Generalized Linear Models – Part IV

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 11 – Lecture 11E

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Revisiting Normal Response

If y_i , $i = 1, \ldots, n$, are independently normally distributed with constant variance and mean

$$E(y_i) = \mathbf{x}_i{}'\boldsymbol{\beta}$$

Then

$$D = \frac{SS_{\text{Res}}}{\sigma^2} \sim \chi^2_{n-p}$$

but σ^2 is unknown and needs to be estimated.

Goodness of Fit Statistic for Binary response

Goodness of fit can also be assessed by Pearson's chi-square statistic

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)} \qquad N = \sum n_i$$

which is chi-square distributed with df $= n - p$. Small values of this statistic imply that the model provides adequate fit to the data.

Goodness of Fit Statistic for Binary response

When there are no replicates on the regressors, the observations can be grouped to perform a goodness-of-fit test (called Homsner-Lemeshow test). Generally, group the observations into g=10 groups. Then, compare the observed number of successes O_j and number of failures, N_j-O_j, with expected frequencies in each group, $N_j \bar{\pi}_j$ and $N_j(1-\bar{\pi}_j)$, where

$$\bar{\pi}_j = \sum_{i=1}^{N_j} \hat{\pi}_{ij} / N_j$$

Goodness of Fit Statistic for Binary response

$$HL = \sum_{j=1}^{g} \frac{(O_j - N_j \bar{\pi}_j)^2}{N_j \bar{\pi}_j (1 - \bar{\pi}_j)}$$

If the fitted logistic regression model is correct, then HL follows a chi-square distribution with g-2 degrees of freedom when sample size is large. Large values of HL imply that the model is not an adequate fit to the data.

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING