Negative Binomial Regression Analysis

And other count models

Asst. Prof. Nikom Thanomsing
Department of Biostatistics & Demography
Faculty of Public Health, Khon Kaen University
Email: nikom@kku.ac.th
Web: http://home.kku.ac.th/nikom

Negative Binomial Regression Analysis

Negative Binomial Regression (NB)

- The earliest definitions of the negative binomial are based on the binomial PDF.
- NB2 (Cameron and Trivedi, 1986), NB2 is derived from a Poisson–gamma mixture distribution.
- NB1, The NB1 model can also be derived as a form of Poisson–gamma mixture, but with different properties resulting in a linear variance.
- The negative binomial model, as a Poisson–gamma mixture model, is appropriate to use when the overdispersion in an otherwise Poisson model is thought to take the form of a gamma shape or distribution.
- A more general class of negative binomial models with mean $\mu_i$ and variance function ($\mu_i + \alpha \mu_i^p$). NB2 with $p = 2$, NB1 with $p=1$.

Negative Binomial Regression Analysis

Negative Binomial Regression (NB2)

- The log-likelihood function for NB2

$$
\ln L(\alpha, \beta) = \sum_{i=1}^{n} \left[ \left( \sum_{j=0}^{y_i} \ln (j + \alpha^{-1}) \right) - \ln y_i! \right] \\
- (y_i + \alpha^{-1}) \ln (1 + \alpha \exp(x_i' \beta)) - y_i \ln \alpha + y_i x_i' \beta
$$

- NB1, The NB1 model can also be derived as a form of Poisson–gamma mixture, but with different properties resulting in a linear variance.
- The negative binomial model, as a Poisson–gamma mixture model, is appropriate to use when the overdispersion in an otherwise Poisson model is thought to take the form of a gamma shape or distribution.

Negative Binomial Regression Analysis

Negative Binomial Regression (NB2): Example


Sample: Licensed Nurse $n=52$
- bed = number of beds in home,
- tdays = annual total patient days (in hundreds)
- pcrev = annual total patient care revenue (in $ millions)
- nsal = annual nursing salaries (in $ millions)
- fexp = annual facilities expenditures (in $ millions)
- rural = (1 = rural; 0 = nonrural)
Negative Binomial Regression Analysis

Negative Binomial Regression (NB2): Interpretation using percentage

For a factor xk, the expected count increases (decreases) by n% \( \exp(\beta_k) \times 100 \), holding all other variables constant.

For a unit change in xk, the expected count increases by a factor of \( \exp(\beta_k) \), holding all other variables constant.

Alternatively, the percentage change in the expected count for a unit change in xk, holding other variables constant.

Methods of interpretation based on \( E(y|x) \)

\[ E(y|x, x_k + \delta) - E(y|x) = \exp(\delta) \times E(y|x) \]

\[ \frac{E(y|x, x_k + \delta)}{E(y|x)} = \exp(\delta) \]

The interpretation

For a change of \( \delta \) in \( x_k \), the expected count increases by a factor of \( \exp(\beta_k \times \delta) \), holding all other variables constant.

For specific values of \( \delta \)

Factor change. For a unit change in \( x_k \), the expected count changes by a factor of \( \exp(\beta_k) \), holding all other variables constant.

Standardize factor change. For a standard deviation change to \( x_k \), the expected count changes by a factor of \( \exp(\beta_k \times \text{std} \text{dev}) \), holding all other variables constant.

\[ \exp(\beta_k \times \text{std} \text{dev}) \]

Negative Binomial Regression Analysis

Negative Binomial Regression (NB2): glm

Methods of interpretation based on \( E(y|x) \)

\[ \frac{E(y|x, x_k + \delta)}{E(y|x)} = \exp(\delta) \times IRR \]

The interpretation

For a change of \( \delta \) in \( x_k \), the expected count increases by a factor of \( \exp(\beta_k \times \delta) \), holding all other variables constant.

Negative Binomial Regression Analysis

Negative Binomial Regression (NB2): Interpretation using the rate

Negative Binomial Regression Analysis
### Negative Binomial Regression Analysis

**Negative Binomial Regression (NB2): Interpretation using the rate**

#### Negative Binomial Regression (NB1)

- **Interpretation based on Incidence rate ratio**
  - Being a annual total patient care revenue decreases the expected number of beds in home by .6792, holding all other variables constant.
- **Interpretation based on percentage**
  - Being a annual total patient care revenue decreases the expected number of beds in home by 32.1%, holding all other variables constant.

#### Negative Binomial Regression (NB2)

| Coef. | Std. Err. | z | P>|z| |
|-------|-----------|---|-----|
| constant | 0.03003 | 0.00815 | nel | 0.000 |

- **e^bStdX = exp(b*SD of X) = change in expected count for SD increase in X**
- **P>|z| = p-value for z-test**
- **b = raw coefficient**
- **alpha | 0.03003 SE(alpha) = 0.00815**

---

### Negative Binomial Regression Analysis

**Negative Binomial Regression (NB2): Interpretation using the rate**

#### Negative Binomial Regression (NB1) using Poisson–gamma mixture

- The NB1 model can also be derived as a form of Poisson–gamma mixture, but with different properties resulting in a linear variance.
- The NB1 model, which sets $p = 1$, is also of interest because it has the same variance function, $(1 + \alpha)\mu$, as the used in the GLM approach.
- The NB1 log-likelihood function is

$$
\ell(\alpha, \beta) = \sum_i \left[ y_i \log(\alpha + \beta y_i) - \alpha - \log(y_i!) \right] = \sum_i \left[ (y_i + 1) \log(\alpha + \beta y_i) - \alpha - \log(y_i!) \right]
$$

- In a linear variance.

---

### Negative Binomial Regression Analysis

**Negative Binomial Regression (NB2): Interpretation using the rate**

#### Negative Binomial Regression (NB1) using Poisson–gamma mixture

- The NB1 model can also be derived as a form of Poisson–gamma mixture, but with different properties resulting in a linear variance.
- The NB1 model, which sets $p = 1$, is also of interest because it has the same variance function, $(1 + \alpha)\mu$, as the used in the GLM approach.
- The NB1 log-likelihood function is

$$
\ell(\alpha, \beta) = \sum_i \left[ y_i \log(\alpha + \beta y_i) - \alpha - \log(y_i!) \right] = \sum_i \left[ (y_i + 1) \log(\alpha + \beta y_i) - \alpha - \log(y_i!) \right]
$$

- In a linear variance.

---
Problem of Zero in Counts Model

Count response models having for more zeros than expected by distributional assumptions of Poisson and Negative binomial models result incorrect & biased.

- Incorrect parameter estimates
- Biased standard Error.
- Cause of Overdispersion

Zero Inflated Poisson Regression Model

Zero-inflated count models were first introduced by Lambert (1992) to provide another method of accounting for excessive zero counts.

- ZIP are two-part models, consisting of both binary and count model sections. (provide for the modeling of zero counts using both binary and count processes.)
- Let the response $Y_i$ denote a non-negative integer count for the $i$th observation, $i = 1, \cdots, N$.

Zero Inflated Poisson Model

The probability of an excess zero is denoted by $\pi_i$, $0 \leq \pi_i \leq 1$, the random variable $Y_i$ follows a ZIP distribution if

$$
\begin{align*}
\Pr(Y_i = y_i) &= \begin{cases} 
\pi_i + (1-\pi_i)e^{-\lambda_i}, & y_i = 0 \\
(1-\pi_i)\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, & y_i = 1, 2, \ldots, 
\end{cases} \\
E(Y_i) &= (1-\pi_i)\lambda_i; \ Var(Y_i) = \mu_i + \left(\frac{\pi_i}{1-\pi_i}\right)\mu_i^2
\end{align*}
$$

