Homework
Applied Logistic Regression

WEEK 3

Exercise 1:

Use the Myopia Study (MYOPIA.dta)

a. Using the results from **Week 2, Exercise 1**, compute 95 percent confidence intervals for the slope coefficient SPHEQ. Write a sentence interpreting this confidence.

b. Use Stata to obtain the estimated covariance matrix. Compute the logit and estimated logistic probability for a subject with SPHEQ = 2. Evaluate the endpoints of the 95 percent confidence intervals for the logit and estimated logistic probability. Write a sentence interpreting the estimated probability and its confidence interval.

Exercise 2:

Use the ICU study (icu.dta)

a. Using the results from **Week 1, Exercise 2, part (d),** compute 95 percent confidence intervals for the slope and constant term. Write a sentence interpreting the confidence interval for the slope.

b. Obtain the estimated covariance matrix for the model fit from **Week 1, Exercise 2, part (d)**. Compute the logit and estimated logistic probability for a 60-year old subject. Compute a 95 percent confidence intervals for the logit and estimated logistic probability. Write a sentence or two interpreting the estimated probability and its confidence interval.

Exercise 3:

Use the ICU study (icu.dta)

Use the ICU data and consider the multiple logistic regression model of vital status, STA, on age (AGE), cancer part of the present problem (CAN), CPR prior to ICU admission (CPR), infection probable at ICU admission (INF), and race (RACE).

a. The variable RACE is coded at three levels. Prepare a table showing the coding of the two design variables necessary for including this variable in a logistic regression model.

b. Write down the equation for the logistic regression model of STA on AGE, CAN, CPR, INF, and RACE. Write down the equation for the logit transformation of this logistic regression model. How many parameters does this model contain?

c. Write down an expression for the likelihood and log likelihood for the logistic regression model in part (b). How many likelihood equations are there? Write down an expression for a typical likelihood equation for this problem.

d. Using a logistic regression package, obtain the maximum likelihood estimates of the parameters of the logistic regression model in part (b). Using these estimates write down the equation for the fitted values, that is, the estimated logistic probabilities.

e. Using the results of the output from the logistic regression package used in part (d), assess the significance of the slope coefficients for the variables in the model using the likelihood ratio test. What assumptions are needed for the p-values computed for this test to be valid? What is the value of the deviance for the fitted model?

f. Use the Wald statistics to obtain an approximation to the significance of the individual slope coefficients for the variables in the model. Fit a reduced model that eliminates those variables with nonsignificant Wald statistics. Assess the joint (conditional) significance of the variables excluded from the model. Present the results of fitting the reduced model in a table.

g. Using the results from part (f), compute 95 percent confidence intervals for all coefficients in the model. Write a sentence interpreting the confidence intervals for the non-constant covariates.