

Topics to be covered this week

Statistics 138
Fall Quarter, 2018

Wednesday, Nov 14 Multicategory logistic regression (Handout 11, and chap 6 from the text), Poisson regression (Handout 12, Section 3.3 from the text).

Friday, Nov 16 Cochran-Mantel-Haenszel statistic, Tests for homogeneous association (Handout 13, Section 4.3 from the text).

Homework 6 (Due: Monday, Nov 19)

You may form a group of three registered in this course and submit one completed homework for the group. The front page should be blank. Please write down the names of the students in the group on the submitted work.

1. For the dataset flu.xls, the columns are
shot ($Y = 1$ indicates flu shot, $Y = 0$ indicates no flu shot),
age (X_1 , age of the subject),
health awareness index (X_2 , higher values indicate higher awareness),
gender ($X_3 = 0$ female, $X_3 = 1$ male).

It is desired to model logit of π , the probability of getting the flu shot, as a linear function of age, awareness and gender, ie, $\pi' = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$, where π' is the logit of π .

- (a) Obtain the maximum likelihood estimates of $\beta_0, \beta_1, \beta_2$ and β_3 and the their standard errors. Write down the estimated logistic regression function.
- (b) Obtain $\exp(\hat{\beta}_1), \exp(\hat{\beta}_2)$ and $\exp(\hat{\beta}_3)$. Interpret these numbers.
- (c) Using part (a), estimate the probability that a randomly chosen 55 year male with health awareness index of 60 will receive a flu shot, along with a 95% confidence interval.
- (d) Carry out the likelihood ratio test to decide if variable X_3 can be dropped from the model. Write down the null and the alternative hypotheses, and then carry out a test at level $\alpha = 0.05$. Find the p-value of your test and state your conclusion.

2. Refer to Problem 1.

- (a) Use the likelihood ratio test to determine whether the following second order terms, square of age, square of health awareness index, and the interaction term between age and awareness index (ie, $X_1 X_2$), should be simultaneously added to the model containing age and awareness index as the first order terms. State the null and the alternative hypotheses, and carry out a test at level $\alpha = 0.05$. Find the p-value of your test and state your conclusion.
- (b) Consider the model which contains all the first order and all the second order terms in age and health awareness (including $X_1 X_2$). Use the backward stepwise regression method and the AIC criterion to select an appropriate model.

Write down the estimated logistic regression as well the standard errors of the estimated parameters.

(c) Use the model selected in part (b) to estimate the probability that a randomly chosen 55 year person with health awareness index of 60 will receive a flu shot, along with a 95% confidence interval. Comment on the difference (if any) in your answers from this part and part (c) of Problem 1.

3. A researcher in geriatrics designed a prospective study to investigate the effects of two interventions on the frequency of falls. One hundred subjects were randomly assigned to one of the two interventions: education only ($X_1 = 0$) and education plus aerobic exercise training ($X_1 = 1$). Subjects were at least 65 years of age and in reasonably good health. Three variables considered to be important as control variables were gender ($X_2 : 0 = \text{female}, 1 = \text{male}$), a balance index (X_3) and a strength index (X_4). The higher the balance index, the more stable is the subject; and higher the strength index, the stronger is the subject. Each subject kept a diary recording the number of falls (Y) during the six months of the study. The data are given in the file GeriatricStudy.xls. A Poisson regression model is considered appropriate here where the mean μ is modeled as $\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)$.

(a) Fit the Poisson regression model, and find the estimated regression coefficients, their standard errors, and the estimated regression function.

(b) Obtain the deviance residuals and plot them against the index ($i = 1, \dots, n$). Do there appear to be any outlying cases?

(c) Use the likelihood ratio test to determine whether gender (X_2) can be dropped from the model. Write down the null and the alternative hypotheses, and carry out a test at level $\alpha = 0.05$. State the null and the alternative hypotheses. Find the p-value of your test and state your conclusion.

(d) For the fitted model containing only X_1, X_3 and X_4 , obtain an approximate 95% confidence interval for β_1 . Does aerobic exercise reduce the frequency of falls when controlling for balance and strength? Explain.