

## 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

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

## 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  $\mu$ . 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  $\mu$  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  $\mu$  is determined by a set of  $k$  regressor variables (the  $X$ 's). The expression relating these quantities is

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

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

## Poisson Regression

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

$$\mu_i = \lambda(\mathbf{x}_i' \boldsymbol{\beta})$$

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

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,  $\alpha$ ,  $\beta$ , and  $B1$  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,  $\phi$  is a measure of over-dispersion,  $\mu_T$  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

$$V(b_1 | \beta_1 = 0) = \frac{1}{\text{Var}(X_1)}$$

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

### Normal

$$V(b_1 | \beta_1 = B1) = \frac{1}{\sigma_{X_1}^2} e^{-\left( B1 \mu_{X_1} + \frac{B1^2 \sigma_{X_1}^2}{2} \right)}$$

$$V(X_1) = \sigma_{X_1}^2$$

### Exponential

$$V(b_1 | \beta_1 = B1) = \frac{(\lambda_{X_1} - B)^3}{\lambda_{X_1}}$$

$$V(X_1) = \lambda_{X_1}^{-2}$$

## Poisson Regression

## 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

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

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

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

Binomial, Parameter  $\pi_{X1}$ 

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

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

## Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

### Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

#### Solve For

##### Solve For

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Power* for a power analysis or *Sample Size* for sample size determination.

Select *Sample Size* when you want to calculate the sample size needed to achieve a given power and alpha level. Select *Power* when you want to calculate the power of an experiment.

## Poisson Regression

---

### Test

#### Alternative Hypothesis

Specify whether the alternative hypothesis is one-sided or two-sided. When a two-sided alternative hypothesis is selected, the specified value of alpha is divided by 2. When a one-sided hypothesis is selected, the value of alpha is left as specified.

In general, the accepted procedure is to use the Two-Sided option unless a one-sided test can be justified.

---

### Power and Alpha

#### Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also very popular.

A single value (e.g., 0.90) may be entered here or a range of values (e.g., 0.8 to 0.95 by 0.05) may be entered.

#### Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

A single value (e.g., 0.05) may be entered here or a range of values (e.g., 0.05 to 0.2 by 0.05) may be entered.

---

### Sample Size

#### N (Sample Size)

This option specifies the total number of observations in the sample. You may enter a single value (e.g., 25) or a list of values (e.g., 10 to 100 by 10).

---

### Effect Size – Response Rates

#### Exp(B0) [Baseline Response Rate]

Exp(B0) is the baseline response rate. This is the response rate that occurs when all covariates are equal to zero. Depending on the dataset being analyzed, the response rate might be the hazard rate, death rate, survival rate, or accident rate.

#### Exp(B1)/Exp(B0) (Rate Ratio)

$B1$  is the value of the regression coefficient under the alternative hypothesis. In Poisson regression, it is more natural to specify a value for  $\exp(B1)/\exp(B0)$  than for  $B1$  because  $\exp(B1)/\exp(B0)$  represents the ratio of the response rate when  $X1$  is increased one unit and all other covariates are constant to its baseline rate. Depending on the dataset being analyzed, the response rate might be the hazard rate, death rate, survival rate, or accident rate.

For example, suppose the baseline response rate (the value of  $\exp(\beta_0)$ ) is 0.70 and you want the study to be large enough to detect a 30% increase in the response rate. The value entered here would be 1.3.

## Poisson Regression

---

**Effect Size – Exposure Time and Over-Dispersion****MuT (Mean Exposure Time)**

This is the mean exposure time,  $\mu_T$ , over which the response rate is calculated. If the response rates are for one year, enter 1 here. If they are for 30 days, enter 30 here.

**Phi**

Phi is the over-dispersion parameter for Poisson regression. When there is no over-dispersion, set this value to one.

---

**Effect Size – Covariates (X1 is the Variable of Interest)****R-Squared of X1 with Other X's**

This is the *R*-Squared that is obtained when  $X_1$  is regressed on the other  $X$ 's (covariates) in the model. Use this to account for the influence on power and sample size of adding other covariates. Note that the number of additional variables does not matter in this formulation. Only their overall relationship with  $X_1$  through this *R*-Squared value is used.

This value is restricted to being greater than or equal to zero and less than one. Use zero when there are no other covariates.

**Distribution of X1**

This option specifies the distribution of the covariate being analyzed. You can choose from the normal, exponential, uniform, and binomial distributions. Once you have selected the anticipated distribution of the covariate, you must also specify values for the corresponding parameter(s) of the distribution.

**Normal(M)**

Specify the mean of the normal covariate distribution. This option is used when Distribution of X1 is set to Normal(M,S).

**Normal(S)**

Specify the standard deviation of the normal covariate distribution. This option is used when Distribution of X1 is set to Normal(M,S).

**Uniform(C)**

Specify the minimum of the uniform covariate distribution. This option is used when Distribution of X1 is set to Uniform(C,D). The uniform distribution of the covariate is assumed to range from  $C$  to  $D$ .

**Uniform(D)**

Specify the maximum of the uniform covariate distribution. This option is used when Distribution of X1 is set to Uniform(C,D). The uniform distribution of the covariate is assumed to range from  $C$  to  $D$ .

**Exponential(L)**

Specify the value of *lambda* associated with the exponential covariate distribution. This option is used when Distribution of X1 is set to Exponential(L). Note that the pdf of the exponential distribution is assumed to be  $(\lambda) \exp(-\lambda X)$ .

**Binomial(P)**

Specify the value of the proportion for the binomial covariate distribution. This option is used when the Distribution of X1 is set to Binomial(P). For example, if the covariate being studied is an indicator variable of treatment and control and you anticipated having an equal number of treatments and controls, this value should be set to 0.50.

## Poisson Regression

## 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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Poisson Regression** procedure window. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

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

### Annotated Output

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

### Numeric Results

Numeric Results when X1 is Normal with Mean = 3.2 and Sigma = 2.1  
And Phi (Over-Dispersion Parameter) = 1.0000

Power	Sample Size (N)	Response Rate Ratio	Baseline Rate Exp(B0)	Mean Exposure Time (MuT)	R-Squared X1 vs Other X's (R2)	Two-Sided Alpha	Beta
0.11604	5	1.3000	1.0000	1.0000	0.0000	0.05000	0.88396
0.36043	10	1.3000	1.0000	1.0000	0.0000	0.05000	0.63957
0.61237	15	1.3000	1.0000	1.0000	0.0000	0.05000	0.38763
0.79600	20	1.3000	1.0000	1.0000	0.0000	0.05000	0.20400
0.90403	25	1.3000	1.0000	1.0000	0.0000	0.05000	0.09597
0.95876	30	1.3000	1.0000	1.0000	0.0000	0.05000	0.04124
0.98355	35	1.3000	1.0000	1.0000	0.0000	0.05000	0.01645
0.99384	40	1.3000	1.0000	1.0000	0.0000	0.05000	0.00616
0.99781	45	1.3000	1.0000	1.0000	0.0000	0.05000	0.00219
0.99926	50	1.3000	1.0000	1.0000	0.0000	0.05000	0.00074
0.44890	5	1.5000	1.0000	1.0000	0.0000	0.05000	0.55110
0.95354	10	1.5000	1.0000	1.0000	0.0000	0.05000	0.04646

## Poisson Regression

(Numeric Results – Continued)

0.99892	15	1.5000	1.0000	1.0000	0.0000	0.05000	0.00108
0.99999	20	1.5000	1.0000	1.0000	0.0000	0.05000	0.00001
1.00000	25	1.5000	1.0000	1.0000	0.0000	0.05000	0.00000
1.00000	30	1.5000	1.0000	1.0000	0.0000	0.05000	0.00000
1.00000	35	1.5000	1.0000	1.0000	0.0000	0.05000	0.00000
1.00000	40	1.5000	1.0000	1.0000	0.0000	0.05000	0.00000
1.00000	45	1.5000	1.0000	1.0000	0.0000	0.05000	0.00000
1.00000	50	1.5000	1.0000	1.0000	0.0000	0.05000	0.00000

