SYLLABUS

Course	**Biostatistics 200A, Applied Linear Regression, Spring 2015**
Meeting Days/Location	**Lecture: Monday, Wednesday and Friday, 11-11:50, in CHS 33-105** **Discussion: Friday, 12-12:50, in CHS 33-105** Lab A: Tuesdays, 4-4:50, A1-241 Lab B: Mondays, 12-12:50, A1-241
Instructor Information	Kate Crespi, Ph.D. Department of Biostatistics, UCLA Fielding School of Public Health Email: ccrespi@ucla.edu Phone: 310-206-9364 Office: CHS A2-125 (within this suite, look for office AL-516) Office hours: Thursdays 11-12, or by appointment. Feel free to make an appointment.
Teaching Assistant	Daniel Conn, email: djconn17@gmail.com Office hours: Thursdays 3-3:50 Location: CHS A1-228
Course Description	The main objectives of this course are to gain a solid understanding of the theory and application of linear regression analysis and practical skills for developing linear regression models for inference and prediction and interpreting the results. We will learn both classical linear regression theory methods as well as some modern statistical learning methods. This course is designed primarily for students pursuing graduate degrees in biostatistics; it includes statistical theory and uses linear algebra.
Prerequisites	Preferred preparation is Biostatistics 110A and 110B. Previous coursework in linear algebra is also required.
Texts	Kutner, Nachtsheim, Neter, Li (2005) Applied Linear Statistical Models. 5th edition. McGraw-Hill. *This is out of print. The course reader has Chapters 1-11.* James, Witten, Hastie, Tibshirani (2013) An Introduction to Statistical Learning with Applications in R. Springer.
Homework Assignments	There will be regular homework assignments that will include a mix of problems, some to be solved by hand and some involving computing.
Computing	There will be computer labs once a week. We will use SAS and R.
Projects	There will be two data analysis projects. Project 1: Inference-based project Project 2: Prediction-based project
Holidays	There will be no lecture or lab on Memorial Day, Monday, May 25.
Exam Dates	Midterm exam: Friday, May 1, 11-12:50 pm Final exam: Tuesday, June 9, 8-11 am
Grading Policy	Grades for the course will be determined based on the following weighting:

15% Homework assignments
20% Midterm
25% Final exam
15% Project 1
20% Project 2
5% Attendance and participation

Topics (subject to change)

Topics	ALSM	ISLR	Other
Purpose of modeling: inference vs prediction		Chapter 2 (2.1)	
Linear regression with one predictor: model and estimation Model statement, meaning of parameters, least squares estimation, properties of fitted regression line, estimation of error variance, normal error regression model	Chapter 1	3.1	
Linear regression with one predictor: inference Sampling distributions, tests and confidence intervals; estimation of mean response; prediction of new observation; anova approach to partitioning sums of squares; R^2	Chapter 2 (2.1-2.10)		
Linear regression with one predictor: diagnostics Residuals; model departures; diagnostics for residuals; remedial measures; transformations; loess	Chapter 3 (3.1-3.6, 3.8-3.11)	3.3.3	
Linear regression with one predictor: other topics Regression through the origin; measurement error	Chapter 4 (4.4-4.5)		
Matrix approach to simple linear regression Matrices; matrix operations; special types of matrices; linear dependence; rank; matrix inverse; random vectors and matrices; SLR in matrix notation; least squares and normal equations; fitted values and residuals; anova results; inferences	Chapter 5		
Multiple regression models: basics General linear regression model; in matrix notation; estimation; anova results; inferences	Chapter 6	3.2	
Multiple regression models: partial effects Extra sums of squares; their use in tests; partial determination and correlation; standardized regression model; multicollinearity	Chapter 7		
Categorical predictors	Chapter 8	3.3.1	
Adjusted means			
Regression to the mean			Barnett article
Interactions	Chapter 8	3.3.2	Hayes article
Modeling nonlinear relationships 1 Polynomial regression; power transformations	Chapter 8	3.3.2	Fox Chap 4
Model diagnostics Added-variable plots; outliers and residuals; leverage; influence; variance inflation factor	Chapter 10		
Remedial measures Unequal error variance and weighted least squares; multicollinearity and ridge regression; influence and robust regression	Chapter 11 (11.1-11.3)		
Statistical learning; models for prediction Assessing model accuracy; bias-variance tradeoff		2.2	
Resampling methods		Chapter 5	

Cross-validation; the bootstrap			
Linear model selection Subset selection; shrinkage methods (ridge regression, the lasso)		Chapter 6 (6.1, 6.2	
Modeling nonlinear relationships 2 Basis functions; regression splines; smoothing splines; local regression; generalized additive models		Chapter 7	
Regression trees (if time permits) Decision tree basics; bagging, boosting, random forests		8.1, 8.2	

Learning objectives and competencies

Learning objectives	MS competencies	DrPH competencies	PhD competencies
Understand the theory and application of linear regression modeling	A2 Formulate a public health question in statistical terms	B2 Formulate a public health question in statistical terms.	A2 Formulate a public health or scientific question in statistical terms
Perform data analyses using linear regression modeling to meet scientific and public health objectives	A3 Identify the strengths and weaknesses of different study designs to address public health and scientific questions; communicate these issues to public health researchers A6 Conduct appropriate statistical analyses of study data and interpret the results. A8 Use statistical software to answer research questions and communicate the results to other research professionals C4 Develop ability to use new and evolving computational and digital technologies into biostatistical work	B1 Collaborate with researchers to formulate the aims of a public health research project. B2 Formulate a public health question in statistical terms. B3 Identify the strengths and weaknesses of different study designs to address public health and scientific questions; communicate these issues to public health researchers B6 Conduct appropriate statistical analyses of study data and interpret the results. E5 Develop ability to evaluate and incorporate new and evolving computational and digital technologies into biostatistical work	A1 Collaborate with researchers to formulate the aims of a public health research project. A3 Identify the strengths and weaknesses of different study designs to address public health and scientific questions; communicate these issues to public health researchers A5 Conduct appropriate statistical analyses of study data and interpret the results A8 Research biostatistical methods and computational resources for collaborative research B5 Develop ability to evaluate and incorporate new and evolving computational and digital technologies into biostatistical work
Communicate the assumptions and results of a linear regression analysis	A7 Effectively communicate the assumptions and results of analyses through oral and written communications to the collaborative team B1 Gauge the statistical skill	B7 Effectively communicate the assumptions and results of analyses through oral and written communications to the collaborative team	A6 Effectively communicate the assumptions and results of analyses through oral and written communications to the collaborative team C1 Gauge the statistical skill set of

	set of an audience to appropriately customize the level of biostatistical presentations. B2 Effectively communicate statistical concepts and reasoning to public health collaborators. B3 Learn to write and disseminate substantive field publications and communicate the statistical portion of the methodology to a substantive field audience. B4 Learn digital tools useful for communication	C4 Effectively communicate statistical concepts and reasoning to public health collaborators D2 Write and present effective and clear reports or publications about the application of statistical methods to health problems. D5 Understand and be able to effectively communicate the public health significance of the problems being addressed	an audience to appropriately customize the level of biostatistical presentations. C2 Effectively communicate statistical concepts and reasoning to public health collaborators. C3 Learn to write and disseminate substantive field publications and communicate the statistical portion of the methodology to a substantive field audience. C4 Learn digital tools useful for communication D4 Effectively communicate statistical concepts and reasoning to public health collaborators D6 Learn to write and publish substantive field publications and communicate the statistical portion of the methodology to a substantive field audience. D7 Learn current and future digital tools useful for communication