To Offset or Not: Using Offsets in Count Models

March 2018

When count data is collected, different observations may be collected with varying levels of “effort”. For example, suppose that the count variable of interest is the number of ants that arrive at a food source. Ideally, each food source would be observed for the same amount of time; i.e., each observation would be collected with equal effort. However, if the food sources were observed for varying amounts of time, the analysis should account for the level of effort expended in the collection of each observation. This can be accomplished using an offset variable in a count model.

Poisson and negative binomial regression models are frequently used to model count data. The Poisson model can be written

$$\log(\mu) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$ 

The exponentiated beta coefficient of a predictor variable from a Poisson model tells us how much the expected count changes multiplicatively for a one unit increase in the predictor variable. For more information on this type of model, please refer to our StatNews #43. Notice that this form of the model assumes that the effort is the same for each observation. If this is not the case, an offset variable should be included in the model.

An offset variable is a variable whose beta coefficient is constrained to be 1. When an offset variable is included in a Poisson or negative binomial regression model, the rate is being modeled instead of the count. In order to successfully model the rate, the log of the effort variable is included in the model as an offset. For the ant arrival example, the offset variable would be the log of the amount of time spent observing each food source. The formulation of the Poisson regression model with a logged effort variable ($A$) as an offset can be expressed as

$$\log(\mu) = 1 \cdot \log(A) + \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$
which is equivalent to,

$$\log \left( \frac{\mu}{A} \right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$ 

When you fit a model with an offset, the exponentiated beta coefficients of a predictor variable tells us how much the expected rate changes multiplicatively for a one unit increase in the predictor variable; the rate is in terms of counts per one unit of the effort.

When using an offset, the assumption is made that doubling the effort will lead to a doubling of the count. If this assumption is not appropriate, controlling for the effort as a covariate instead of an offset may be more appropriate. A likelihood ratio test could be used to determine if the fit of these two model is significantly different.

By not controlling for exposure or effort, the model may have over-dispersion due to a lack of fit. For more information on over-dispersion in Poisson regression, please refer to our StatNews #86.

For studies that involve running a series of trials where each trial has a binary outcome (success or failure), it may be tempting to use a count model where the response is the count of the number of successes and the number of trials is considered the effort. However, since response and the count are in the same units, the response has an upper bound (the number of trials). This type of data would be better analyzed using a binomial count model.

If you need assistance fitting count models with offsets, or if you have any other statistical consulting questions, please feel free to contact the statistical consultants at CSCU.

Author: Stephen Parry