140.654 Methods in Biostatistics IV
Generalized Linear Regression Models
Syllabus
Fourth Term, 2020-2021

Virtual Format:

Synchronous lecture (same lecture offered at two times):
Tuesdays, 7:30-8:50am EST, 10:30-11:50am EST

Asynchronous lecture replacing scheduled Thursday class session: Posted Wednesday

Synchronous lab (same lab offered at two times):
Tuesdays, 3:30-4:20pm EST, 9:30-10:20pm EST

Office hour: Thursday 10:30-11:50am EST and by appointment

Tentative class schedule:

In the meeting column, S indicates a synchronous meeting, R indicates recorded lecture

The date for the recorded session is the recommended date when students should listen to the recorded lecture (Thursdays), consistent with the in-person structure of the course. Recorded lectures will be posted on Wednesday.

Meeting	Date	Topic	Readings
1:R	Mar 23	Introduction to logistic regression Bernoulli distribution Logistic model Interpretation of logistic coefficients Connections to 2x2 tables	FEH 10.1 HTF 4.4 MN 4 Gill: Chapters 1 - 4
2:R	Mar 25	Continuation of: Connections to 2x2 tables. Evaluating confounding in generalized linear models, continuous exposures, invariance properties	FEH 10.1 HTF 4.4 MN 4 Gill: Chapters 1 – 4
	Mar 30	Breakday!	
3:R	Apr 1	Statistical inference for logistic regression models: Likelihood function, maximum likelihood estimation by iterative weighted least squares (IWLS)	FEH 9.1, 10.4 Gill: Chapter 5-6
4:S	Apr 6	Statistical inference for logistic regression models continued: Asymptotic results, Likelihood ratio tests, Wald tests	FEH 10.4-10.8 Gill: Chapter 5-6
5:R	Apr 8	Predictions/classification using logistic regression; Receiver-operator characteristic (ROC) curves; Cross-validated errors; Bootstrapping for error assessment	FEH 11 or 12
6:S	Apr 13	Classification and regression trees (CART):	FEH 11 or 12

		Application to predicting major smoking caused disease	HTF 9.2
7:R	Apr 15	Random forests: Application to predicting major smoking caused disease	FEH 11 or 12 HTF 15
8:S	Apr 20	Conditional logistic regression models; Conditional likelihood; Application to matched case-control studies	
	Apr 22	Breakday!	
9:S	Apr 27	Continuation of conditional logistic regression	
10:R	Apr 29	Log-linear regression for counted data; Poisson model; Log-linear regression; Interpretation of coefficients; Likelihood-based inference	
11:S	May 4	Case-Study II: Hurricane deaths	
12:R	May 6	Time-to-event data Log-linear analysis of interval data	
13:S	May 11	Survival analysis in continuous time Hazard and survival functions and their relationship to the density and distribution functions Kaplan-Meier estimate of the survival curve Inferences about a survival function Log-rank test of equality to two survival functions	FEH 16.1-16.2
14:R	May 13	Survival analysis continued: Cox proportional hazards model; interpretation; partial likelihood; connection to conditional logistic regression	FEH 16.3-5
15:S	May 18	Multiple comparisons with applications to statistical genetics	
16:R	May 20	Time to ask questions for final or outstanding questions/topics of interest from 653-4	

Tentative Lab Schedule:

Lab	Date	Work Pending	Data Analytic Skills Covered
1	Mar 23	Prob Set 1	Generalized linear models in R: continuation of in class example computing interaction model (Model C); use of *glm*
	Mar 30		Breakday!
2	Apr 6	Prob Set 1	Evaluating the fit of a generalized linear model: independence, predicted values -vs- predictors, residuals vs. predicted values
3	Apr 13	Prob Set 2	Evaluating predictions in R: Receiver operator characteristic curves (ROC curves); cross-validation and bootstrapping for unbiased assessments;
4	Apr 20	Prob Set 2	R eq5q analysis using random forests and classification trees
5	Apr 27	Prob Set 3	Application of marginal regression models
6	May 4	Prob Set 3	Log-linear models
7	May 11	Prob Set 3	Cox models and log rank test in R; Checking survival models
8	May 18	Prob Set 4 / Final project	Making 654 final projects something you are proud to include in your data science portfolio

Key Due Dates:
Problem Set 1: Friday April 9th
Quiz 1: Monday April 12th
Problem Set 2: Thursday April 29th
Quiz 2: Monday May 3rd
Problem Set 3: Friday May 14th
Quiz 3: Sunday May 16th
Problem Set 4: Friday May 21st

Office Hours

Elizabeth Colantuoni: Thursday, 10:30-11:50 AM EST and by request

Books for Reference

1. (HTF): Hastie, T, Tibshirani, R, Friedman, J. 2013. The Elements of Statistical Learning. Springer. http://statweb.stanford.edu/~tibs/ElemStatLearn/index.html

2. (FEH): Harrell, FE Jr. (2015). Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer. http://link.springer.com/book/10.1007%2F978-3-319-19425-7

3. (MN): McCullagh, P, Nelder, J. 1989. Generalized Linear Models – 2nd ed. Chapman and Hall. London. https://www.crcpress.com/Generalized-Linear-Models-Second-Edition/McCullagh-Nelder/p/book/9780412317606

4. (Gill): Jeff Gill. Generalized Linear Models. 2011. Sage. https://methods-sagepub-com.proxy1.library.jhu.edu/book/generalized-linear-models

The Statistical Sleuth: A course in methods of data analysis, 3rd edition. 2013
Fred Ramsey and Dan Schafer. Brooks/Cole Cengage Learning.