the completion of assignments is prohibited unless explicitly specified. Students must acknowledge any collaboration and its extent in all submitted work. This requirement applies to collaboration on editing as well as collaboration on substance. Should academic misconduct occur, the student(s) may be subject to disciplinary action as outlined in the Student Handbook. See the Student Handbook for additional policies related to academic integrity and disciplinary actions.

***Accommodations for Students with Disabilities**
Harvard University provides academic accommodations to students with disabilities. Any requests for academic accommodations should ideally be made before the first week of the semester, except for unusual circumstances, so arrangements can be made. Students must register with the Local Disability Coordinator in the Office for Student Affairs to verify their eligibility for appropriate accommodations. Contact Colleen Cronin ccronin@hsph.harvard.edu in all cases, including temporary disabilities.

***Religious Holidays, Absence Due to**
According to Chapter 151c, Section 2B, of the General Laws of Massachusetts, any student in an educational or vocational training institution, other than a religious or denominational training institution, who is unable, because of his or her religious beliefs, to attend classes or to participate in any examination, study, or work requirement on a particular day shall be excused from any such examination or requirement which he or she may have missed because of such absence on any particular day, provided that such makeup examination or work shall not create an unreasonable burden upon the School. See the student handbook for more information.

***Grade of Absence from Examination**
A student who cannot attend a regularly scheduled examination must request permission for an alternate examination from the instructor in advance of the examination. See the student handbook for more information.

***Final Examination Policy**
No student should be required to take more than two examinations during any one day of finals week. Students who have more than two examinations scheduled during a particular day during the final examination period may take their class schedules to the director for student affairs for assistance in arranging for an alternate time for all exams in excess of two. Please refer to the student handbook for the policy.

Course Evaluations

- Constructive feedback from students is a valuable resource for improving teaching.
- The feedback should be specific, focused and respectful.
- It should also address aspects of the course and teaching that are positive as well as those which need improvement.
- Instructors take student feedback for their courses very seriously.
- Completion of the evaluation is a requirement for each course. Your grade will not be available until you submit the evaluation. In addition, registration for future terms will be blocked until you have completed evaluations for courses in prior terms.