Study Questions For Predict 410

Topic: Logistic Regression

Our learning format requires that you complete the assigned readings efficiently and "intelligently". In order to help you focus your attention on important concepts in the course reading, we have constructed a list of study questions for each topic covered in PREDICT 410. You should preview each list of study questions before you begin your reading, and then answer the questions in a notebook while you are performing you reading. If you cannot answer a question, then you should look up the answer. If you cannot find the answer, then you should post a question in your Blackboard course shell.

(25) What is the response variable in a logistic regression? To what distribution does the response variable belong? When a logistic regression model is fit, what quantity serves as the left hand side of the regression equation? Is the logistic regression model a 'linear model', why or why not?

(26) Why is fitting a binary response variable with OLS regression statistically incorrect? What problems would this OLS model exhibit, i.e. how would the fitted model violate the OLS assumptions?

(27) How do you interpret the results from fitting a logistic regression model? In particular how do we interpret the predictor coefficient in logistic regression? How do we compute the p-values for each predictor coefficient, and how does this differ from the case of OLS regression?

(28) How do you compute and interpret the odds ratio for a predictor variable in logistic regression?

(29) What types of residuals are associated with logistic regression? Is the standard OLS residual r(i) = y(i) - yhat(i) a valid residual for logistic regression?

(30) How do we interpret the goodness-of-fit statistics output by SAS? This includes the percent concordant, the percent discordant, Somer's D, Gamma, and Tau-a.

(31) In logistic regression what quantity follows a binomial distribution and what quantity follows a logistic distribution?

(32) For logistic regression what is the analogue of the overall F-test? What is the statistic and what distribution does it follow?

(33) What graphical diagnostics can be used to assess model adequacy?

(34) What is the G statistic, what distribution does it have, and how is it used to compare logistic regression models?

(35) Three techniques for variable selection in logistic regression include: (1) a likelihood ratio test for nested models, (2) AIC, and (3) BIC. What are these methods? How are they defined, what statistical distributions do any test statistics follow, and how are they used for model selection?

(36) Overall Model Adequacy For Logistic Regression: How do we assess the overall effectiveness of a logistic regression model? In OLS regression we have the R-squared measure of overall adequacy. However, in logistic regression no proper definition of a R-squared statistic exists. Are there any statistics to address the question of overall model adequacy for logistic regression? Are there any "modeling approaches" that would address the overall effectiveness of a fitted logistic regression model?