Chapter 870

Poisson Regression

Introduction

Poisson regression is used when the dependent variable is a count. Following the results of Signorini (1991), this procedure calculates power and sample size for testing the hypothesis that $\beta_1 = 0$ versus the alternative that $\beta_1 = B$. Note that $e^{\beta 1}$ is the change in the rate for a one-unit change in X_1 when the rest of the covariates are held constant. The procedure assumes that this hypothesis will be tested using the score statistic

$$z = \frac{\hat{\beta}_1}{\sqrt{\mathrm{Var}(\hat{\beta}_1)}}$$

The Poisson Distribution

The Poisson distribution models the probability of *y* events (i.e., failure, death, or existence) with the formula

$$\Pr(Y = y|\mu) = \frac{e^{-\mu}\mu^y}{y!} \quad (y = 0,1,2,\dots)$$

Notice that the Poisson distribution is specified with a single parameter μ. This is the mean incidence rate of a rare event per unit of *exposure*. Exposure may be time, space, distance, area, volume, or population size. Because exposure is often a period of time, we use the symbol *t* to represent the exposure. When no exposure value is given, it is assumed to be one.

The parameter μ may be interpreted as the risk of a new occurrence of the event during a specified exposure period, *t*. The probability of *y* events is then given by

$$\Pr(Y = y|\mu, t) = \frac{e^{-\mu t}(\mu t)^y}{y!} \quad (y = 0,1,2,\dots)$$

The Poisson distribution has the property that its mean and variance are equal.

The Poisson Regression Model

In Poisson regression, we suppose that the Poisson incidence rate μ is determined by a set of *k* regressor variables (the *X's*). The expression relating these quantities is

$$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$$

The regression coefficients $\beta_0, \beta_1, \beta_2, \cdots, \beta_k$ are unknown parameters that are estimated from a set of data. Their estimates are labeled $b_0, b_1, \cdots, b_k$.

Using this notation, the fundamental Poisson regression model for an observation i is written as

$$\Pr(Y_i = y_i | \mu_i, t_i) = \frac{e^{-\mu_i t_i}(\mu_i t_i)^{y_i}}{y_i!}$$

where

$$\mu_i = \lambda(\mathrm{x}_i^{'}\beta)$$

$$\lambda(\mathbf{x}_i^{'}\boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki})$$

That is, for a given set of values of the regressor variables, the outcome follows the Poisson distribution.

Power Calculations

Suppose you want to test the null hypothesis that $\beta_1 = 0$ versus the alternative that $\beta_1 = B1$. Signorini (1991) gives the formula relating sample size, α, β, and *B*1 when X_1 is the only covariate of interest as

$$N = \phi \frac{\left(z_{1-\alpha/2}\sqrt{V(b_1|\beta_1 = 0)} + z_{1-\beta}\sqrt{V(b_1|\beta_1 = B1)}\right)^2}{\mu_T e^{\beta_0} B1^2}$$

where *N* is the sample size, ϕ is a measure of over-dispersion, μ_T is the mean exposure time, and *z* is the standard normal deviate. Following the extension used in Hsieh, Block, and Larsen (1998) and Hsieh and Lavori (2000), when there are other covariates, **PASS** uses the approximation

$$N = \phi \frac{\left(z_{1-\alpha/2}\sqrt{V(b_1|\beta_1 = 0)} + z_{1-\beta}\sqrt{V(b_1|\beta_1 = B1)}\right)^2}{\mu_T e^{\beta_0} B1^2 (1 - R^2)}$$

where R^2 is the square of the multiple correlation coefficient when the covariate of interest is regressed on the other covariates. The variance in the null case is given by

$$V(b_1|\beta_1 = 0) = \frac{1}{\mathrm{Var}(X1)}$$

The variance for the non-null case depends on the underlying distribution of *X*. Common choices are given next.

Normal

$$V(b_1|\beta_1 = B1) = \frac{1}{\sigma_{X1}^2} e^{-\left(B1\mu_{X1} + \frac{B1^2\sigma_{X1}^2}{2}\right)}$$

$$V(X1) = \sigma_{X1}^2$$

Exponential

$$V(b_1|\beta_1 = B1) = \frac{(\lambda_{X1} - B)^3}{\lambda_{X1}}$$

$$V(X1) = \lambda_{X1}^{-2}$$

Uniform, on the Interval [C,D]

$$V(b_1|\beta_1 = B1) = \frac{m}{m(m_{11}) - m_1^2}$$

$$V(X1) = \frac{(D - C)^2}{12}$$

where

$$m = \frac{e^{B1D} - e^{B1C}}{(D - C)B1}$$

$$m_1 = \frac{e^{B1D}(B1D - 1) - e^{B1C}(B1C - 1)}{(D - C)B1^2}$$

$$m_{11} = \frac{e^{B1D}(2 - 2B1D + B1^2D^2) - e^{B1C}(2 - 2B1C + B1^2C^2)}{(D - C)B1^3}$$

