December 4, 2020

Assignment 3

CS 624, Fall 2020

Due on December 10, 2020

Problem 1. We are interested in finding and assessing the effect sizes of the important predictors of the housing price classes. Analyze the Boston house price dataset available at http://lib.stat.cmu.edu/datasets/boston to answer this question. In particular, create an outcome factor variable of interest with three categories (Low, Midium and High) by using the 25^{th} and 75^{th} sample percentiles of the median value of owner occupied homes on 1000s.

Problem 2. We need to develop a likelihood ratio statistic for testing the hypothesis of proportional odds in ordinal logistic regression models with k ordered factor levels using a sample of n subjects and p covariates. These steps can help:

a) The general ordinal logistic regression model assumes different effect sizes for all distinct logistic regression models for cumulative log-odds:

$$\log\left(\frac{P(Y_i \le j \mid x_i)}{P(Y_i > j \mid x_i)}\right) = \beta_{oj} + \beta_{1j} x_{i1} + \ldots + \beta_{pj} x_{ip}, \; j = 1, 2, \ldots, k .$$

b) The proportional odds ordinal logistic regression model assumes different intercepts and same effect sizes for all distinct logistic regression models:

$$\log\left(\frac{P(Y_i \le j \mid x_i)}{P(Y_i > j \mid x_i)}\right) = \beta_{oj} + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}, \; j = 1, 2, \ldots, k .$$

c) The likelihood of the multinomial data with subject specific probabilities for each person for each factor level is: $L(\beta) = \prod_{i=1}^{n} [\pi_{0i}(\beta)]^{y_{0i}} [\pi_{1i}(\beta)]^{y_{1i}} \ldots [\pi_{ki}(\beta)]^{y_{ki}}$, where one of the y_{ji} equals one only if the observed outcome for the i-th subject was in the j-th factor level and zero otherwise.

d) We want to test the hypothesis:

$H_0:$ the model from part b) is correct

$H_A:$ the model from part a) is correct

e) The likelihood ratio statistics allows to test the hypothesis from part d):

$LRT = -2\ln\left(\frac{\max_{H_0} L(\beta)}{\max_{H_A} L(\beta)}\right) \sim \chi^2(d)$, where d equals the difference of the number of regression coefficients between models a) and b).