LINEAR REGRESSION ANALYSIS

MODULE – XIV

Lecture - 41

Logistic and Poisson Regression Models

Dr. Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Interpretation of parameters

To understand the interpretation of the related $\beta's$ in the logistic regression model, first consider a simple case with only one variable as

$$\eta(x) = \beta_0 + \beta_1 x.$$

After fitting of model, $\hat{\beta}_0$ and $\hat{\beta}_1$ are obtained as the estimators of β_0 and β_1 respectively. Then the fitted linear predictor at $x = x_i$ is

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

which is the log-odds at $x = x_i$. The fitted value at $x = x_i + 1$ is

$$\hat{\eta}(x_i + 1) = \hat{\beta}_0 + \hat{\beta}_1 (x_i + 1)$$

which is the log-odds at $x = x_i + 1$.

Thus

$$\begin{aligned}
\hat{\beta}_1 &= \hat{\eta}(x_i + 1) - \hat{\eta}(x_i) \\
&= \ln\left[\text{odds}(x_i + 1)\right] - \ln\left[\text{odds}(x_i)\right] \\
&= \ln\left[\frac{\text{odds}(x_i + 1)}{\text{odds}(x_i)}\right] \\
\Rightarrow \frac{\text{odds}(x_i + 1)}{\text{odds}(x_i)} &= \exp(\hat{\beta}_1).
\end{aligned}$$

This is termed as **odd ratio** which is the estimated increase in the probability of success when value of explanatory variable changes by one unit.

When there are more than one explanatory variables in the model, then the interpretation of β_j's is similar as in the case of single explanatory variable case. The odds ratio is $\exp(\hat{\beta}_j)$ associated with explanatory variable x_j keeping other explanatory variables constant. This is similar to the interpretation of β_j in multiple linear regression model.

If there is a m unit change is the explanatory variable, then the estimated increase in odds ratio is $\exp(m\hat{\beta}_j)$.

Test of hypothesis

The test of hypothesis for the parameters in the logistic regression model is based on asymptotic theory. It is a large sample test based on likelihood ratio test statistic termed as **deviance.**

A model with exactly p parameters that perfectly fits to the sample data is termed as **saturated model.**

The statistic that compares the log-likelihoods of fitted and saturated models is called as **model deviance**. It is defined as

$$\lambda(\beta) = 2\ln L(\text{saturated model}) - 2\ln L(\hat{\beta})$$

where $\ln L(\cdot)$ is the log-likelihood and $\hat{\beta}$ is the maximum likelihood estimate of β.

In case of logistic regression model, $y_i = 0$ or 1 and π_i's are completely unrestricted. So the likelihood will be maximum at $\pi_i = y_i$ and the maximum value of L(saturated model) is

$$Maximum\ L(\text{saturated model}) = 1$$

$$\Rightarrow \ln Maximum\, L(\text{saturated model}) = 0.$$

Let $\hat{\beta}$ be the maximum likelihood estimator of β, then log-likelihood is maximum at $\beta = \hat{\beta}$, and

$$\ln L(\hat{\beta}) = \sum_{i=1}^{n} y_i x_i' \hat{\beta}_i - \sum_{i=1}^{n} \ln\left[1 + \exp(x_i'\beta)\right]$$

$$\geq \ln L(\text{saturated model}).$$

Assuming that the logistic regression function is correct, the large sample distribution of likelihood ratio test statistic $\lambda(\beta)$ is approximately distributed as $\chi^2(n-p)$, when n is large.

Assuming that the logistic regression function is correct, the large sample distribution of likelihood ratio test statistic $\lambda(\beta)$ is approximately distributed as $\chi^2(n-p)$, when *n* is large.

Large value of $\lambda(\beta)$ implies model is incorrect. Small value of $\lambda(\beta)$ implies that model is well fitted and is as good as the saturated model. Note that generally the fitted model will be having smaller number of parameters than the saturated model that is based on all the parameters. Thus at $\alpha\%$ level of significance.

$\lambda(\beta) \leq X^2_{n-p}(\alpha)$ => fitted model is adequate.

$\lambda(\beta) > X^2_{n-p}(\alpha)$ => fitted model is not adequate.

Poisson regression model

The usual regression model is based on the assumption that the random errors are normally distributed and hence the study variable is normally distributed. In case, the study variable is a dichotomous variable taking only binary values, viz., 0 and 1, then logistic regression is used where study variable follows a Bernoulli distribution.