Zero Inflated Negative Binomial Model

The probability of an excess zero is denoted by $\pi_i$, $0 \leq \pi_i \leq 1$, the random variable $Y_i$ follows a ZINB distribution if

$$
\begin{align*}
\Pr(Y_i = y_i) &= \frac{\pi_i}{1-k\lambda_i^k} \left(\frac{1}{1+k\lambda_i^k}\right)^{y_i}, \ y_i = 0 \\
&= \frac{\pi_i}{1-k\lambda_i^k} \left(\frac{1}{1+k\lambda_i^k}\right)^{y_i} + \frac{1-\pi_i}{1-k\lambda_i^k} \left(\frac{1}{1+k\lambda_i^k}\right)^{y_i}, \ y_i > 0
\end{align*}
$$

where $K$ is an overdispersion parameter

$$
\begin{align*}
E(Y_i) &= (1-\pi_i)\lambda_i, \ Var(Y_i) = (1-\pi_i)\lambda_i(1+(K-1)\lambda_i),
\end{align*}
$$

Zero Inflated Negative Binomial Model

Let the response $Y_i$ denote a non-negative integer count for the $i$th observation, $i = 1, \cdots, N$, then ZINB distribution

$$
\Pr(Y_i = y_i) = \left(\frac{\pi_i}{1-k\lambda_i^k} \left(\frac{1}{1+k\lambda_i^k}\right)^{y_i} \right) + \left(\frac{1-\pi_i}{1-k\lambda_i^k} \left(\frac{1}{1+k\lambda_i^k}\right)^{y_i} \right), \ y_i > 0
$$

ZIP & ZINB Model

ZIP & ZINB: example

<table>
<thead>
<tr>
<th>y1</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20,596</td>
<td>41.19</td>
<td>41.19</td>
</tr>
<tr>
<td>1</td>
<td>12,657</td>
<td>25.31</td>
<td>66.51</td>
</tr>
<tr>
<td>2</td>
<td>7,126</td>
<td>14.25</td>
<td>80.76</td>
</tr>
<tr>
<td>3</td>
<td>4,012</td>
<td>8.02</td>
<td>88.78</td>
</tr>
<tr>
<td>4</td>
<td>2,270</td>
<td>4.54</td>
<td>93.32</td>
</tr>
<tr>
<td>5</td>
<td>1,335</td>
<td>2.67</td>
<td>95.99</td>
</tr>
<tr>
<td>6</td>
<td>781</td>
<td>1.56</td>
<td>97.55</td>
</tr>
<tr>
<td>7</td>
<td>479</td>
<td>0.96</td>
<td>98.51</td>
</tr>
<tr>
<td>8</td>
<td>278</td>
<td>0.56</td>
<td>99.07</td>
</tr>
<tr>
<td>9</td>
<td>172</td>
<td>0.35</td>
<td>99.42</td>
</tr>
<tr>
<td>10</td>
<td>106</td>
<td>0.21</td>
<td>99.63</td>
</tr>
<tr>
<td>11</td>
<td>60</td>
<td>0.12</td>
<td>99.75</td>
</tr>
<tr>
<td>12</td>
<td>33</td>
<td>0.06</td>
<td>99.81</td>
</tr>
<tr>
<td>13</td>
<td>23</td>
<td>0.04</td>
<td>99.85</td>
</tr>
<tr>
<td>14</td>
<td>20</td>
<td>0.04</td>
<td>99.90</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>0.03</td>
<td>99.93</td>
</tr>
<tr>
<td>16</td>
<td>14</td>
<td>0.02</td>
<td>99.95</td>
</tr>
<tr>
<td>17</td>
<td>11</td>
<td>0.02</td>
<td>99.97</td>
</tr>
<tr>
<td>18</td>
<td>7</td>
<td>0.01</td>
<td>99.99</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>0.01</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Total | 50,000 | 100.00

---

Example: Synthetic NB2 data :STATA (Hilbe,2011)
Zero Inflated Poisson Model

Zero inflated Poisson Example: `zip` y1 x1 x2, inflate(x1 x2)

Log likelihood = -81687.51                       Prob > chi2 =     0.0000
Inflation model = logit LR chi2(2)      =    5673.14

Iteration 4:   log likelihood = -81687.514
Iteration 4:   log likelihood = -84524.083
...
Iteration 0:   log likelihood = -93719.413

`. zip  y1 x1 x2, inflate(x1 x2)`

Zero inflated Poisson Model (ZIP): Example Interpretation

Interpretation based on Poisson Model
- Poisson Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Poisson Regression Model.

