Generalized Linear Models – Part VII

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 12 – Lecture 12D

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Generalized Linear Model for Count Data

y: a count variable,

$y = 0, 1, 2, \ldots\ldots$ with probability density function:

$$f(y) = \frac{e^{-\mu}\mu^{y}}{y!}, \quad y = 0, 1, \ldots \qquad \mu > 0$$

$$E(y) = \mu \quad \text{and} \quad \mathrm{Var}(y) = \mu$$

Poisson Regression Model

$$y_i = E(y_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

$$E(y_i) = \mu_i$$

And there is a function g (called link function) such that

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = \mathbf{x}_i'\boldsymbol{\beta}$$

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i'\boldsymbol{\beta})$$

Poisson Regression Model

One of the link functions for Poisson regression is the **identity link**

$$g(\mu_i) = \mu_i = \mathbf{x}_i'\beta$$

Another popular link function is the **log link**:

$$g(\mu_i) = \ln(\mu_i) = \mathbf{x}_i'\beta$$

$$\mu_i = g^{-1}(\mathbf{x}_i'\beta) = e^{\mathbf{x}_i'\beta}$$

which ensures all the predicted values of the response variable will be nonnegative

Maximum Likelihood Estimation in Poisson Regression Model

$$L(\mathbf{y},\beta)=\prod_{i=1}^{n} f_i(y_i)=\prod_{i=1}^{n}\frac{e^{-\mu}\mu_i^{y_i}}{y_i!}$$

$$=\frac{\prod_{i=1}^{n}\mu_i^{y_i}\exp\left(-\sum_{i=1}^{n}\mu_i\right)}{\prod_{i=1}^{n}y_i!}$$

$$\mu_i=g^{-1}(\mathbf{x}_i'\beta)$$

$$\ln L(\mathbf{y},\boldsymbol{\beta})=\sum_{i=1}^{n}y_i\ln(\mu_i)-\sum_{i=1}^{n}\mu_i-\sum_{i=1}^{n}\ln(y_i!)$$

Maximum Likelihood Estimation in Poisson Regression Model

After obtaining MLE of $\boldsymbol{\beta}$,

$$\hat{y}_i = g^{-1}\left(\mathbf{x}_i'\hat{\boldsymbol{\beta}}\right)$$

The Aircraft Damage Data (Ex. 13.8, page 446)

Explore the relationship between y = # of locations where damage was inflicted on the aircraft and three possible regressors:

x_1= 1 (A-6 type of aircraft), 0 (A-4 type)
x_2 = bomb load (in tons)
x_3 = total months of aircrew experience

Data in Table 13.6 (p.447)

The Aircraft Damage Data (Ex. 13.8, page 446)

Poisson regression model with log link:

$$\ln(\mu_i) = -0.38 + 0.88x_1 + 0.14x_2 - 0.013x_3$$

	Estimate	SE	p-value
$\hat{\beta}_0$	-0.38	0.86	0.66
$\hat{\beta}_1$	0.88	0.50	0.079
$\hat{\beta}_2$	0.14	0.065	0.039
$\hat{\beta}_3$	-0.013	0.008	0.12

The Aircraft Damage Data (Ex. 13.8, page 446)

Use deviance to compare all possible models.

Model	Deviance	Difference in Deviance Compared to Full Model	P Value
$x_1x_2x_3$	28.4906		
x_1x_2	31.0223	2.5316	0.1116
x_1x_3	32.8817	4.3911	0.0361
x_2x_3	31.6062	3.1155	0.0775
x_1	38.3497	9.8591	0.0072
x_2	33.0137	4.5251	0.1041
x_3	54.9653	26.4747	<0.0001

The Aircraft Damage Data (Ex. 13.8, page 446)

$$\hat{y} = e^{-1.6491+0.2282x_2}$$

$D(\beta) = 33.0137$ with 28 degrees of freedom

P value is 0.2352

This model is an adequate fit

Generalized Linear Model in Exponential Family

$$y_i = E(y_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n \qquad E(y_i) = \mu_i$$

y_i distribution is a member of the exponential family with probability density function

$$f(y_i, \theta_i, \phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + h(y_i, \phi)\}$$

where ϕ is a scale parameter and θ_i is called the natural location parameter.

$$\mu = E(y) = \frac{db(\theta_i)}{d\theta_i}$$

$$\mathrm{Var}(y) = \frac{d^2 b(\theta_i)}{d\theta_i^2} a(\phi) = \frac{d\mu}{d\theta_i} a(\phi)$$

Generalized Linear Model in Exponential Family

Link predictor

$$\eta_i = g[E(y_i)] = g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$$

$$E(y_i) = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i'\boldsymbol{\beta})$$

If we choose $\eta_i = \theta_i$, then

we say that η_i is the **canonical link.**

Generalized Linear Model in Exponential Family

TABLE 13.8 Canonical Links for the Generalized Linear Model

Distribution	Canonical Link
Normal	$\eta_i = \mu_i$ (identity link)
Binomial	$\eta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$ (logistic link)
Poisson	$\eta_i = \ln(\lambda)$ (log link)
Exponential	$\eta_i = \frac{1}{\lambda_i}$ (reciprocal link)
Gamma	$\eta_i = \frac{1}{\lambda_i}$ (reciprocal link)

Parameter Estimation and Inference in Generalized Linear Model

Method of maximum likelihood is the basis for parameter estimation in general linear model. But it requires numerical algorithm to compute. The MLE $\hat{\beta}$ has the following properties:

$$E(\hat{\beta}) = \beta \quad \text{and} \quad \mathrm{Var}(\hat{\beta}) = a(\phi)(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$$

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING