Count Data Models

Jacob Hochard & Ramjee Acharya

1

Count Data Models

- What is Count Data Model?

Very often data take the form of non-negative integer values such as number of children, number of accidents, visits to doctor, or number of students. Count Data Model exploits this feature of data for estimation. The most common model is Poisson.

2

Count Data Models

- Types
  - Poisson Model
  - Negative Binomial Model
  - Zero-Inflated Count Model
  - Zero-Truncated Count Model
  - Hurdle Model
  - Random-Effects Count Model

.

3

Why Count Data Model?

- Count data variables are dependent variables.
- OLS may give non-integer values or negative number.
- Poisson and negative Binomial are the popular Count Data Models'.

4

Poisson Distribution

What is a Poisson Distribution?

Eg: Suppose a man gets four calls days on an average; sometimes he gets more and sometime none. What is likely that the count will be five?

- Assumption- The variance of the number of occurrences equals the expected number of occurrences: E(Y)=VAR(Y)=λ

5

Poisson Models

$$\Pr(Y = y_i) = \frac{EXP^{-\lambda_i}\lambda_i^{y_i}}{y_i!};\ y = 0,1,2,...$$

where;

$$E[y_i] = \lambda_i = EXP(\beta X_i)$$

$$\ln(\lambda_i) = \beta' X_i$$

6

Poisson Models

By substituting E[y] = EXP (βX) in the expression, one easily obtains the likelihood function for all observations:

$$L(\beta) = \prod_i \frac{EXP\left[-EXP(\beta X_i)\right]\left[EXP(\beta X_i)\right]^{y_i}}{y_i!}$$

And the log-likelihood is simply:

$$LL(\beta) = \sum_{i=1}^{n}\left[-EXP(\beta X_i) + y_i \beta X_i - LN(y_i!)\right]$$

7

Estimates

- Consistent
- Inefficient

Therefore, robust standard error is used to correct the standard error of the parameters.

8

Poisson Models, GOF Measures

A measure similar to R-square is given as

$$R_p^2 = 1 - \frac{\sum_{i=1}^{n}\left[\frac{y_i - \hat{\lambda}_i}{\sqrt{\hat{\lambda}_i}}\right]^2}{\sum_{i=1}^{n}\left[\frac{y_i - \bar{y}}{\sqrt{\bar{y}}}\right]^2}$$

- Why not R-square?
  - Since the conditional mean function is non-linear and the regression is heteroscedastic

9

Poisson's Restriction

The Poisson distribution has one parameter, λ, which represent the distribution mean and variance.

- Often in real data the variance is not equal to the mean (e.g. statistically), and the Poisson model is not appropriate for the count process.
- For variance>mean, the over-dispersion, we need to use Negative Binomial Regression Model.

10

Test for Over-dispersion

A test by Cameron and Trivedi (1990)

$$H_0: \ VAR[y_i] = E[y_i]$$
$$H_A: \ VAR[y_i] = E[y_i] + \alpha g(E[y_i])$$

Do t-test for testing the over-dispersion.

$$Z_i = \frac{(y_i - E(y_i))^2 - y_i}{E(y_i)\sqrt{2}}$$

11

Negative Binomial Models:

$$\lambda_i = EXP(\beta' x_i + \varepsilon_i)$$

where;

EXP^{ε_i} is gamma distributed with mean = 1 and variance α

We introduce unobserved heterogeneity in the error term

12

Negative Binomial Model:

The model has an additional parameter alpha, such that:

$$VAR(y_i) = E[y_i]\{1 + \alpha E(Y_i)\}$$

When α= 0, the model "collapses" to the Poisson model.

13

Does Green want me to do this?

http://www.youtube.com/watch?v=5N4EFVgtB0Y&feature=player_embedded