Interpretation based on Binary Logit Model
- Binary Logit Model, contains coefficients for the factor change in the odds of being in the Always Zero group compared with the Not Always Zero group.
- The coefficients interpreted in the same way as coefficients for a binary logit model

Interpretation based on Negative Binomial Model
- NB Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Negative Binomial Model.

Zero Inflated Poisson Regression Model

Zero inflated Poisson Model (ZIP): Interpretation

Zero inflated Negative Binomial Model

Zero inflated Negative Binomial Model Example: `zinh`

Zero inflated Negative Binomial Model (ZINB): Interpretation

Interpretation based on Negative Binomial Model
- NB Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Negative Binomial Model.

Interpretation based on Binary Logit Model
- Binary Logit Model, contains coefficients for the factor change in the odds of being in the Always Zero group compared with the Not Always Zero group.
- The coefficients interpreted in the same way as coefficients for a binary logit model

Interpretation based on Negative Binomial Model
- NB Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Negative Binomial Model.

Zero Inflated Negative Binomial Model (ZINB): Interpretation

Interpretation based on Negative Binomial Model
- NB Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Negative Binomial Model.

Interpretation based on Binary Logit Model
- Binary Logit Model, contains coefficients for the factor change in the odds of being in the Always Zero group compared with the Not Always Zero group.
- The coefficients interpreted in the same way as coefficients for a binary logit model

Zero Inflated Negative Binomial Model

Zero inflated Negative Binomial Model (ZINB): Interpretation

Interpretation based on Negative Binomial Model
- NB Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Negative Binomial Model.

Interpretation based on Binary Logit Model
- Binary Logit Model, contains coefficients for the factor change in the odds of being in the Always Zero group compared with the Not Always Zero group.
- The coefficients interpreted in the same way as coefficients for a binary logit model

Zero Inflated Negative Binomial Model

Zero inflated Negative Binomial Model (ZINB): Interpretation

Interpretation based on Negative Binomial Model
- NB Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Negative Binomial Model.

Interpretation based on Binary Logit Model
- Binary Logit Model, contains coefficients for the factor change in the odds of being in the Always Zero group compared with the Not Always Zero group.
- The coefficients interpreted in the same way as coefficients for a binary logit model

Zero Inflated Negative Binomial Model

Zero inflated Negative Binomial Model (ZINB): Interpretation

Interpretation based on Negative Binomial Model
- NB Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Negative Binomial Model.

Interpretation based on Binary Logit Model
- Binary Logit Model, contains coefficients for the factor change in the odds of being in the Always Zero group compared with the Not Always Zero group.
- The coefficients interpreted in the same way as coefficients for a binary logit model

Zero Inflated Negative Binomial Model

Zero inflated Negative Binomial Model (ZINB): Interpretation

Interpretation based on Negative Binomial Model
- NB Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Negative Binomial Model.

Interpretation based on Binary Logit Model
- Binary Logit Model, contains coefficients for the factor change in the odds of being in the Always Zero group compared with the Not Always Zero group.
- The coefficients interpreted in the same way as coefficients for a binary logit model

Zero Inflated Negative Binomial Model

Zero inflated Negative Binomial Model (ZINB): Interpretation

Interpretation based on Negative Binomial Model
- NB Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Negative Binomial Model.

Interpretation based on Binary Logit Model
- Binary Logit Model, contains coefficients for the factor change in the odds of being in the Always Zero group compared with the Not Always Zero group.
- The coefficients interpreted in the same way as coefficients for a binary logit model

Zero Inflated Negative Binomial Model

Zero inflated Negative Binomial Model (ZINB): Interpretation

Interpretation based on Negative Binomial Model
- NB Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Negative Binomial Model.

Interpretation based on Binary Logit Model
- Binary Logit Model, contains coefficients for the factor change in the odds of being in the Always Zero group compared with the Not Always Zero group.
- The coefficients interpreted in the same way as coefficients for a binary logit model

Zero Inflated Negative Binomial Model

Zero inflated Negative Binomial Model (ZINB): Interpretation

Interpretation based on Negative Binomial Model
- NB Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Negative Binomial Model.

Interpretation based on Binary Logit Model
- Binary Logit Model, contains coefficients for the factor change in the odds of being in the Always Zero group compared with the Not Always Zero group.
- The coefficients interpreted in the same way as coefficients for a binary logit model

Zero Inflated Negative Binomial Model

Zero inflated Negative Binomial Model (ZINB): Interpretation

Interpretation based on Negative Binomial Model
- NB Model, contains coefficients for the factor change in expected count for those in the Not Always Zero group.
- The coefficients can be interpreted in the same way as coefficient from the Negative Binomial Model.

Interpretation based on Binary Logit Model
- Binary Logit Model, contains coefficients for the factor change in the odds of being in the Always Zero group compared with the Not Always Zero group.
- The coefficients interpreted in the same way as coefficients for a binary logit model
Zero Inflated Negative Binomial Model

Zero Inflated Negative Binomial (ZINB): Example Interpretation

```
. zinb y1 x1 x2, inflate(x1 x2) vuong zip

Test of Comparative fit

Comparison models present

Combinatorial model: Factor Change in Expected Count

Count Equation: Factor Change in Expected Count for Those Not Always 0

```

Test of Comparative Fit

Comparison of Models

```

Summary statistics across models: BIC, AIC, likelihood Ratio Test, Vuong test

Graph Difference between the observed and predicted probability for the PRM, NB2, ZIP & ZINB models

(Long & Freese, 2006)
```
Comparison of Models

Comparison model: countfit (Graph & statistics across models)

- Summary statistics across models: BIC, AIC, likelihood Ratio Test, Young test
- Graph Difference between the observed and predicted probability for the PMR, NB2, ZIP & ZINB models
  
  . countfit y1 x1 x2, gen(base_) inflate(x1 x2) maxcount(10) /// pm storreg zip zinb model.

Summary of Mean Observed and Predicted Count

<table>
<thead>
<tr>
<th>Model</th>
<th>Difference</th>
<th>Value</th>
<th></th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base_NB2M</td>
<td>-0.114</td>
<td>2</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Base_NBRM</td>
<td>0.124</td>
<td>2</td>
<td>0.103</td>
<td></td>
</tr>
<tr>
<td>Base_ZINB</td>
<td>0.119</td>
<td>1</td>
<td>0.016</td>
<td></td>
</tr>
</tbody>
</table>

Tests and Fit Statistics

| Base_NB2M | BIC= -1311.572  AIC= 3.566 | Prefer Over Evidence |
| Base_NB2M | BIC= -1387.037  AIC= 3.390 | Prefer Over Evidence |
| Base_ZIP  | BIC= -1466.037  AIC= 3.249 | Prefer Over Evidence |
| Base_ZINB | BIC= -1448.970  AIC= 3.258 | Prefer Over Evidence |

Comparison model: zinb (Graph & statistics across models)

Comparison of Models

Comparison of Mean Observed and Predicted Count

... countfit y1 x1 x2, gen(Base_) inflate(x1 x2) maxcount(10)  ///

Fitting zinb model:

```stata
zinb y1 x1 x2, inflate(x1 x2) vuong zip
```

Vuong test of zinb vs. standard negative binomial: z = 0.52  Pr>z = 0.3016

Likelihood-ratio test of alpha=0: chibar2(01) = 68.15 Pr>=chibar2 = 0.0000

Comparison of Models

Comparison model: countfit (Graph & statistics across models)

Comparison of Models

Comparison model: zinb (Young test)

Comparison of Models

Comparison model: countfit (Graph & statistics across models)

Other Count Data Models

- Zero truncated Poisson & Zero truncated negative binomial
- Truncated Poisson & truncated negative binomial (zap & zanb)
- Censored Poisson & censored negative binomial
- Generalized Poisson Regression
- Generalized Negative Binomial
- etc

Reference

Reference: Negative Binomial & other Count Models

- Mullahy, J. (1986). Truncated Poisson & truncated negative binomial etc
- Overdispersion test. Available at http://home.kku.ac.th/nikom