

# Keck School of Medicine of USC

## **PM592: Regression Analysis for Health Data Science**

Units: 4  
Term: Fall 2024, 1:30PM – 3:30PM  
Location: SSB 115  
Instructor: Trevor A. Pickering (tpickeri@usc.edu)  
TA: TBD

Office hour times and locations will be posted on Blackboard.

### **Course Overview**

#### **Course Overview**

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 to:

- Explain the basic statistical concepts underlying regression methods, such as random variables and probability distributions, estimation, and hypothesis testing.
- Conduct an ordinary linear regression, including how to formulate regression hypotheses, interpret intercept and slope estimates, and form confidence intervals
- Execute appropriate analytic methods for analysis of categorical outcome/response data, including logistic regression in relation to binary outcome data.
- Apply foundational techniques of model building, variable selection, model diagnostics, goodness of fit, and interpretation of model estimates.
- Appropriately analyze and interpret analyses of continuous and categorical outcome data, and present statistical methods, results, and conclusions sections.
- Implement and interpret output from other types of generalized linear models, such as Cox Proportional Hazards, Poisson, and Negative Binomial regression, and describe the assumptions of such models.

#### **Foundational Knowledge Needed**

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 in-person lab session and complete a check-in quiz before the live session. The lab session will consist of lab-type activities such as programming and problem solving.

# Course Requirements

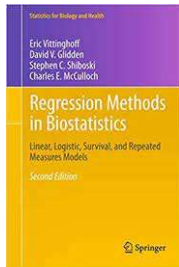
## Communication

Brightspace (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, 4<sup>th</sup> Edition, 2019. Available at: <https://leanpub.com/openintro-statistics>.



Vittinghoff E, Glidden D, Shiboski S, McCulloch C. Regression Methods in Biostatistics, 2<sup>nd</sup> 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>).

## USC Technology Support Links

- [Zoom information for students](#)
- [Software available to USC Campus](#)

## Classroom Norms

This class provides a rigorous, hands-on approach to regression methods in health data science. Students are encouraged to collaborate with each other in and out of class and use all communication platforms available to them, such as the course Blackboard and Slack. Students should feel free to ask questions during class, clarify concepts they are unsure of, and share their views. As part of group work, students should treat their teammates with respect and make an attempt to be a productive member of their team.

Synchronous (in-person) sessions will typically not be recorded or available via Zoom unless explicitly requested on a week-by-week basis.

## 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-in quizzes, and labs will receive no credit. To account for any unforeseen circumstances, the lowest 2 lab scores, lowest quiz score, and lowest homework score will be dropped.

The USC Student Health policy is to not provide medical notes for absences, (<https://studenthealth.usc.edu/policy-on-medical-excuses-for-class-absence>), but if you will be absent then please let the instructor know before class.

### 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. Students will work in groups to present responses to select lab questions during the lab.
- 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.

## Alignment Grid

<b>Learning Objective</b> By the end of this course, students are expected to be able to:	<b>Learning Activities</b> The learning objective will be facilitated by:	<b>Assignment/Assessment</b> This learning objective skill is measured by:
Explain the basic statistical concepts underlying regression methods, such as random variables and probability distributions, estimation, and hypothesis testing.	Lecture and lab activities from Weeks 1-3	Quizzes and homework from Weeks 1-3
Conduct an ordinary linear regression, including how to formulate regression hypotheses, interpret intercept and slope estimates, and form confidence intervals	Lecture and lab activities from Weeks 3-7	Quizzes and homework from Weeks 3-7, Exam 1
Execute appropriate analytic methods for analysis of categorical outcome/response data, including logistic regression in relation to binary outcome data.	Lecture and lab activities from Weeks 8-10	Quizzes and homework from Weeks 8-10, Exam 2
Apply foundational techniques of model building, variable selection, model diagnostics, goodness of fit, and interpretation of model estimates.	Lecture and lab activities from Weeks 4-7 and 9-10	Quizzes and homework from Weeks 4-7 and 9-10, Exams 1 and 2, Project
Appropriately analyze and interpret analyses of continuous and categorical outcome data, and present statistical methods, results, and conclusions sections.	Lecture and lab activities will incorporate information and examples about statistical writing	All homework and exams may assess this, Project
Implement and interpret output from other types of generalized linear models, such as Cox Proportional Hazards, Poisson, and Negative Binomial regression, and describe the assumptions of such models.	Lecture and lab activities from Weeks 11-12	Quizzes and homework from Weeks 11-12, Exam 2

