Generalized Linear Models – Part III

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 11 – Lecture 11D

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Interpretation of Parameters in Logistic Regression Model

Suppose that there is only one regressor.

$$\hat{\eta}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

This image cannot currently be displayed.

$$\Rightarrow \quad \hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \hat{\beta}_1$$

Interpretation of Parameters in Logistic Regression Model

$\hat{\eta}(x_i)$ is the log odds when the regressor is x_i .

$$\hat{\eta}(x_i+1)-\hat{\eta}(x_i)=\ln(\text{odds}_{x_i+1})-\ln(\text{odds}_{x_i})$$

$$=\ln\left(\frac{\text{odds}_{x_i+1}}{\text{odds}_{x_i}}\right)=\hat{\beta}_1$$

$$\hat{O}_R=\frac{\text{odds}_{x_i+1}}{\text{odds}_{x_i}}=e^{\hat{\beta}_1}$$

Pneumoconiosis Data (Ex. 13.2)

$$\hat{y} = \frac{1}{1 + e^{+4.7965 - 0.0935x}}$$

$$\hat{O}_R = e^{\hat{\beta}_1} = e^{0.0935} = 1.10$$

Every additional year of exposure increases the odds of contracting a severe case of pneumoconiosis by 10%. If exposure increases by 10 years, the odds ratio becomes (more than double) $e^{10\hat{\beta}_1} = e^{10(0.0935)} = 2.55$

Statistical Inference on Model Parameters

Likelihood ratio (LR) tests

Can be used to compare a “full” model with a “reduced” model. This is analogous to the “extra sum-of-squares” technique.

The LR test procedure compares twice the logarithm of the value of the likelihood function for the full model (FM) to twice the logarithm of the value of the likelihood function of the reduced model (RM) to obtain a test statistic.

Likelihood Ratio Statistic

$$LR = 2\ln\frac{L(FM)}{L(RM)} = 2[\ln L(FM) - \ln L(RM)]$$

For large samples, when the reduced model is correct, the test statistic LR follows a chi-square distribution with degrees of freedom equal to the difference in the number of parameters between the full and reduced models. So, if LR exceeds $\chi^2_{\alpha,df}$, then we would reject the claim that the reduced model is appropriate.

Likelihood Ratio Statistic for Logistic Regression

For logistic regression, LR test can be used to provide a test for significance in logistic regression. For simplicity, let us say that we consider whether the logistic regression with two regressors x_1 and x_2 is significant, relative to that without the two regressors.

Without the two regressors, the model is the constant-probability-of-success model

$$E(y) = \pi = \frac{e^{\beta_0}}{1+e^{\beta_0}}$$

Likelihood Ratio Statistic for Logistic Regression

$$\ln L(RM) = y\ln(y) + (n-y)\ln(n-y) - n\ln(n)$$

With the two regressors,

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

$$i = 1, \ldots, n$$

After β's are estimated, $\hat{\pi}_i = \dfrac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}}}$

Likelihood Ratio Statistic for Logistic Regression

$$LR = 2\left\{\sum_{i=1}^{n} y_i \ln \hat{\pi}_i + \sum_{i=1}^{n} (n_i - y_i) \ln(1 - \hat{\pi}_i) - [y \ln(y) + (n - y)\ln(n - y) - n \ln(n)]\right\}$$

If $LR > \chi^2_{\alpha,2}$ then conclude that at least one of the regressors is important.

Likelihood Ratio Statistic for Logistic Regression

Testing Goodness of Fit

Can also be assessed using a LR test procedure. This test compares the current model to a saturated model where each observation (or group of observations when $n_i > 1$) is allowed to have its own parameter (that is, a success probability). These parameters or success probabilities are y_i/n_i, where y_i is # of successes and n_i is # of observations.

Deviance for Logistic Regression

Deviance is twice the difference in log-likelihoods between this saturated model and the full model (which is the current model) that has been fit to the data with estimated probability $\hat{\pi}_i = \exp\left(\mathbf{x}_i'\hat{\boldsymbol{\beta}}\right) / \left[1 + \exp\left(\mathbf{x}_i'\hat{\boldsymbol{\beta}}\right)\right]$

$$D = 2\ln \frac{L(\text{saturated model})}{L(FM)} = 2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{n_i\pi_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i(1-\hat{\pi}_i)}\right)\right]$$

Deviance for Logistic Regression

When the logistic regression model is an adequate fit to the data and the sample size is large, the deviance D has a chi-square distribution with $n - p$ degrees of freedom, where p is the number of parameters in the model. Small values of the deviance (or a large p-value) imply that the model provides a satisfactory fit to the data, while large values of D imply the current model is not adequate.

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING

© The Johns Hopkins University 2017, All Rights Reserved.