Introduction to Poisson Regression with Robust Standard Errors - Part 1

Christopher Liu, Siyu Zhao, Zheming Yu, Runnan Liu, Shiming Zhang

11/29/2018

San Diego State University, Stats 610

The Poisson Distribution

A random variable Y is said to have a Poisson distribution with parameter λ if its probability is given by the probability mass function

$$Pr(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!}$$

for $\lambda > 0$ and y = 0, 1, 2, ...

The mean and variance of this distribution can be shown to be

$$E(Y) = Var(Y) = \lambda$$

The Poisson Distribution

A random variable Y is said to have a Poisson distribution with parameter λ if its probability is given by the probability mass function

$$Pr(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!}$$

for $\lambda > 0$ and y = 0, 1, 2, ...

The mean and variance of this distribution can be shown to be

$$E(Y) = Var(Y) = \lambda$$

Introduction to Poisson Regression

In Poisson Regression:

- Model used when the desired response variable, Y_i, is a count (eg. Number of vehicle accidents per year, number of visits to a website over a certain time span, etc)
- We can also have the response variable be Y_i/t, the rate at which the event happens with t being an interval representing time, space, or some other grouping of interest

Introduction to Poisson Regression

- The regression model with the log link function:

$$log(\lambda_i|X_i) = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} = X_i\beta$$

  where $E(Y_i|X_i) = \lambda_i = e^{X_i\beta}$

- Predictor variables are estimated by maximizing the likelihood function:

$$L(\beta) = \prod_{i=1}^{n} f(Y_i) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{Y_i}}{Y_i!}$$

Take the simple case:

$$log(\lambda_i|x) = \beta_0 + \beta_1 x$$

Consider the difference between the mean response given $(x + 1)$ and the mean response given x:

$$log(\lambda_i|x + 1) - log(\lambda_i|x)$$

$$= \beta_0 + \beta_1(x + 1) - (\beta_0 + \beta_1 x) = \beta_1$$

$$\implies \frac{(\lambda_i|x + 1)}{(\lambda_i|x)} = e^{\beta_1}$$

Poisson Regression with Robust Standard Errors

Our focus, however, is to discuss **Poisson Regression with Robust Standard Errors**

- Modified Poisson Regression that can work with response variables with binary outcomes
- Addresses problems with overdispersion

Poisson Regression with Robust Standard Errors

Main problem is with the Poisson assumption of

$$E(Y_i) = Var(Y_i)$$

- With binomial data, Poisson regression usually underestimates variance of data
- We need a way to address this problem when making inferences

Poisson Regression with Robust Standard Errors

The only adjustment to make:

- The standard errors of the estimated predictors ($\hat{\beta}$) are replaced with "robust" standard errors from the sandwich estimator

$$Var(\hat{\beta}) = (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1}$$

where Σ is the covariance matrix of the residuals

Poisson Regression with Robust Standard Errors

- Simple to implement through software such as R with the "sandwich" library
- Everything else is kept the same, such as the values of $\hat{\beta}$ (themaximum likelihood estimates of β) as well as their interpretations

Regression for Data with Binary Outcomes

Multiple methods for regression of data with binary outcomes

- Logistic Regression is a popular alternative

Many real-world scenarios can be modeled with binary regression

- Health-related problems such as association of disease and certain factors
- Widely used methods in Epidemiology and Public Health

Topics of Videos in this Series

- Example of implementing Poisson Regression with Robust Standard Errors on real data
- Interpretation of analysis and model parameters
- Diagnosis of models- how do we know it is working?