What are some examples of Poisson regression in SPSS for data analysis?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What are some examples of Poisson regression in SPSS for data analysis?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=158018

Poisson regression is a statistical analysis method used in SPSS software to model count data, where the dependent variable represents the number of occurrences of a particular event. Some common examples of Poisson regression in SPSS for data analysis include:

1. Analysis of traffic accidents: In this scenario, the number of accidents occurring in a specific location or time period can be modeled using Poisson regression to determine the factors that contribute to an increase or decrease in the number of accidents.

2. Healthcare research: Poisson regression can be used to analyze the number of hospital visits or medical procedures performed by patients, and identify any significant predictors such as age, gender, or medical conditions.

3. Marketing analysis: Poisson regression can be applied to analyze the number of sales or customer purchases in a specific time period, and identify the factors that influence these patterns.

4. Social science research: Researchers can use Poisson regression to analyze the number of criminal offenses in a particular area, and determine the impact of various socio-economic factors on crime rates.

Overall, Poisson regression in SPSS is a powerful tool for analyzing count data and identifying significant predictors, making it a valuable technique in various fields of research and data analysis.

Poisson Regression | SPSS Data Analysis Examples

Poisson regression is used to model count variables.

Please note: The purpose of this page is to show how to use various data analysis commands. It does not cover all aspects of the research process which researchers are expected to do. In particular, it does not cover data cleaning and checking, verification of assumptions,

model diagnostics or
potential follow-up analyses.

This page is done using SPSS 19.

Examples of Poisson regression

Example 1. The number of persons killed by mule or horse kicks in the Prussian army per year. Ladislaus Bortkiewicz collected data from 20 volumes of Preussischen Statistik. These data were collected on 10 corps of the Prussian army in the late 1800s over the course of 20 years.

Example 2. The number of people in line in front of you at the grocery store. Predictors may include the number of items currently offered at a special discounted price and whether a special event (e.g., a holiday, a big sporting event) is three or fewer days away.

Example 3. The number of awards earned by students at one high school.

Predictors of the number of awards earned include the type of program in which the student was enrolled (e.g., vocational, general or academic) and the score on their final exam in math.

Description of the Data

For the purpose of illustration, we have simulated a data set for Example 3 above: https://stats.idre.ucla.edu/wp-content/uploads/2016/02/poisson_sim.sav. In this example, num_awards is the outcome variable and indicates the number of awards earned by students at a high school in a year, math is a continuous predictor variable and represents students' scores on their math final exam, and prog is a categorical predictor variable with three levels indicating the type of program in which the students were enrolled.

Let's start with loading the data and looking at some descriptive statistics.

GET
FILE='D:datapoisson_sim.sav'.

DESCRIPTIVES
VARIABLES=math num_awards
/STATISTICS=MEAN STDDEV VAR MIN MAX .

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
math score	200	33	75	52.65	9.368	87.768
num_awards	200	0	6	.63	1.053	1.109
Valid N (listwise)	200					

Each variable has 200 valid observations and their distributions seem quite reasonable. The unconditional mean and variance of our outcome variable are not extremely different. Our model assumes that these values, conditioned on the predictor variables, will be equal (or at least roughly so).

Let's continue with our description of the variables in this dataset. The table below shows the average numbers of awards by program type and seems to

suggest that program type is a good candidate for predicting the number of awards, our outcome variable, because the mean value of the outcome appears to vary by prog. Additionally, the means and variances within each level of prog-the conditional means and variances-are similar.

MEANS tables = num_awards by prog.

