Zero-Inflated Poisson Models

- In some settings, the incidence of zero counts will be much greater than expected for the Poisson distribution.
- Poisson regression models will exhibit overdispersion when they are fit to data with an excess number of zeros.
- Zero-inflated Poisson (ZIP) models might be a better fit to the data.
Count Data with Many Zeros

ZIP Models

- The population that can be modeled with the zero-inflated Poisson distribution is considered to consist of two types of responses.
- The first type gives Poisson distributed counts, which can produce the zero outcome or some other positive outcome.
- The second type always gives a zero count.
- Therefore, the relevant distribution is a mixture of a Poisson distribution and a distribution that is constant at zero.
## A Biological Example

### Table: Photoperiod and Concentration

<table>
<thead>
<tr>
<th>Photoperiod (hour)</th>
<th>Concentration (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.2</td>
</tr>
<tr>
<td>8</td>
<td>Number of roots</td>
</tr>
<tr>
<td>16</td>
<td>Number of roots</td>
</tr>
</tbody>
</table>

### Distribution

![Histogram of Number of Roots]

- X-axis: Number of roots
- Y-axis: Percent
- Data shows a bell-shaped distribution with the peak at around 4 roots.
MCMC Information

• ZIP models can be fit by specifying two models in PROC MCMC. The first one is a model for the Poisson mean.

\[ \mu = e^{\beta_0 + \beta_1 \text{photo} + \beta_2 \text{bap} + \beta_3 \text{photo} \_\text{bap}} \]

• The second model is a logistic model for the probability of the excess number of zeros.

\[ p_0 = \text{logistic}(\gamma_0 + \gamma_1 \text{photo}) \]

MCMC Information

• The log likelihood function is defined as:

\[ llike = \log(p_0 \text{ (roots = 0)}) + (1 - p_0) \text{ pdf("Poisson", roots, mu))} \]

• You can use the DGENERAL function to specify the mixture likelihood function.
Fitting a Zero-Inflated Poisson Model in PROC MCMC

This demonstration illustrates the concepts discussed previously.

Questions?