

Keck School of Medicine of USC

PM592: Regression Analysis for Health Data Science

Units: 4
Term: Fall 2022, 1:30PM – 5:30PM
Location: SSB 114
Instructor: Trevor A. Pickering (tpickeri@usc.edu)
TA: TBD

Office hour times and locations will be posted on Blackboard.

Course Overview

Course Description

This course provides a rigorous introduction to statistical methods for analyzing data with a focus on regression modeling for continuous and binary outcomes. Exploratory data analysis including summary statistics and plots, probability distribution functions, hypothesis testing, estimation, multivariate models, and prediction will all be covered. The R language will be used.

Learning Objectives

Through this course, students will become familiar with data analysis and regression using R. Students will learn:

- Basic statistics, random variables and probability distributions, estimation, hypothesis testing
- Linear regression, finding the best fitting line, interpreting intercept and slope estimates, testing hypotheses and forming confidence intervals
- Statistical methods for analysis of categorical outcome/response data.
- Logistic regression in relation to binary outcome data.
- General techniques of model building, variable selection, model diagnostics, goodness of fit, and interpretation of model estimates.
- To appropriately analyze and interpret analyses of continuous and categorical outcome data, and present statistical methods, results, and conclusions section

Course Preparation

Prerequisites	Admission to the MS PHDS program or permission of instructor.
Co-Requisites	It is recommended to take this class concurrently with PM566.
Recommended Preparation	The course assumes a background in statistics covering material through ANOVA, and familiarity with R or a similar programming language.

Course Format

This course will follow a “flipped course” format. Students are expected to view the week’s lecture prior to the live session and complete a check-in before the live session. The live session will consist of lab-type activities such as programming and problem solving.

Course Requirements

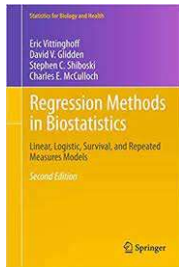
Communication

Blackboard (lecture slides, labs, data sets, assignment submission), USC e-mail (formal communication), Slack (informal communication)

Required Textbooks



Diez D, Cetinkaya-Rundel M, Barr CD. OpenIntro Statistics, 4th Edition, 2019. Available at: <https://leanpub.com/openintro-statistics>.



Vittinghoff E, Glidden D, Shiboski S, McCulloch C. Regression Methods in Biostatistics, 2nd Edition, 2012. E-copy available free through the USC Libraries.

Required Software



R and RStudio

This course will be taught in R, freely available online (<http://cran.r-project.org>).

Assessments

Grading Breakdown

Category	Points	Grade Range (%)	Letter Grade	Grade Range (%)	Letter Grade
Labs	10%	[93, 100]	A	[73, 77]	C
Check-In Quizzes	10%	[90, 93)	A-	[70, 73)	C-
Assignments	20%	[87, 90)	B+	[67, 70)	D+
Exam 1	20%	[83, 87)	B	[63, 67)	D
Exam 2	20%	[80, 83)	B-	[60, 63)	D-
Final Project	20%	[77, 80)	C+	[0, 60)	F

Late Submission Policy

Late assignments, check-ins, and labs are unacceptable and will receive no credit. To account for any unforeseen circumstances, the lowest 2 lab scores and the lowest quiz score will be dropped. The USC Student Health policy is to not provide medical notes for absences, but if you will be absent please fill out and email the self-verification medical absence form (<https://studenthealth.usc.edu/policy-on-medical-excuses-for-class-absence>).

Description of Assessments

- Labs** Labs will provide hands-on applications for the material learned in class and will be taught in the R programming language. There will be a brief assignment to confirm completion of the week's lab. Attendance in the live lab session, while optional, is strongly suggested.
- Assignments** Students will work to complete assignments involving the practical application of class concepts on actual data and research. Students may discuss problems with one another but must submit their own solutions. Assignments must be submitted to Blackboard by the indicated time.
- Check-Ins** There will be a short quiz to complete before each live session. These check-ins serve to ensure an individual's comprehension of the week's material and should be straightforward if the student has viewed the week's lecture.
- Exams** There will be two open-note exams, which will assess students' ability to use the techniques learned in class to answer research-related questions. These exams will be structured to integrate class-related knowledge in various ways.
- Project** Students will work on a project that will involve the analysis of real-world data and will provide a demonstration for your "portfolio" of work. Students concurrently enrolled in PM566 may use the same final project topic for both classes, and analyses for this class will complement the work they do in PM566. Otherwise, students should identify a project topic they would like to use for the final project.