## Course Schedule

	Topics	VGSM	DCB	HW Due	Quiz Due
<b>Class 1</b> M 8/26	<b>Course Introduction.</b> Variable types, sampling principles, summarizing data, visualizing data Lab: RStudio, Tidyverse, variable types, data import	1, 2	1, 2		
M 9/2	<b>Labor Day</b>				
<b>Class 2</b> M 9/9	<b>Probability &amp; Hypothesis Testing.</b> 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
<b>Class 3</b> M 9/16	<b>Linear Regression I.</b> 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
<b>Class 4</b> M 9/23	<b>Linear Regression II.</b> 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
<b>Class 5</b> M 9/30	<b>Multiple Regression.</b> 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
<b>Class 6</b> M 10/7	<b>Effects of a Third Variable.</b> Confounding, mediation, effect modification Lab: The interaction package	4.4, 4.5, 4.6	*A	HW4	Q6
<b>Class 7</b> M 10/14	<b>Complex Coding Schemes.</b> Splines, polynomial terms, dose-response coding, overfitting, adjusted r-squared Lab: Extra practice, interpreting parameter estimates	4.7.1	*B	HW5	Q7
M 10/21	<b>Exam 1</b>				
<b>Class 8</b> M 10/28	<b>Logistic Regression I.</b> 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
<b>Class 9</b> M 11/4	<b>Logistic Regression II.</b> Assessing linearity (grouped smooth, LOESS, fractional polynomials), goodness of fit, diagnostics, selection procedures (caret package) Lab: LR test, diagnostics practice	5.4		HW7	Q9
<b>Class 10</b> M 11/11	<b>Prediction Models.</b> Building a prediction model, classification sensitivity/specificity, ROC Lab: Model development, training vs. testing splits	10	9.2	HW8	Q10
<b>Class 11</b> M 11/18	<b>Generalized Linear Models.</b> Poisson, negative binomial Lab: Poisson modeling, estimated marginal means	8		HW9	Q11
<b>Class 12</b> M 11/25	<b>Survival Analysis.</b> Kaplan-Meier, Cox proportional hazards Lab: Survival package	3.7, 6		HW10	Q12
M 12/2	<b>Exam 2</b>			HW11	
M 12/16	<b>Final Project</b> due to Blackboard by midnight				

\*A: [https://www.openintro.org/go/?id=stat\\_extra\\_interaction\\_effects](https://www.openintro.org/go/?id=stat_extra_interaction_effects)

\*B: [https://www.openintro.org/go/?id=stat\\_extra\\_nonlinear\\_relationships](https://www.openintro.org/go/?id=stat_extra_nonlinear_relationships) c

## **USC Policies**

### **Sharing of Course Materials Outside of the Learning Environment**

USC has a policy that prohibits sharing of any synchronous and asynchronous course content outside of the learning environment.

Recording a university class without the express permission of the instructor and announcement to the class, or unless conducted pursuant to an Office of Student Accessibility Services (OSAS) accommodation. Recording can inhibit free discussion in the future, and thus infringe on the academic freedom of other students as well as the instructor. (Living our Unifying Values: The USC Student Handbook, page 13).

Distribution or use of notes, recordings, exams, or other intellectual property, based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study. This includes but is not limited to providing materials for distribution by services publishing course materials. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relationship to the class, whether obtained in class, via email, on the internet, or via any other media. (Living our Unifying Values: The USC Student Handbook, page 13).

### **Academic Integrity**

The University of Southern California is foremost a learning community committed to fostering successful scholars and researchers dedicated to the pursuit of knowledge and the transmission of ideas. Academic misconduct is in contrast to the university's mission to educate students through a broad array of first-rank academic, professional, and extracurricular programs and includes any act of dishonesty in the submission of academic work (either in draft or final form).

This course will follow the expectations for academic integrity as stated in the USC Student Handbook. All students are expected to submit assignments that are original work and prepared specifically for the course/section in this academic term. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s). Students suspected of engaging in academic misconduct will be reported to the Office of Academic Integrity.

Other violations of academic misconduct include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see the student handbook or the Office of Academic Integrity's website, and university policies on Research and Scholarship Misconduct.

## Statement on Academic Conduct and Support Systems

### Academic Integrity

The University of Southern California is a learning community committed to developing successful scholars and researchers dedicated to the pursuit of knowledge and the dissemination of ideas. Academic misconduct, which includes any act of dishonesty in the production or submission of academic work, compromises the integrity of the person who commits the act and can impugn the perceived integrity of the entire university community. It stands in opposition to the university's mission to research, educate, and contribute productively to our community and the world.

All students are expected to submit assignments that represent their own original work, and that have been prepared specifically for the course or section for which they have been submitted. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s).

Other violations of academic integrity include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), collusion, knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university. All incidences of academic misconduct will be reported to the Office of Academic Integrity and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see the student handbook or the Office of Academic Integrity's website, and university policies on Research and Scholarship Misconduct.

Please ask your instructor if you are unsure what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

### Students and Disability Accommodations

USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at [osas.usc.edu](http://osas.usc.edu). You may contact OSAS at (213) 740-0776 or via email at [osasfrontdesk@usc.edu](mailto:osasfrontdesk@usc.edu).

### 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>