Binomial, Parameter π_{X1}

$$V(b_1|\beta_1 = B1) = \frac{1}{1 - \pi_{X1}} + \frac{1}{\pi_{X1}e^{B1}}$$

$$V(X1) = \pi_{X1}(1 - \pi_{X1})$$

Example 1 – Power for Several Sample Sizes

Poisson regression will be used to analyze the power for a study of the relationship between the number of flaws on a manufactured article and the experience (measured in years) of the operator. The researchers want to evaluate the sample size needs for detecting ratios in response rates of 1.3 and 1.5. The experience of an operator is assumed to be normally distributed with mean 3.2 and standard deviation 2.1. No other covariates will be included in the analysis. The researchers will test their hypothesis using a 5% significance level with a two-sided Wald test. They decide to calculate the power at sample sizes between 5 and 50.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**Power**
Alternative Hypothesis	**Two-Sided**
Alpha	**0.05**
N (Sample Size)	**5 to 50 by 5**
Exp(B0)	**1.0**
Exp(B1) / Exp(B0)	**1.3 1.5**
MuT (Mean Exposure Time)	**1.0**
Phi	**1.0**
R-Squared of X1 with Other X's	**0.0**
Distribution of X1	**Normal(M,S)**
Normal(M)	**3.2**
Normal(S)	**2.1**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Reports

Numeric Results

Solve For:	Power
Alternative Hypothesis:	Two-Sided
Distribution of X1:	Normal(M,S)
Mean (M):	3.2
Sigma (S):	2.1
Phi (Over-Dispersion Parameter):	1

Power	Sample Size N	Response Rate Ratio Exp(B1)/Exp(B0)	Baseline Rate Exp(B0)	Mean Exposure Time MuT	R-Squared of X1 with Other X's R^2	Alpha
0.11604	5	1.3	1	1	0	0.05
0.36043	10	1.3	1	1	0	0.05
0.61237	15	1.3	1	1	0	0.05
0.79600	20	1.3	1	1	0	0.05
0.90403	25	1.3	1	1	0	0.05
0.95876	30	1.3	1	1	0	0.05
0.98355	35	1.3	1	1	0	0.05
0.99384	40	1.3	1	1	0	0.05
0.99781	45	1.3	1	1	0	0.05
0.99926	50	1.3	1	1	0	0.05
0.44890	5	1.5	1	1	0	0.05
0.95354	10	1.5	1	1	0	0.05
0.99892	15	1.5	1	1	0	0.05
0.99999	20	1.5	1	1	0	0.05
1.00000	25	1.5	1	1	0	0.05
1.00000	30	1.5	1	1	0	0.05
1.00000	35	1.5	1	1	0	0.05
1.00000	40	1.5	1	1	0	0.05
1.00000	45	1.5	1	1	0	0.05
1.00000	50	1.5	1	1	0	0.05

Phi	The over-dispersion parameter used when the Poisson model does not fit.
Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N	The size of the sample drawn from the population.
Exp(B1)/Exp(B0)	The response rate ratio due to a one-unit change in X1.
Exp(B0)	The response rate when all covariates have a value of zero.
MuT	The mean exposure time of the subjects of the study.
R^2	The R-squared achieved when X1 is regressed on the other covariates.
Alpha	The probability of rejecting Exp(B1)/Exp(B0) is one.

Summary Statements

A Poisson regression (count response Y versus X's) design will be used to test whether the response rate (Exp(B1)) of the variable of interest (X1) is different from the baseline response rate (Exp(B0)) of 1. The comparison will be made using a two-sided Poisson regression slope test of B1, with a Type I error rate (α) of 0.05. The mean exposure time is 1. The over-dispersion parameter is assumed to be 1 (no over-dispersion is assumed). The variable of interest (X1) is assumed to follow a normal distribution with a mean of 3.2 and a standard deviation of 2.1. To detect a response rate ratio (Exp(B1) / Exp(B0)) of 1.3 (for a unit increase in X1) with a sample size of 5, the power is 0.11604.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	5	7	2
20%	10	13	3
20%	15	19	4
20%	20	25	5
20%	25	32	7
20%	30	38	8
20%	35	44	9
20%	40	50	10
20%	45	57	12
20%	50	63	13

Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N The evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N' The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula N' = N / (1 - DR), with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D The expected number of dropouts. D = N' - N.

Dropout Summary Statements

Anticipating a 20% dropout rate, 7 subjects should be enrolled to obtain a final sample size of 5 subjects.