Course Schedule

	Topics	VGSM	DCB	HW Due	Quiz Due
Class 1 M 8/21	Course Introduction. Variable types, sampling principles, summarizing data, visualizing data Lab: RStudio, Tidyverse, variable types, data import	1, 2	1, 2		
Class 2 M 8/28	Probability & Hypothesis Testing. Distributions of random variables, the sampling distribution, central limit theorem. Lab: R projects, merging, setting, Z & t tests, labels, factors	3.1	4 (3, 5, 6, 7)		Q2
M 9/4	Labor Day				
Class 3 M 9/11	Linear Regression I. Correlation, simple linear regression, variation in Y, the least-squares line, model coefficients, residuals, R-squared, model assumptions Lab: Variable manipulation (cut, quantile), functions, ggplot2	3.2, 3.3,	8.1, 8.2	HW1	Q3
Class 4 M 9/18	Linear Regression II. Model assumptions, ANOVA, log transformations, categorical IVs, dummy coding Lab: diagnostics (ggfortify), introduction to Stringr	4.3, 4.7.2-4.7.3	8.3, 8.4	HW2	Q4
Class 5 M 9/25	Multiple Regression. Multiple regression, multiple R-squared, collinearity, model diagnostics: influential values Lab: ANOVA, sums of squares, extra sums of squares test	4.2, 4.7.4-4.7.6	9.1, 9.3, 9.4	HW3	Q5
Class 6 M 10/2	Effects of a Third Variable. Confounding, mediation, effect modification Lab: The interaction package	4.4, 4.5, 4.6	*A	HW4	Q6
Class 7 M 10/9	Complex Coding Schemes. Splines, polynomial terms, dose-response coding, overfitting, adjusted r-squared Lab: Extra practice, interpreting parameter estimates	4.7.1	*B	HW5	Q7
M 10/16	Exam 1				
Class 8 M 10/23	Logistic Regression I. Intro to binary outcomes, contingency tables, the odds ratio, the logit link, logistic regression coefficients, maximum likelihood estimation Lab: Contingency tables, odds ratios, pseudo R-squared	5.1-5.3, 5.6, 5.7	9.5	HW6	Q8
Class 9 M 10/30	Logistic Regression II. Assessing linearity (grouped smooth, LOESS, fractional polynomials), goodness of fit, diagnostics, selection procedures (caret package) Lab: LR test, diagnostics practice	5.4		HW7	Q9
Class 10 M 11/6	Prediction Models. Building a prediction model, classification sensitivity/specificity, ROC Lab: Model development, training vs. testing splits	10	9.2	HW8	Q10
Class 11 M 11/13	Generalized Linear Models. Poisson, negative binomial Lab: Poisson modeling, estimated marginal means	8		HW9	Q11
Class 12 M 11/20	Survival Analysis. Kaplan-Meier, Cox proportional hazards Lab: Survival package	3.7, 6		HW10	Q12
M 11/27	Exam 2				
M 12/11	Final Project due to Blackboard by midnight				

*A: https://www.openintro.org/go/?id=stat_extra_interaction_effects

*B: https://www.openintro.org/go/?id=stat_extra_nonlinear_relationships

Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Part B, Section 11, “Behavior Violating University Standards” <https://policy.usc.edu/scampus-part-b/>. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, <http://policy.usc.edu/scientific-misconduct>.

Support Systems:

Student Counseling Services (SCS) - (213) 740-7711 – 24/7 on call

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention. <https://engemannshc.usc.edu/counseling/>

National Suicide Prevention Lifeline - 1-800-273-8255

Provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week. <http://www.suicidepreventionlifeline.org>

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 - 24/7 on call

Free and confidential therapy services, workshops, and training for situations related to gender-based harm. <https://engemannshc.usc.edu/rsvp/>

Sexual Assault Resource Center

For more information about how to get help or help a survivor, rights, reporting options, and additional resources, visit the website: <http://sarc.usc.edu/>

Office of Equity and Diversity (OED)/Title IX compliance – (213) 740-5086

Works with faculty, staff, visitors, applicants, and students around issues of protected class. <https://equity.usc.edu/>

Bias Assessment Response and Support

Incidents of bias, hate crimes and microaggressions need to be reported allowing for appropriate investigation and response. <https://studentaffairs.usc.edu/bias-assessment-response-support/>

The Office of Disability Services and Programs

Provides certification for students with disabilities and helps arrange relevant accommodations. <http://dsp.usc.edu>

Student Support and Advocacy – (213) 821-4710

Assists students and families in resolving complex issues adversely affecting their success as a student EX: personal, financial, and academic. <https://studentaffairs.usc.edu/ssa/>

Diversity at USC

Information on events, programs and training, the Diversity Task Force (including representatives for each school), chronology, participation, and various resources for students. <https://diversity.usc.edu/>

USC Emergency Information

Provides safety and other updates, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible, <http://emergency.usc.edu>

USC Department of Public Safety – 213-740-4321 (UPC) and 323-442-1000 (HSC) for 24-hour emergency assistance or to report a crime

Provides overall safety to USC community. <http://dps.usc.edu>