Topics covered, STAT 615, Regression and Linear Models

1. Simple Linear Regression (SLR)
    a. Formal statement of the model and model assumptions
    b. Definition of first-order linear model
    c. Method of Least Squares (LS) Estimation (LSE) of regression parameters
    d. LS Normal Equations and solutions, in scalar form
    e. Alternative SLR Models

2. Multiple Linear Regression
    a. Extension of all above to multiple regression, in vector-matrix form
    b. Hat matrix and properties

3. Properties of Least Squares Estimators / Estimates
    a. Gauss-Markov Theorem
    b. Definition of linear estimator
    c. Variance of \hat\beta_j; Cov(\hat\beta_0, \hat\beta_1) in regular LSE model and alternative SLR model 2; variance of residuals
    d. Distribution of \hat\beta_j under normal error assumptions
    e. Estimation of variance \hat\sigma^2, and the effect of using \hat\sigma^2 instead of \sigma^2 in 3.d

4. Hypothesis Testing
    a. Conceptual framework, definitions
    b. The decision rule: test-statistic
    c. Student t-test
    d. Confidence interval

5. ANalysis Of Variance (ANOVA) for simple regression
    a. Conceptual setup
    b. ANOVA for hypothesis testing (testing the significance of regression)
    c. Cochran's theorem, Chi-square and F-distributions
    d. Coefficient of determination (R-squared) and properties

6. Interval Estimation of the Mean Response
    a. Problem statement, assumptions
    b. Confidence interval for the mean response
    c. Prediction of a future value

7. Maximum Likelihood Estimation (MLE)
    a. The likelihood function
    b. MLE for regression
    c. MLE vs LSE
    d. Properties of ML estimators

8. (More) Multiple Linear Regression, MLR
    a. ANOVA for multiple linear regression
    b. Hypothesis testing, for significance of multiple linear regression
    c. Maximum Likelihood Estimation for MLR
    d. Geometry of LS estimates

e. Coefficient and adjusted coefficient of multiple determination (R-squared and adjusted R-squared), their properties, relationship to each other and to F-statistic
f. Quadratic forms and hat matrix
g. Derivation of degrees of freedom for ANOVA sums of squares

9. Model Selection with Hypothesis Testing
    a. Problem setup, assumptions
    b. Choosing between smaller and larger model using ANOVA F-test

10. Checking the Assumptions of Regression Models
    a. Model assumptions for MLR
    b. Checking error assumptions (homoscedasticity, normality, uncorrelatedness)
        i. Residual plots,
        ii. Q-Q plots, theory behind (quantiles), interpretation
    c. Checking for unusual observations (leverage points, outliers)
        i. Leverage points and hat matrix
        ii. Assessing the influence of outliers using hat matrix, Cook's Distance, PRESS residuals; Bonferroni correction, DFFITS and DFBETAS
    d. Checking uncorrelatedness (coefficient of correlation, AR(1) model, Durbin-Watson test, Autocorrelation Function)
    e. Checking structural assumptions (Partial Regression or Added Variable Plots)

11. Joint Confidence Intervals, and Multicollinearity
    a. Bonferroni joint CI-s
    b. Confidence ellipse for joint CI-s
    c. Eigenvealues and eigenvectors of X^TX, relationship to confidence ellipse
    d. Multicollinearity, and detection from scatterplots, partial regressions, condition number of X^TX, and Variance Inflation Factors

12. Permutation Test
    a. Motivation and formal setup for hypothesis testing
    b. Algorithm
    c. Assessing the reliability of the p-value obtained from the permutation test

13. Testing For Lack of Fit (LOF)
    a. Formal problem statement, replicates
    b. Lack Of Fit test
    c. Remedial Transformations
        i. Transforms of the response, and consequences
        ii. Box-Cox method, and constraints / limitations
        iii. Transforms of the regressors

14. Generalized Least Squares and Robust Regression – and what they remedy
    a. Generalized Least Squares (GLS)
    b. Weighted Least Squares (WLS)
    c. Robust Regression and its relation to ordinary Least Squares (OLS) and WLS
        i. LAD, Huber, Bisquare, LTS;
        ii. Method of Iterative (Re)weighted Least Squares
    d. Ridge regression
        i. Thikhonov regularization, Bias-Variance trade-off
        ii. Graphical interpretation of ridge regression, ridge trace

15. Variable Selection
    a. General framework for variable selection
    b. Stepwise procedures (forward and backward elimination, and combination)
    c. Criterion-based procedures
        i. Mallow's C_p statistic
        ii. Akaike Information Criterion (AIC)
        iii. Bayesian Information Criterion (BIC)
    d. LASSO

16. Qualitative Predictors
    a. Indicator variables for encoding qualitative predictors
    b. ANCOVA (mixed quantitative and qualitative predictors)
        i. Models with and without interaction between quantitative predictor and indicator variables
    c. ANOVA (with purely qualitative predictors): Single-Factor ANOVA Model I.
    d. Treatments, effects, and evaluation of the effects of treatments
        i. Pairwise t-tests, Bonferroni CI-s, Tukey's Honest Significant Difference

17. R implementation / simulations of all above (as the course proceeds) using R Studio, and knitr to generate write-ups.

The course has weekly home works with a mix of theoretical and computational problems.

We offer weekly optional review / advising sessions for extra help.

Tests: Two "in-class" closed-notes quizzes with theoretical questions, 45 minutes each. Two take-home exams, open notes and books, about 1/3 theoretical and 2/3 computational problems.