### Report Definitions

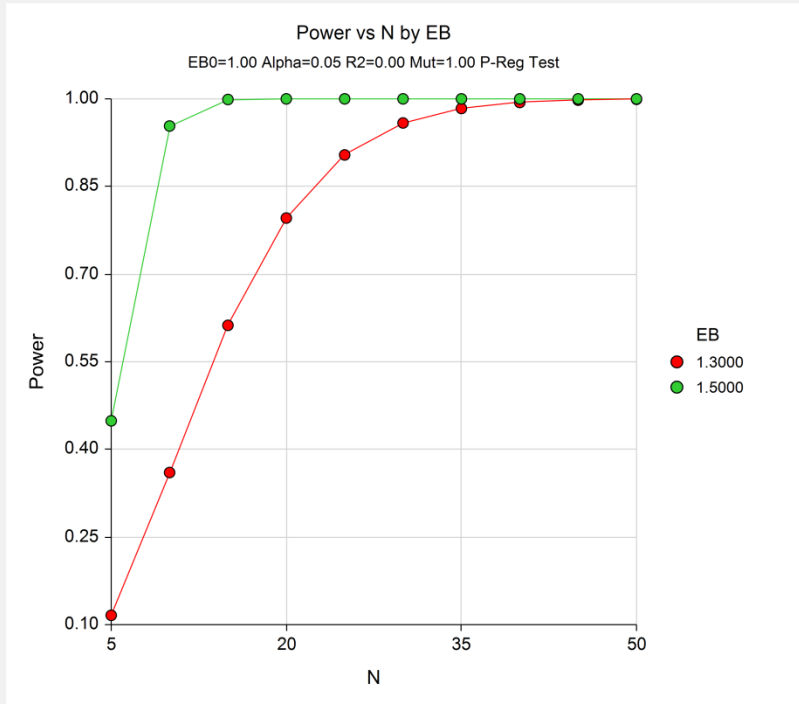
Power is the probability of rejecting a false null hypothesis. It should be close to one.  
 N is the size of the sample drawn from the population.  
 Exp(B1)/Exp(B0) is the response rate ratio due to a one-unit change in X1.  
 Exp(B0) is the response rate when all covariates have a value of zero.  
 Phi is the over-dispersion parameter used when the Poisson model does not fit.  
 R2 is the R-squared achieved when X1 is regressed on the other covariates.  
 Alpha is the probability of rejecting Exp(B1)/Exp(B0) is one.  
 Beta is the probability of accepting a false null hypothesis.

