Text Version

Expand Print

Poisson Regression for Rates

When outcomes occur over time, space, or some other index of size, it might be more useful to model the rate of occurrence rather than the counts.

Rates are simply counts divided by the measure of exposure. If the size measures vary across observations, modeling the rates provides the necessary standardization to ensure that you can compare outcomes. For example, consider the study of disease incidence by county. Because more at-risk subjects result in more occurrences of the disease, you need to adjust for the number of subjects at risk in each county. Simply modeling the number of occurrences of disease by county would be misleading if the number at risk in each county varies. Modelling rates provides the standardization needed.

Modeling the rate of an event is common in many fields of study, including in the medical sciences, biological sciences, social sciences, engineering and business. For example, an epidemiologist might want to examine how the county's cancer rate is related to certain demographic variables and health policy programs in a county. A business analyst might want to examine how the rate of insurance claims is related to demographic variables and the number of years since the last claim on the policy or how the rate of loan defaults is related to a region of the country. A government agency might want to know how crime rates are related to a city's unemployment rate. A market researcher might want to study how response rates to marketing campaigns are related to known characteristics of the recipients.

To model Poisson outcomes as rates rather than counts, you assume that the numerator variable (that is, the counts) follows a Poisson distribution and that the denominator standardizes the counts. Both the counts (numerator variable) and size measures (denominator variable) are required to model the rates. If the numerator and denominator are not known (that is, just the rates are known) another possibility for modeling rates is the gamma distribution.

To model rates using the Poisson regression model, you use the OFFSET= option in the MODEL statement in PROC GENMOD. For Poisson regression models for rates, the log of the incidence (where T is a measure of exposure) is modeled as a linear function of the explanatory variables. Rearranging the terms in the model, you can see that the log of the mean can be modeled as a linear function of the explanatory variables and the log of the measure of exposure. The log of the measure of exposure is called the offset variable. If the measure of exposure is the same for every subject, then the offset variable can be absorbed into the intercept.

If you exponentiate both sides of the model expression, you obtain the expected number of events. Notice that the expected number of events is proportional to the index of exposure times the marginal effects of the explanatory variables. The scale of the index of exposure does not affect the parameter estimates for the explanatory variables. For example, expressing disease rate by 1,000 people or by 100,000 people does not matter because the differences in exposure scales are reflected in the different values of the intercept.

Close