Chapter 15

Poisson Regression Models

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

$$\begin{aligned} g(\lambda_i) &= \eta_i \\ &= \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \\ &= x_i'\beta \end{aligned}$$

and

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

The **identity link function** is

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

The **log-link function** is

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

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 $\beta's$ are estimated through the link function.

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

$$L(y,\lambda) = \prod_{i=1}^{n} p_i(y_i)$$
$$= \prod_{i=1}^{n} \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}$$
$$= \frac{\left(\prod_{i=1}^{n} \lambda_i^{\mu_i}\right)\left(\exp\left(-\sum_{i=1}^{n}\lambda_i\right)\right)}{\prod_{i=1}^{n} y_i!}$$

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

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.