### Summary Statements

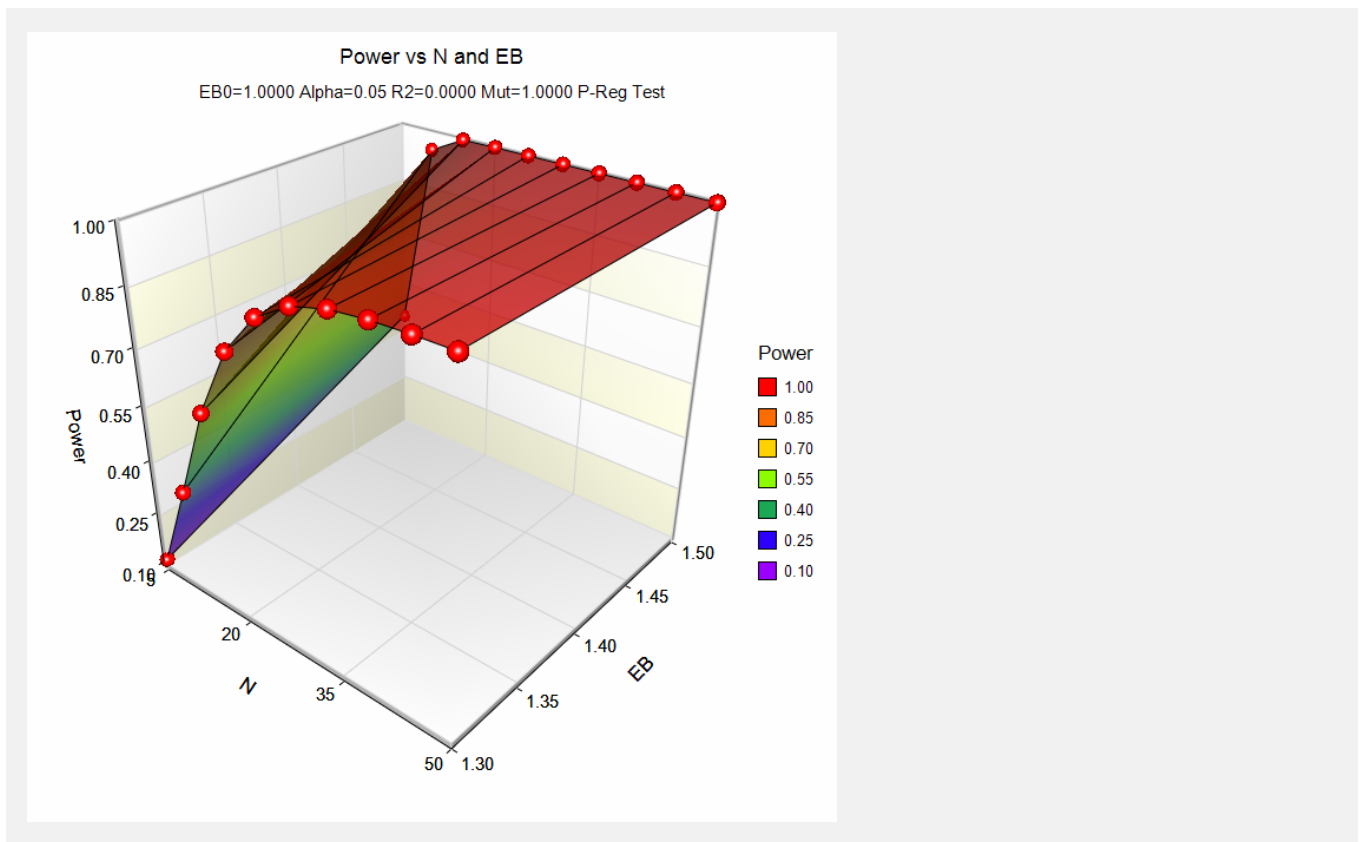
A Poisson regression of a dependent variable of counts on normally distributed independent variable with mean = 3.2 and standard deviation = 2.1 using a sample of 5 observations achieves 12% power at a 0.05000 significance level to detect a response rate ratio of at least 1.3000 due to a one-unit change in the IV. The baseline rate is 1.0000 and the mean exposure time is 1.0000.

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

## Plots Section



Poisson Regression



These plots show the relationship between power and sample size.



## Poisson Regression

## Example 2 – Validation using Signorini

Signorini (1991), page 449, presents an example which we will use to validate this program. In the example,  $\text{Exp}(B1)/\text{Exp}(B0) = 1.3$ ,  $\text{Exp}(B0) = 0.85$ ,  $R^2 = 0.0$ ,  $\text{MuT} = 1.0$ , and  $\text{Phi} = 1.0$ . 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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Poisson Regression** procedure window. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

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

### Output

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

### Numeric Results

**Numeric Results when X1 is Binomial with Proportion = 0.5  
And Phi (Over-Dispersion Parameter) = 1.0000**

Power	Sample Size (N)	Response Rate Ratio	Baseline Rate Exp(B0)	Mean Exposure Time (MuT)	R-Squared X1 vs Other X's (R2)	Two-Sided Alpha	Beta
0.80000	406	1.3000	0.8500	1.0000	0.0000	0.05000	0.20000
0.90000	556	1.3000	0.8500	1.0000	0.0000	0.05000	0.10000
0.95000	697	1.3000	0.8500	1.0000	0.0000	0.05000	0.05000

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