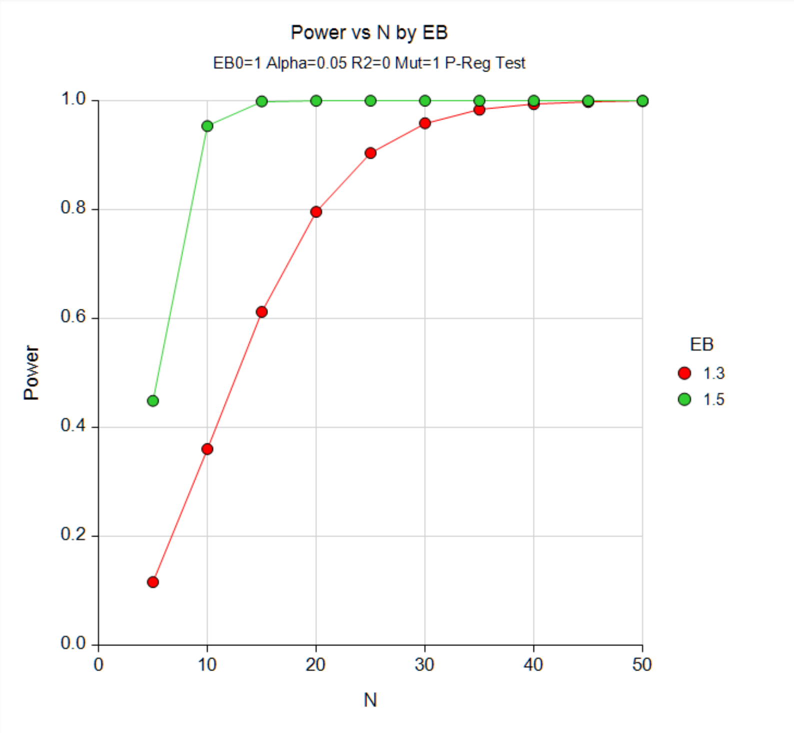
References

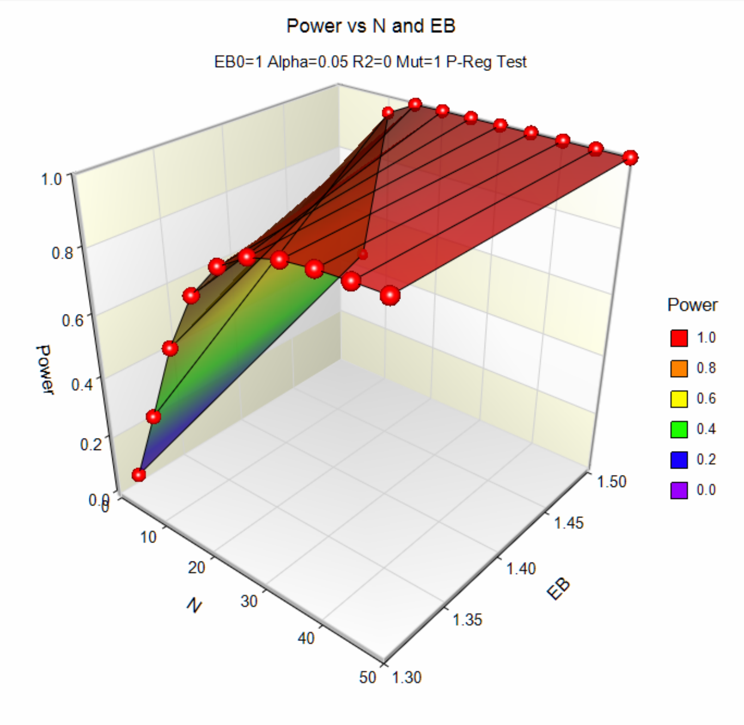
Signorini, David. 1991. 'Sample size for Poisson regression', Biometrika, Volume 78, 2, pages 446-450.

This report shows the power for each of the scenarios. Note that if you were interested in $B1$ instead of Exp($B1$), you would simply take the natural logarithm of the value of Exp($B1$).

Plots Section

Plots





These plots show the relationship between power and sample size.

Example 2 – Validation using Signorini (1991)

Signorini (1991), page 449, presents an example which we will use to validate this program. In the example, Exp(*B*1)/Exp(B0)= 1.3, Exp(*B*0)= 0.85, *R*2 = 0.0, *Mu T* = 1.0, and *Phi* = 1.0. The independent variable is assumed to be binomial with proportion 0.5. A one-sided test with alpha = 0.05 will be used. Sample sizes for power = 0.80, 0.90, and 0.95 are calculated to be 406, 555, and 697.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Sample Size**
Alternative Hypothesis .... **One-Sided**
Power.. **0.80 0.90 0.95**
Alpha.. **0.05**
Exp(B0)... **0.85**
Exp(B1) / Exp(B0)... **1.3**
MuT (Mean Exposure Time) **1.0**
Phi... **1.0**
R-Squared of X1 with Other X's..................... **0.0**
Distribution of X1 .... **Binomial(P)**
Binomial(P) **0.5**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Solve For: Sample Size
Alternative Hypothesis: One-Sided
Distribution of X1: Binomial(P)
  Proportion (P): 0.5
Phi (Over-Dispersion Parameter): 1

Power	Sample Size N	Response Rate Ratio Exp(B1)/Exp(B0)	Baseline Rate Exp(B0)	Mean Exposure Time MuT	R-Squared of X1 with Other X's R^2	Alpha
0.80	406	1.3	0.85	1	0	0.05
0.90	556	1.3	0.85	1	0	0.05
0.95	697	1.3	0.85	1	0	0.05

Note that **PASS** calculated 556 rather than the 555 calculated by Signorini (1991). The discrepancy is due to rounding.