

Biostatistics 651: Applied Statistics II: Extensions for Linear Regression

Course Overview

Meeting Time and Place

M/W: 1:00-2:30 in M 1170, SPH II.

Office Hours: M/W 2:30-3:30, or by appointment

Course Summary

Classic linear regression assumes that observed outcome data of interest is a linear combination of observed predictor variables and unknown parameter values, to which independent and identically distributed normal error terms are added. This course will extend the concept of linear regression to the broader family of exponential distributions via generalized linear models (GLMs), which include the normal model as a special case but also allow for modeling the mean of dichotomous or integer-valued (count) outcomes as a function of covariates. Maximum likelihood estimation and likelihood inference will also be introduced in the GLM context, and iterative fitting algorithms described. Model fit and model selection procedures will be discussed. Use of standard software packages (SAS, R) will be described, with a focus on epidemiological and clinical research applications.

Texts

Required: *Generalized Linear Models*, 2nd Edition, McCullagh and Nelder

Background: *An Introduction to Generalized Linear Models*, Dobson

Categorical Data Analysis, Agresti

Applied Logistic Regression, Hosmer and Lemeshow

Applying Generalized Linear Models, Lindsey

Analysis of Longitudinal Data, Diggle, Liang and Zeger

Grades

Homework (~5 during semester): 30%

Midterm (Open book, in class): 20%

Final Project: 20%

Final (Open book, in class): 30%

Instructors

Michael Elliott, Assistant Professor of Biostatistics

SPH II, M4140

647-5160

Assistant Research Professor, Survey Methodology Program

ISR, 4065

647-5563

mrelliot@umich.edu

Tentative Schedule

1/9-1/22

Linear algebra review, brief overview of maximum likelihood estimation, multivariate normal distribution, assumptions of linear models, derivation of ordinary least squares estimator.

Review of linear regression: maximum likelihood estimation and linear model inference.

No Class 1/16: Martin Luther King Symposium.

1/23-1/29

Review of linear regression: linear model checking and model fit criteria, generalized least squares estimator.

1/30-2/5

Introduction to generalized linear models: exponential family distributions and canonical parameterization link functions and interpretation of parameters.

2/6-2/12

Estimation of GLMs: iterative reweighted least squares, Newton-Raphson algorithm.

2/13-2/19

Inference with GLMs: Fisher information; score, Wald, and likelihood ratio statistics.

Computation of confidence intervals: profile likelihood.

2/20-2/26

Review of common epidemiological measures and relationship to logistic and Poisson regression.

Binary/proportion data: logistic regression, probit regression, complementary log-log models.

2/27-3/5

“Spring” vacation

3/6-3/12

Count/rate data: Poisson, negative binomial regression. Overdispersion.

Midterm exam

3/13-3/19

Modeling considerations: confounding, interaction. Cochran-Mantel-Haenzel test.

Case-control and matched case-control studies: conditional logistic regression

3/20-3/26

Model fit in GLMs: deviance, Pearson chi-square, omnibus tests, ROC curves.

3/27-4/2

Model checking in GLMs: outliers, influential observations.

4/3-4/9

Robust estimators: quasi-likelihood models.

4/10-4/16

Repeated measures: generalized estimating equations (GEE), mixed-effect models.

4/17-4/22

Review. Last class 4/17; final exams begin 4/20.

Project due.

4/23-4/27

Final exams.