Generalized Linear Models – Part VI

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 12 – Lecture 12C

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Testing Individual Model Coefficients

$$H_0 : \beta_j = 0, \quad H_1 : \beta_j \neq 0$$

can be tested using a difference in deviance described previously.

Alternatively, t-like statistic can be constructed based on the estimated β_j and its standard error. The variances and covariances of the β's estimates can be obtained from the second partial derivatives of log-likelihood function.

Testing Individual Model Coefficients

$$Var(\hat{\beta}) = (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$$

The test statistic (Wald test) for H_0:

$$Z_0 = \frac{\hat{\beta}_j}{\mathrm{se}(\hat{\beta}_j)}$$

which is approximately standard normal

Interval Estimation

100(1-α) percent confidence interval for β_j :

$$\hat{\beta}_j - Z_{\alpha/2}\text{se}\left(\hat{\beta}_j\right) \le \beta_j \le \hat{\beta}_j + Z_{\alpha/2}\text{se}\left(\hat{\beta}_j\right)$$

100(1-α) percent confidence interval for O_{Rj} :

$$\hat{O}_{Rj} = \exp(\hat{\beta}_j)$$

$$\exp(\hat{\beta}_j - z_\alpha se(\hat{\beta}_j)) \le O_{Rj} \le \exp(\hat{\beta}_j + z_\alpha se(\hat{\beta}_j))$$

Pneumoconiosis Data (Ex 13.6, page 438)

$$\hat{\beta}_1 = 0.0934629$$

$$O_{R1} = \exp(\hat{\beta}_1) = 1.10$$

$$se(\hat{\beta}_1) = 0.0154258$$

95% CI for O_{R1} :

$$\exp[0.0934629 - 1.96(0.0154258)] \le O_R \le \exp[0.0934629 - 1.96(0.0154258)]$$

$$\exp(0.063228) \le O_R \le \exp(0.123697)$$

$$1.07 \le O_R \le 1.13$$

Linear Predictor

$$\mathbf{x}_0' = [1, x_{01}, x_{02}, \ldots, x_{0k}]$$

The linear predictor evaluated at $\mathbf{x}_0$ is

$$\hat{\eta} = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$$

$$\operatorname{Var}\left(\mathbf{x}_0'\hat{\boldsymbol{\beta}}\right) = \mathbf{x}_0'\operatorname{Var}\left(\hat{\boldsymbol{\beta}}\right)\mathbf{x}_0 = \mathbf{x}_0'(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_0$$

Interval Estimation for Linear Predictor

100(1-α) percent CI for η is

$$L(\mathbf{x}_0) \leq \mathbf{x}_0'\boldsymbol{\beta} \leq U(\mathbf{x}_0)$$

$$L(\mathbf{x}_0) = \mathbf{x}_0'\hat{\boldsymbol{\beta}} - Z_{\alpha/2}\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_0}$$

$$U(\mathbf{x}_0) = \mathbf{x}_0'\hat{\boldsymbol{\beta}} + Z_{\alpha/2}\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_0}$$

Interval Estimation for Response Probability

100(1-α) percent CI for π_0 is

$$L(\mathbf{x}_0) = \mathbf{x}_0'\hat{\boldsymbol{\beta}} - Z_{\alpha/2}\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_0}$$

$$U(\mathbf{x}_0) = \mathbf{x}_0'\hat{\boldsymbol{\beta}} + Z_{\alpha/2}\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_0}$$

$$\frac{\exp[L(\mathbf{x}_0)]}{1+\exp[L(\mathbf{x}_0)]} \le \pi_0 \le \frac{\exp[U(\mathbf{x}_0)]}{1+\exp[U(\mathbf{x}_0)]}$$

$$\hat{\pi}_0 = \exp\left(\mathbf{x}_0'\hat{\boldsymbol{\beta}}\right) / \left[1+\exp\left(\mathbf{x}_0'\hat{\boldsymbol{\beta}}\right)\right]$$

Pneumoconiosis Data (Ex 13.7, page 439)

A point estimate of the probability of miners with x = 40 years of exposure contracting pneumoconiosis is

$$\hat{\pi}_0 = \frac{e^{-4.7965+0.0935(40)}}{1+e^{-4.7965+0.0935(40)}} = \frac{e^{-1.0565}}{1+e^{-1.0565}} = 0.2580$$

$$\begin{aligned}\mathrm{Var}\left(\mathbf{x}_0'\hat{\boldsymbol{\beta}}\right) &= \mathbf{x}_0'\left(\mathbf{X}'\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{x}_0 \\ &= \begin{bmatrix}1 & 40\end{bmatrix}\begin{bmatrix}0.32383 & -0.0083480 \\ -0.0083480 & 0.0002380\end{bmatrix}\begin{bmatrix}1 \\ 40\end{bmatrix} = 0.036243\end{aligned}$$

