Homework
Applied Logistic Regression

WEEK 6

Exercise 1:

From the ICU data, use as the outcome variable vital status (STA) and CPR prior to ICU admission (CPR) as a covariate.

a.   Demonstrate that the value of the log-odds ratio obtained from the cross-classification of STA by CPR is identical to the estimated slope coefficient from the logistic regression of STA on CPR. Verify that the estimated standard error of the estimated slope coefficient for CPR obtained from the logistic regression package is identical to the square root of the sum of the inverse of the cell frequencies from the cross-classification of STA by CPR. Use either set of computations to obtain 95% CI for the odds ratio. What aspect concerning the coding of the variable CPR makes the calculations for the two methods equivalent?

b.   For purposes of illustration, use a data transformation statement to recode, for this problem only, the variable CPR as follows: 4 = no and 2 = yes. Perform the logistic regression of STA on CPR (recoded). Demonstrate how the calculation of the logit difference of CPR = yes versus CPR = no is equivalent to the value of the log-odds ratio obtained in problem 1(a). Use the results from the logistic regression to obtain the 95% CI for the odds ratio and verify that they are the same limits as obtained in Exercise 1(a).

c.   Consider the ICU data and use as the outcome variable vital status (STA) and race (RACE) as a covariate. Prepare a table showing the coding of the two design variables for RACE using the value RACE = 1, white, as the reference group. Show that the estimated log-odds ratios obtained from the cross-classification of STA by RACE, using RACE = 1 as the reference group, are identical to estimated slope coefficients for the two design variables from the logistic regression of STA on RACE. Verify that the estimated standard errors of the estimated slope coefficients for the two design variables for RACE are identical to the square root of the sum of the inverse of the cell frequencies from the cross-classification of STA by RACE used to calculate the odds ratio. Use either set of computations to compute the 95% CI for the odds ratios.

d.   Create design variables for RACE using the method typically employed in ANOVA. Perform the logistic regression of STA on RACE. Show by calculation that the estimated logit differences of RACE = 2 versus RACE = 1 and RACE = 3 versus RACE = 1 are equivalent to the values of the log-odds ratio obtained in problem 1(c). Use the results of the logistic regression to obtain the 95% CI for the odds ratios and verify that they are the same limits as obtained in Exercise 1(c). Note that the estimated covariance matrix for the estimated coefficients is needed to obtain the estimated variances of the logit differences.

e.   Consider the logistic regression of STA on CRN and AGE. Consider CRN to be the risk factor and show that AGE is a confounder of the association of CRN with STA. Addition of the interaction of AGE by CRN presents an interesting modeling dilemma. Examine the main effects only and interaction models graphically. Using the graphical results and any significance tests you feel are needed, select the best model (main effects or interaction) and justify your choice. Estimate relevant odds ratios. Repeat this analysis of confounding and interaction for a model that includes CPR as the risk factor and AGE as the potential confounding variable.

f.   Consider an analysis for confounding and interaction for the model with STA as the outcome, CAN as the risk factor, and TYP as the potential confounding variable. Perform this analysis using logistic regression modeling and Mantel-Haenszel analysis. Compare the results of the two approaches.