STAT 213: EXAM 2 STUDY GUIDE

COLIN REIMER DAWSON, SPRING 2016

(1) Multiple Regression

- Model-building and Assessment. You should be able to

  (a) write down regression models in the standard form from a verbal description of the model (see, e.g., many reading quizzes)

  (b) interpret coefficients, including those for binary indicator variables, interactions between indicators and quantitative predictors, interactions between two quantitative variables

  (c) identify degrees of freedom for models and residuals

  (d) reconstruct ANOVA tables from regression output

  - understand relationship between "residual standard error" and MSE in ANOVA

  - understand relationship between R^2 and SS components

  - understand relationship between adj. R^2, MSE, and $MS_{Total} = SS_{Total}/df_{Total}$

  (e) Conduct and interpret overall and nested F-tests

  - Will not need to compute SS from raw observations, but should understand relationships between different SS components

  (f) Identify regression conditions, and assess (some of) them from residual plots

  (g) Correctly interpret t-tests of individual predictors

  - concept of "controlling for" other predictors

  - "overall value" vs. "added value" of a predictor

(h) Interpret confidence and prediction intervals for the response variable in the MLR setting

(i) Recognize relationship between (and convert between) a MLR model with mutually exclusive indicator variables and a one-way ANOVA model.

- Multicollinearity

  (a) Understand what can happen to confidence intervals of individual coefficients if predictors are multicollinear

  (b) Use variance inflation factors to identify multicollinearity

  (c) Interpret a confidence ellipse

  (d) Suggest remedies for multicollinearity

(2) Model Selection

(a) Recognize and explain the problem with using unadjusted R^2 to compare models with different numbers of predictors

(b) Explain how adjusted R^2 attempts to address this issue

(c) Explain how Mallow's C_p addresses this issue, and how it differs from adjusted R^2 (Hint: is sensitive to predictors that have been "left out", not just those that are still in)

(d) Identify and explain differences between the four search algorithms for examining subsets of predictors to keep in the model

(e) Explain the goal of cross-validation and identify difference between in-sample and out-of-sample prediction error. What is overfitting and why is out-of-sample prediction error more reliable (think about the practice test analogy)

(3) Two-Way ANOVA: Basic concepts

(a) What does the two-way ANOVA model look like (in the $Y = f(\text{Predictors}) + \varepsilon$ form)?

(b) What is the interpretation of the α and β terms?

(c) What is the interpretation of the γ terms in an interaction model?

(d) In words, what does it mean for there to be an interaction?

(e) How many degrees of freedom for each factor, for the interaction, and for residuals, and why?