BIA652D Multivariate Data Analysis

Logistic Regression of HATCO Dataset

GROUP5 Homework 6

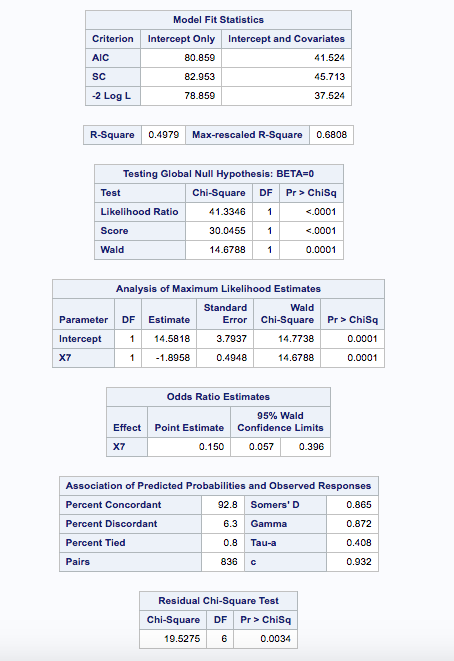
MEHRNOOSH OGHBAIE

2016

# **Summary of Logistic regression Analysis**

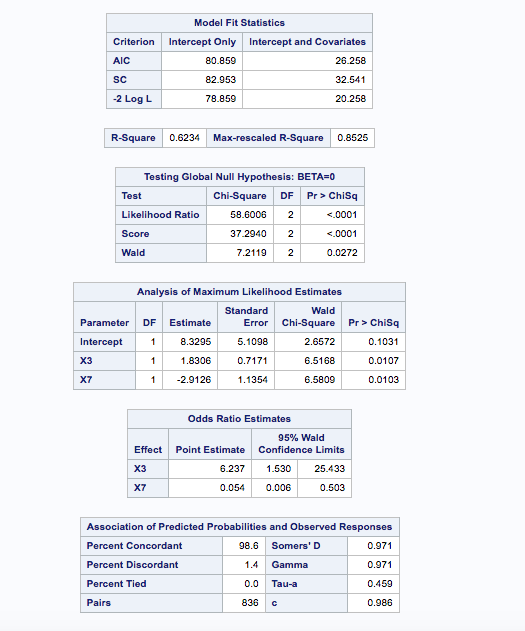
Here is the report on Logistic regression analysis on eight variables from HATCO database. Numerical values in X1-X7 and two levels factor in X11. In the first part of the assignment, the table is divided to 40 and 60 rows based on the value of column Split60.

Applying Proc Logistic, the value of -2 times the log of the likelihood is equal to 78.859. Wald logistic test P-value is (0.04) and the logistic coefficient is significance.

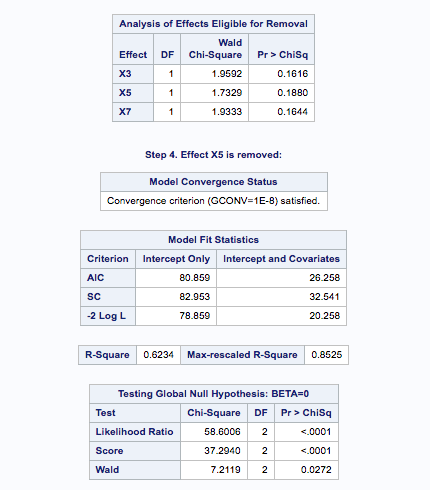


X7, X1, X3, X2 and X5 seem to be significant in eligible test. After entering X7 to the model, AIC, Schwarz Criterion and log likelihood are 41.524, 45.713 and 37.524. The Nagelkerke R-Square is 0.4979. The testing global Null hypothesis p-value for likelihood ratio, score and Wald is around zero showing that at least one variable is not equal to zero in overall model. The association of predicted probability and observed responses are pretty high (Sommers’D is 0.865).

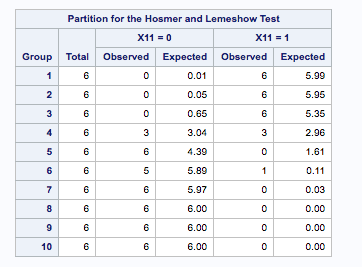
After removing the X7, the analysis of eligible entry shows that X3, X1, X4 and X5 are significant in addition to the model based on X7. After adding X3, the model fit statistics improved to AIC=26.254, SC= 32.541 and -2LogL =20.258. The R-Square is bigger (0.6234). The null hypothesis testing show very significance values (the likelihood ration around zero), that suggests there are not zero variables in the model. Both X3 and X7 seems to be significant in the maximum likelihood estimate. The associated of predicted probabilities and observed responses are all higher than previous model.



The analysis of eligible for entry shows X5 eligible. After adding X5 to the model, the model fit improves, but chi-square value of maximum likelihood changes doesn’t show significant p-value. Therefor X5 is about to be removed.



The hosmer and lemeshow doesn’t show significant p-value (0.24). So the model seems to be good. And it also shows when X11 expected value of 0 grows more than 1.



Finally the percentage of 0 and 1 are calculated for X11 for logisti60 and logistic40 from odd ratios and printed in the final table.

