STAT R324: Applied Regression Analysis **Winter 2016**

Prereqs: IME 326, STAT 252, STAT 302, STAT 312, or STAT 313

This course is the non-matrix version of STAT 334. It is a prerequisite for all current courses with a STAT 324 or STAT 334 prerequisite (STAT 416, 418, & 434).

Course Description

Data analysis using linear regression models. Model equations, inference and prediction, interpretation of results, model assumptions and diagnostics, data transformations, model selection and validation, model building using indicator variables, interactions, polynomials. Logistic regression models for binary data. Not open to students with credit in STAT 334. 4 lectures. Prerequisite: IME 326, STAT 252, STAT 302, STAT 312, or STAT 313.

Course Learning Objectives

- Describe linear regression models using model equations
- Analyze data using linear regression models and interpret the results
- Recognize when data violate the assumptions of the linear regression model
- Apply transformations to data and interpret the results of models with transformed data
- Select the best model from a number of possibilities using several different criteria
- Use polynomials, indicator variables, and interactions to model relationships in data and interpret the results
- Analyze binary data using the logistic regression model and interpret the results

Unit 1: Simple and multiple regression side-by-side with model equation and graphics

Estimated time: 6 hours

I. Review of simple and multiple regression with model equations
II. Interpretations (& graphics): slopes (scatterplots) and partial slopes (partial regression plots)
III. Descriptive measures of model fit: R^2, adjusted R^2, SSE, AIC, BIC
IV. Formal testing: variable, partial, and global *F* tests
V. Bonferroni adjustment for multiple tests: advantages and disadvantages
VI. Estimation: confidence and prediction intervals

Unit 2: Diagnostics

Estimated time: 4 hours

I. Formal diagnostics of assumptions
    a. Review of graphics
    b. Lack of fit test
    c. Tests of normality and equal variance
    d. Detecting correlated errors
II. Collinearity: VIF
III. Outliers, leverage, influence

Unit 3: Transformation

Estimated time: 7 hours

I. Transformations of *x* only, *y* only, both *x* and *y*
II. Box-Cox idea
III. Transformations to fix collinearity
IV. Weighted least squares to fix nonconstant variance
V. Transformations to fix correlated errors
    a. First differences, Cochrane-Orcutt method
    b. Forecasting with correlated errors
VI. Interpretations under transformed models (heavy emphasis)

Unit 4: Model selection and validation

Estimated time: 3 hours

I. Stepwise regression (brief)
II. "Best" subsets
III. Comparing models (partial F-test, Mallow's C_p, AIC, BIC, R^2 adjusted, PRESS)
IV. Cross-validation

Unit 5: Higher-order models and indicator variables

Estimated time: 8 hours

I. Polynomial regression
II. Indicator variables in regression
    a. Different parameterizations of indicator variables
    b. Connection to ANOVA & ANCOVA
III. Interactions: categorical x categorical, categorical x quant, quant x quant
IV. Centering and standardizing to address variance inflation
V. Piecewise linear and polynomial models
VI. Model selection using indicators, interactions, polynomials, etc.

Unit 6: Logistic regression

Estimated time: 4 hours

I. Simple logistic regression model
    a. Logit transformation
    b. Model equation
II. Interpretation of odds, log-odds, odds ratio
III. Computing predicted probabilities
IV. Formal inference for both quantitative and categorical predictors

Unit 7: Instructor's discretion

Estimated time: 4 hours

Total time: 36 hours