

Poisson Regression

Poisson regression is a type of generalized linear model often used to analyze count data. It is most appropriate when the response variable represents a count of some relatively rare event, such as defects in a unit of manufactured product, errors or “bugs” in software, or a count of particulate matter or other pollutants in the environment. Additional examples of possible Poisson Regression response variables include: number of ear infections in infants number of equipment failures colony counts for bacteria or viruses counts of a rare disease in a population number of fatal crashes at an intersection.

Poisson regression assumes that the response variable follows a Poisson distribution that is conditional on the values of the predictor variables. The Poisson is another example of a probability distribution where the mean and variance are related. In fact, both the mean and variance of the Poisson distribution are equal. This distribution is the benchmark distribution for count data in much the same way that the normal distribution is the benchmark for continuous data. Although ordinary least squares regression can be used to analyze count data, Poisson regression has the advantage of being precisely customized to the discrete, often skewed distribution of count data.

In addition to being skewed, the sample distribution should have a fairly small mean if Poisson regression is the method of choice. The mean should certainly be below 10, preferably below 5, and ideally close to 1. For highly skewed data with large mean values, the gamma distribution or the lognormal distribution might be more appropriate.

Count data can include a zero value. In some situations, the number of zeros observed may cause a problem with the equality of the mean and variance. When count data have an incidence of zeros greater than expected for the underlying probability distribution, you must use a zero-inflated Poisson model. You can learn more about the zero-inflated Poisson model by taking additional SAS training. A list of related SAS training is available in the Help and Resources section of this course.