Assignment #7: Problem Set for Logistic Regression (50 points)

This assignment will be made available in both pdf and Microsoft docx format. Answers should be typed into the docx file, saved, and converted into pdf format for submission into Blackboard. **Color your answers in green so that they can be easily distinguished from the questions themselves.**

Throughout this assignment keep all decimals to four places, i.e. X.xxxx.

Foundations of Logistic Regression:

(1) (5 points) What values can the response variable Y take in logistic regression, and hence what statistical distribution does Y follow?
The values can take on a binary response, thus belonging to a binomial distribution, or they can take on a multiple responses (ordinal) and be part of a multinomial distribution.

(2) (5 points) How are the parameters estimated in logistic regression? Is this different from how the parameters are estimated in Ordinary Least Squares (OLS) regression?
The parameters are estimated in logistic regression using maximum likelihood estimation. The estimators agree most closely with the data (8). Yes, this is different from OLS.

(3) (5 points) How do we define a "residual" in logistic regression, and how is it computed?
A "residual" in logistic regression is defined as the difference between actual Y, which is 0 or 1, and the predicted probability for Y, and it is computed as 1-estimated probability.

Model 1: Let's consider the logistic regression model, which we will refer to as Model 1, given by

$$\log(pi / [1-pi]) = 0.25 + 0.32*X1 + 0.70*X2 + 0.50*X3 \qquad \text{(M1)},$$

where X3 is an indicator variable with X3=0 if the observation is from Group A and X3=1 if the observation is from Group B. The likelihood value for this fitted model on 100 observations is 850.

(4) (6 points) For X1=2 and X2=1 compute the log-odds ratio for each group, i.e. X3=0 and X3=1.

Group A
0.25 + 0.32*2 + 0.70*1 + 0.50*1 =
.25 + .64 + .70 + .5 = 2.09

Group B
0.25 + 0.32*2 + 0.70*1 + 0.50*0 =
.25 + .64 + .70 = 1.59

(5) (6 points) For X1=2 and X2=1 compute the odds ratio for each group, i.e. X3=0 and X3=1.

Group A

0.25 + 0.32*2 + 0.70*1 + 0.50*1

(exp.25) +(exp.64) + (exp.70) + (exp.5) = 6.8430

Group B

0.25 + 0.32*2 + 0.70*1 + 0.50*0 =

(exp.25) +(exp.64) + (exp.70) = 5.1943

(6) (6 points) For X1=2 and X2=1 compute the probability of an event for each group, i.e. X3=0 and X3=1. (Page 32)

Group A

0.25 + 0.32*2 + 0.70*1 + 0.50*1 =

.25 + .64 + .70 + .5 = 2.09

e^2.09 / 1 + e^2.09

8.0849 / 9.0849 = .89

Group B

0.25 + 0.32*2 + 0.70*1 + 0.50*0 =

.25 + .64 + .70 = 1.59

e^1.59/ 1 + e^1.59 =

4.9037 / 5.9037 = .8306

(7) (2 points) Using the equation for M1, compute the relative odds associated with X3, i.e. the relative odds ratio of Group B compared to Group A.

=e^B1

= e^.5

= 1.6487

(8) (5 points) Use the odds ratios for each group to compute the relative odds of Group B to Group A. How does this number compare to the result in Question #7. Does this make sense?

Group A

(exp.25) +(exp.64) + (exp.70) + (exp.5) = 6.8430

Group B

(exp.25) +(exp.64) + (exp.70) = 5.1943

6.8430 - 5.1943 = 1.6487

Both calculations get the same answer, which is the odds ratio of each group compared to each other.

Model 2: Now let's consider an alternate logistic regression model, which we will refer to as Model 2, given by

$$\log(pi / [1-pi]) = 0.25 + 0.32*X1 + 0.70*X2 + 0.50*X3 + 0.1*X4 \quad (M2),$$

where X3 is an indicator variable with X3=0 if the observation is from Group A and X3=1 if the observation is from Group B. The likelihood value from fitting this model to the same 100 observations as M1 is 910.

(9) (10 points) Use the G statistic to perform a likelihood ratio test of nested models for M1 and M2. State the hypothesis that is being tested, compute the test statistic, and test the statistical significance using a critical value for alpha=0.05 from Table A.3 on page 357 in *Regression Analysis By Example*. From these results should we prefer M1 or M2?

.1 * X4 is the covariate that is being tested and will be represented by b4. The null hypothesis that is being tested is H0: b4=0

Formula: G = -2ln [likelihood without the variable / likelihood with the variable]
Model G = -2ln [850/910] =.1364
6.7452- 6.8134 = .1364

The alpha values for this model are 9.35 and 11.14, and seeing that .1364 is far less than these values we cannot reject are null hypothesis, thus the smaller model is better and we prefer model M1.