Similarly, we consider the situations where the study variable is a count variable that represents the count of some relatively rare event. For example, the study variable can be a count of patients with some rare type of disease with one or more explanatory variables like age of variables, hemoglobin level, blood sugar etc. In an another example, the study variable can be the number of defects in the car engine of a reputed car maker which again depends on one or more explanatory variables.

Assumption of normal or Bernoulli distribution for study variable will not be appropriate in such situations. The Poisson distribution describes such situations more appropriately. So we assume that the study variable y_i is a count variable and follows a Poisson distribution with parameter $\lambda > 0$ as

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

Note that the mean and variance of a Poisson random variable are same and related as

$$E(y) = \lambda, Var(y) = \lambda.$$

Based on a sample $y_1, y_2, ..., y_n$, we can write

$$E(y_i) = \lambda$$

and express the Poisson regression model as

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

where ε_i's are disturbance terms.

We can define a link function g that relates to the mean of study variable to a linear predictor as

$$\begin{aligned} g(\lambda_i) &= \eta_i \\ &= \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k \\ &= x_i' \beta \end{aligned}$$

and

$$\begin{aligned} \lambda_i &= g^{-1}(\eta_i) \\ &= g^{-1}(x_i' \beta). \end{aligned}$$

where $x_{1i} = 1$ for all $i = 1, 2, \ldots, n$ will denote the intercept term.

The **identity link function** is

$$g(\lambda_i) = \lambda_i = x_i' \beta.$$

The **log-link function** is

$$\begin{aligned} & g(\lambda_i) = \ln(\lambda_i) = x_i' \beta \\ \Rightarrow\ & \lambda_i = g^{-1}(x_i' \beta) = \exp(x_i' \beta). \end{aligned}$$

Note that in identity link function, the predicted values of *y* can be negative but in log-link function, the predicted values of *y* are nonnegative.

Maximum likelihood estimation of parameters

We use the method of maximum likelihood estimation to estimate the parameters of the Poisson regression model. The likelihood function is based on Poisson distribution with parameter λ and then β 's are estimated through the link function.

The likelihood function of $y_1, y_2, ..., y_n$ is

$$L(y,\lambda) = \prod_{i=1}^{n} p_i(y_i)$$

$$= \prod_{i=1}^{n} \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}$$

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

$$\ln L(y,\lambda) = \sum_{i=1}^{n} y_i \ln(\lambda_i) - \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n} \ln(y_i!) .$$

The parameter λ_i can be related to $\beta's$ through the link function

$$\lambda_i = g^{-1}(x_i'\beta).$$

After choosing the proper link function, the log-likelihood function can be maximized using some numerical optimization techniques for a given set of data. Let $\hat{\beta}$ be the obtained maximum likelihood estimator of β.

Then the fitted Poisson regression model is

$$\hat{y}_i = g^{-1}(x_i'\hat{\beta}).$$

In case of **identity link,**

$$\hat{y}_i = g^{-1}(x_i'\beta) = x_i'\beta.$$

In case of **log-link,**

$$\hat{y}_i = g^{-1}(x_i'\hat{\beta}) = \exp(x_i'\hat{\beta}).$$

Testing of hypothesis

The test of hypothesis is case of Poisson regression model is similar to the case of logistic regression model. It is constructed as **model deviance** which is based on large sample test using likelihood ratio test statistic.

The model deviance is defined as

$$\lambda^*(\beta) = 2\ln L(\text{saturated model}) - 2\ln L(\hat{\beta})$$

where saturated model is based on all the *p* parameters of the model and it fits to the data perfectly.

The statistic $\lambda^*(\beta)$ has approximately $\chi^2(n-p)$ distribution when *n* is large. The large value of $\lambda^*(\beta)$ indicates that the model is not correctly fitted to the given data whereas small values of $\lambda^*(\beta)$ indicate that model is well fitted to the given set of data in the sense that it is as good as the saturated model.

If $\lambda(\beta) \le \chi^2_{n-p}(\alpha) \Rightarrow$ fitted model is adequate

and if $\lambda(\beta) > \chi^2_{n-p}(\alpha) \Rightarrow$ fitted model is not adequate

at $\alpha\%$ level of significance.