

Homework

Applied Logistic Regression

WEEK 1

Exercise 1:

Use the Myopia Study (MYOPIA.dta)

One variable that is clearly important is the initial value of spherical equivalent refraction (SPHEQ).

- a. Write down the equation for the logistic regression model of SPHEQ on MYOPIA. Write down the equation for the logit transformation of this logistic regression model. What characteristic of the outcome variable, MYOPIA, leads us to consider the logistic regression model as opposed to the usual linear regression model to describe the relationship between MYOPIA and SPHEQ?
- b. Form a scatterplot of MYOPIA vs. SPHEQ.
- c. Write down an expression for the likelihood and log likelihood for the logistic regression model in part (a) using the ungrouped, $n = 618$, data. Obtain expressions for the two likelihood equations.
- d. Using Stata, obtain the maximum likelihood estimates of the parameters of the logistic regression model in part (a). These estimates should be based on the ungrouped, $n = 618$, data. Using these estimates, write down the equation for the fitted values, that is, the estimated logistic probabilities. Plot the equation for the fitted values on the axes used in the scatterplots in parts (b) and (c).

Exercise 2:

Use the ICU study (icu.dta)

The ICU data set consists of a sample of 200 subjects who were part of a much larger study on survival of patients following admission to an adult intensive care unit (ICU). The major goal of this study was to develop a logistic regression model to predict the probability of survival to hospital discharge of these patients. A number of publications have appeared which have focused on various facets of this problem.

- a. Write down the equation for the logistic regression model of STA on AGE. Write down the equation for the logit transformation of this logistic regression model. What characteristic of the outcome variable, STA, leads us to consider the logistic regression model as opposed to the usual linear regression model to describe the relationship between STA and AGE?
- b. Form a scatterplot of STA versus AGE.

- c. Write down an expression for the likelihood and log likelihood for the logistic regression model in part (a) using the ungrouped, $n = 200$, data. Obtain expressions for the two likelihood equations.
- d. Using Stata, obtain the maximum likelihood estimates of the parameters of the logistic regression model in part (a). These estimates should be based on the ungrouped, $n = 200$, data. Using these estimates, write down the equation for the fitted values, that is, the estimated logistic probabilities. Plot the equation for the fitted values on the axes used in the scatterplots in part (b).
- e. Summarize (describe in words) the results presented in the plot obtained from parts (b) and (d).