Vanderbilt University
Leadership, Policy and Organizations
Class Number 9553
Spring 2021

**Models for Limited Outcomes**

# Limited Outcomes

Some data present not as a continuous variable, but in a limitd number of outcomes.

One type of limited data is count data: values can take only integer values, and typically over a quite limited range.

Another type is truncated data, where we only observe the values of the dependent variable when they fall within a range.

# Count Outcomes

## Econometric Model

$$Prob(Y_i = y_i|x_i) = \frac{e^{-\lambda_i}\lambda_i^y}{y_i!}, y_i \in 0,1,2 \tag{1}$$

$\lambda$ is our expression of the mean, given by:

$$ln\lambda_i = \boldsymbol{x}_i\boldsymbol{\beta} \tag{2}$$

This model assumes that variance is equal to the mean, which is super weird. In general, we favor the negative binomial model, which allows for overdispresion (mean not equal to variance).

$$ln\mu_i = \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i = ln\lambda_i + \ln\mu_i \tag{3}$$

The functional form for the negative binomial regression becomes:

$$f(y_i|\boldsymbol{x}_i,\mu_i) = \frac{e^{-\lambda_i,\mu_i}(\lambda_i,\mu_i)_i^y}{y_i!}, y_i = 0,1,2 \tag{4}$$

## Estimation

The key for interpretation is to think about how to use the outcome: are you interested in the probability of a given number of events or a count, or both? We'll review marginal effects, $pr(Y_i = y_i)$ and $pr(y_j < Y_i < y_k)$.

### Quick Exercise

Using the full negative binomial estimates (nbreg_full) provide predicted probabilities of applying to at least 2 colleges across the range of math scores(bynels2m) for both white and Hispanic students.

There are two other options when you have a lot of 0, each of which are analagous to the other two.

The zero-inflated poisson model estimates a two-part model, one for 0s and another for counts above 0.

## Truncated Outcomes

### Econometric Model

The tobit models assumes a latent variable $y^*$ which can only be seen when it crosses a given value $a$, where $a$ is many times 0. It can be generalized to be in a range, where the observade value $y_i$ is greater than $a$ but less than $b$. For the case where it's limited from below by $a$:

$$y_i^* = \boldsymbol{x}_i \boldsymbol{\beta} + \epsilon_i \tag{5}$$

$$y_i = \begin{cases} a & \text{if } y^* <= a \\ y_i & \text{if } y^* > a \end{cases} \tag{6}$$

To get the expected value of $y$:

$$E(y_i|x) = \Phi(\frac{\boldsymbol{x}_i \boldsymbol{\beta}}{\sigma})(\boldsymbol{x}_i \boldsymbol{\beta} + \sigma\lambda) \tag{7}$$

Where:

$$\lambda_i = \frac{\phi(\boldsymbol{x}_i \boldsymbol{\beta}/\sigma)}{\Phi(\boldsymbol{x}_i \boldsymbol{\beta}/\sigma)} \tag{8}$$

### Estimation

The key for interpretation here is whether you're interested in the probability that $y_i$ is above $a$, or the expected value of $y_i$ if it is above $a$ , or $y^*$. Each has a different meaning. We'll talk about how to get estimates of all 3.