Poisson Regression

1

Stat 566
5-29-25

Assume that the response variable Y is a count

Let Y have a Poisson distribution:

$$p(y) = \frac{e^{-\mu}\mu^{y}}{y!}, \quad y = 0, 1, 2, \ldots$$

Then $E[Y] = \mu$ and $Var[Y] = \mu$

Then we will fit the model

$$y_i = E[y_i] + \varepsilon_i \quad , i = 1, \ldots, n$$

2

Also, assume that some function of μ_i is a linear function of the predictors

So $\eta_i = g(\mu_i) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k = \vec{x}'\vec{\beta}$

$g(\mu_i)$ is the link function and must have an inverse.

Then $\mu_i = g^{-1}(\eta_i)$

2 common link functions:
$$\begin{cases} \text{Identity} : & g(\mu_i) = \mu_i \\ \text{Log} : & g(\mu_i) = \ln(\mu_i) \end{cases}$$

The log link function is often preferred,

since $\mu_i = e^{g(\mu_i)} = e^{\vec{x}_i' \vec{\beta}}$, which is

always positive (and Y represents a count)

To estimate the parameters, we use the

maximum likelihood method

$$f(y_1, \dots, y_n) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$$

$$= \prod_{i=1}^{n} \mu_i^{y_i} e^{-\sum_{i=1}^{n} \mu_i} \Big/ \prod_{i=1}^{n} (y_i!) = L(\vec{\beta})$$

Then $\ell(\vec{\beta}) = \ln L(\vec{\beta})$

$$= \sum_{i=1}^{n} y_i \ln(\mu_i) - \sum_{i=1}^{n} \mu_i - \sum_{i=1}^{n} \ln(y_i!),$$

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

As in logistic regression, there is no analytic solution, so the maximization is performed numerically in the stat software.

(5)

Two measures of goodness-of-fit:

① <u>Deviance</u>

$$D = 2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{\hat{y}_i}\right) - (y_i - \hat{y}_i)\right]$$

② <u>Pearson χ^2</u>

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i}$$

(6)

<u>Notes</u>: D is related to the likelihood ratio χ^2 statistic $2\sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{\hat{y}_i}\right)$

Both D and the Pearson χ^2 statistics have approximate χ^2 distributions with $n-p = n-(k+1)$ df