Pneumoconiosis Data (Ex 13.7, page 439)

$$L(\mathbf{x}_0) = -1.0565 - 1.96\sqrt{0.036343} = -1.4296$$

$$U(\mathbf{x}_0) = -1.0565 + 1.96\sqrt{0.036343} = -0.6834$$

95% CI for this probability is

$$\frac{\exp(-1.4296)}{1+\exp(-1.4296)} \le \pi_0 \le \frac{\exp(-0.6834)}{1+\exp(-0.6834)}$$

$$0.1932 \le \pi_0 \le 0.3355$$

Diagnostic Checking in Logistic Regression

Three residuals (ordinary, deviance, Pearson):

$$e_i = y_i - \hat{y}_i = y_i - n_i\hat{\pi}_i, \quad i = 1, 2, \ldots, n$$

$$d_i = \pm\left\{2\left[y_i \ln\left(\frac{y_i}{n_i\hat{\pi}_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i(1-\hat{\pi}_i)}\right)\right]\right\}^{1/2}, \quad i = 1, 2, \ldots, n$$

$$r_i = \frac{y_i - n_i\hat{\pi}_i}{\sqrt{n_i\hat{\pi}_i(1-\hat{\pi}_i)}}, \quad i = 1, 2, \ldots, n$$

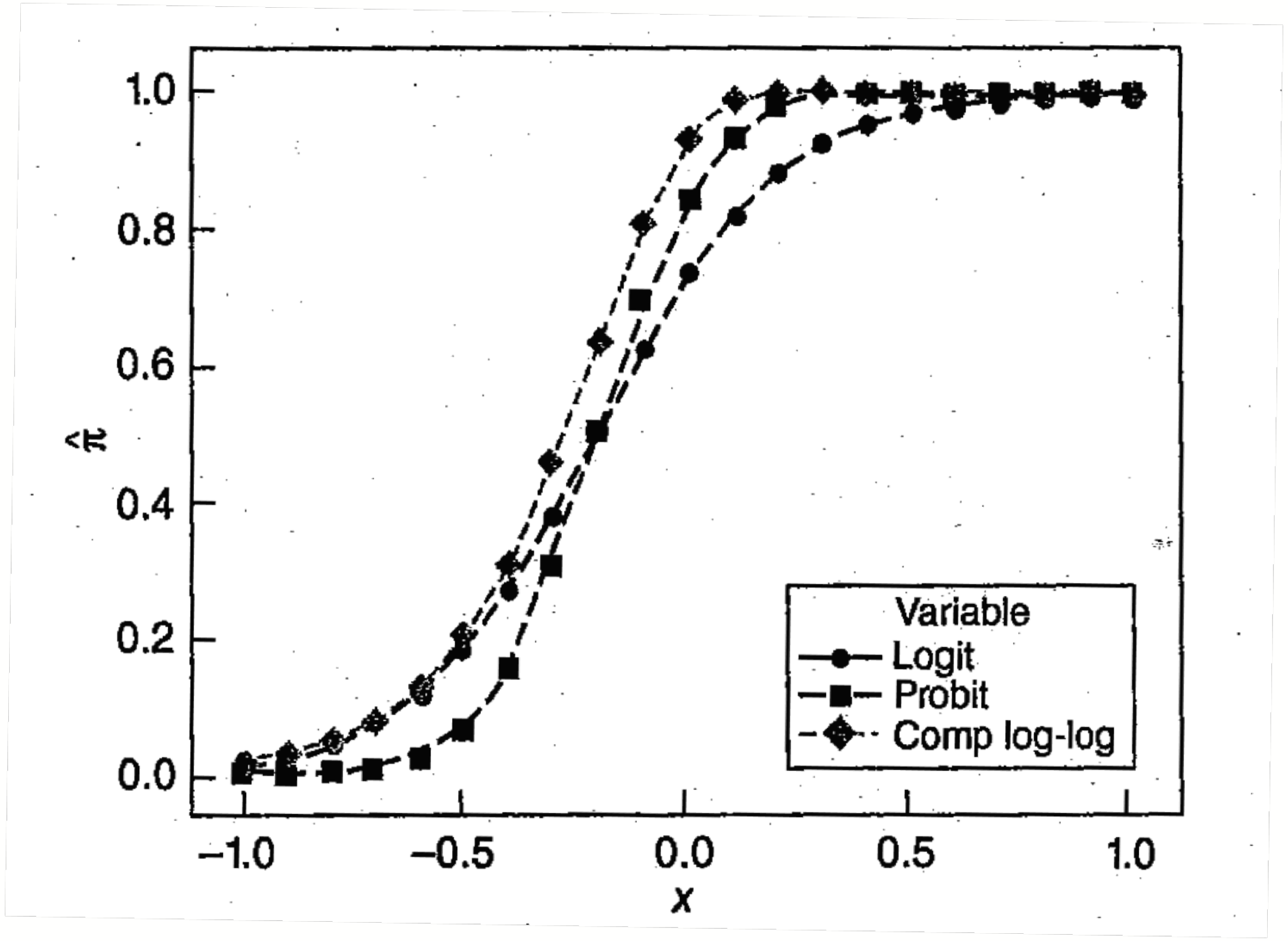
$$sr_i = \frac{r_i}{\sqrt{1-h_{ii}}} = \frac{y_i - n_i\hat{\pi}_i}{\sqrt{(1-h_{ii})n_i\hat{\pi}_i(1-\hat{\pi}_i)}}, \quad i = 1, 2, \ldots, n$$

Other Models for Binary Variable

Probit: $\pi = \Phi(\mathbf{x}'\boldsymbol{\beta})$

Complimentary log-log:

$$\pi = 1 - \exp[-\exp(\mathbf{x}'\boldsymbol{\beta})]$$



JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING