Bayesian Analysis: A Deep Dive into ZIP Models

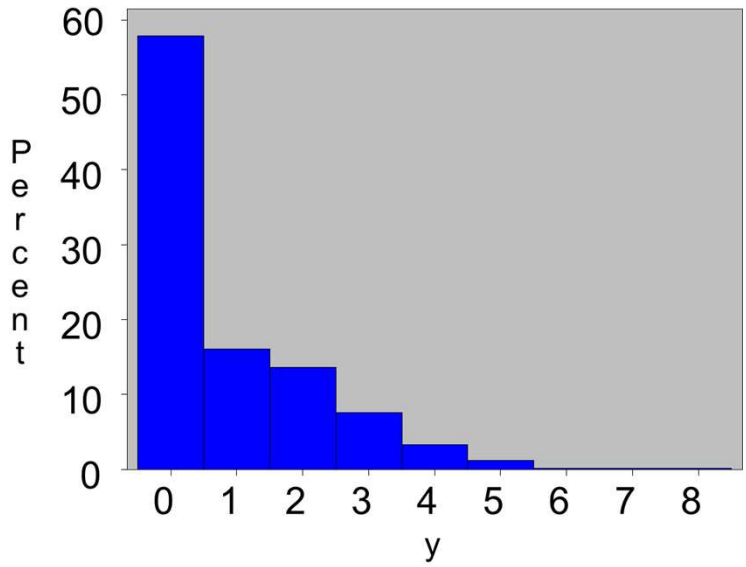
Danny Modlin

Sr. Analytical Training Consultant – Cary, NC

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



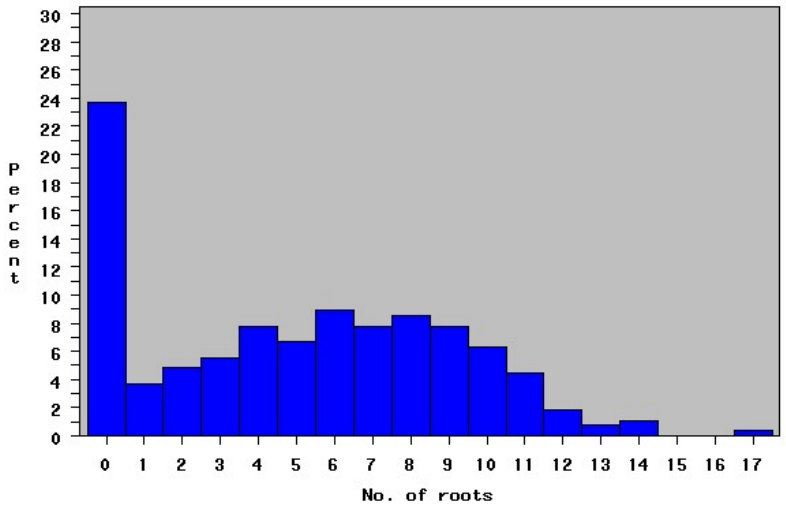
ZIP Models

- 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

photoperiod (hour)	concentration (μM)			
	2.2	4.4	8.8	17.6
8	Number of roots	Number of roots	Number of roots	Number of roots
16	Number of roots	Number of roots	Number of roots	Number of roots

Distribution



MCMC Information

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

$$mu = e^{\beta_0 + \beta_1 * photo + \beta_2 * bap + \beta_3 * photo\_bap}$$

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

$$p_0 = \text{logistic}(\gamma_0 + \gamma_1 * photo)$$

MCMC Information

- The log likelihood function is defined as:

$$llike = \log(p_0 * (roots = 0) + (1 - p_0) * pdf("Poisson", roots, mu))$$

- 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?