Introduction to Poisson Regression with Robust Standard Errors - Part 2

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

11/29/2018

San Diego State University, Stats 610

Objectives of the Video

- Develop a theoretical basis for the Poisson Regression Model
- Understand the interpretation of the model's parameters
- Understand the uses and limitations of this specific regression model

Recap of the Poisson Distribution

Recall $Y \sim Poisson(\lambda)$ if:

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

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

Also its mean and variance is given by:

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

Introduction to the Poisson Regression Model

The Poisson Regression model:

- Used to model situations where the desired output, Y_i, is a count of something
- Y_i is a random variable with a Poisson distribution
- Examples include the number of vehicle accidents per year or the number of visits to a website over a certain time span

Generalized Linear Models

The Poisson Regression Model is a **Generalized Linear Model** (GLM) with the general form:

$$f(\beta, X) = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} = X_i \beta$$

$f(\beta, X)$ is called the **Link Function** and relates the expected value of the output variable, Y_i, to the linear equation

Poisson Regression as a Generalized Linear Model

The Poisson Regression model uses the log link function:

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

where $log(\lambda_i|X_i) = log[E(Y_i|X_i)]$

Estimating the Predictor Variables

The predictor variables for the regression equation are estimated by maximizing the likelihood function:

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

We call them the maximum likelihood estimates and denote as $\hat{\beta}$

Interpreting the Parameters

Given our regression model

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

with the predictor variables $\beta_0, \beta_1, ..., \beta_p$

- How do we interpret the β_i's with respect to our data?

Interpreting the Parameters

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

Interpreting the Parameters

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

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

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

$$= \beta_1$$

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

Parameter Interpretation Example

Ex. Assume our model gives:

$$\frac{(\lambda_i|x+1)}{(\lambda_i|x)} = 1.15$$

- $(\lambda_i|x+1)$ is 15 percent greater than $(\lambda_i|x)$
- An increase in x by 1 unit, with every other predictor held constant, increases the expected value/count by .15

In the next video

In the next video:

- Tutorial of how to apply the Poisson Regression model to real data in R
- Working example of interpreting the output in terms of the theory just discussed