Case Processing Summary

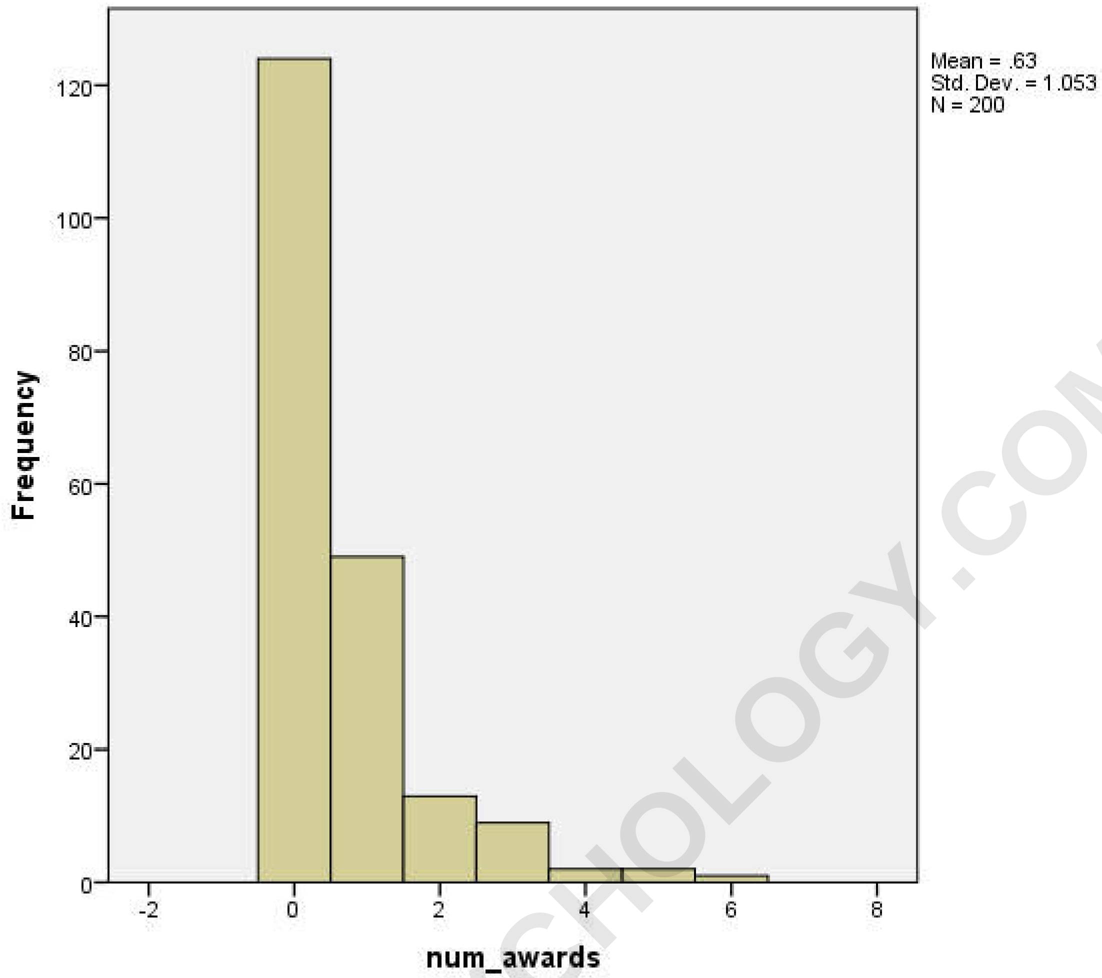
	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
num_awards * type of program	200	100.0%	0	.0%	200	100.0%

Report

num_awards

type of program	Mean	N	Std. Deviation
general	.20	45	.405
academic	1.00	105	1.279
vocation	.24	50	.517
Total	.63	200	1.053

GRAPH /HISTOGRAM=num_awards.



Analysis methods you might consider

Below is a list of some analysis methods you may have encountered. Some of the methods listed are quite reasonable, while others have either fallen out of favor or have limitations.

Poisson regression

Below we use the genlin command to estimate a Poisson regression

model. We have one continuous predictor and one categorical predictor. In the genlin line, we list our categorical predictor prog after "by" and our continuous predictor math after "with". Both appear in the model line. We use the covb=robust option in the criteria line to obtain robust standard errors for the parameter estimates as recommended by Cameron and Trivedi (2009) to control for mild violation of the distribution assumption that the variance equals the mean. Finally, we ask SPSS to print out the model fit statistics, the summary of the model effects, and the parameter estimates.

```
GENLIN num_awards BY prog WITH math
/MODEL prog math INTERCEPT=YES
DISTRIBUTION=POISSON LINK=LOG
/CRITERIA COVB=ROBUST
/PRINT FIT SUMMARY SOLUTION.
```

Goodness of Fit[b]

	Value	df	Value/df
Deviance	189.450	196	.967
Scaled Deviance	189.450	196	
Pearson Chi-Square	212.144	196	1.082
Scaled Pearson Chi-Square	212.144	196	
Log Likelihood[a]	-182.752		
Akaike's Information Criterion (AIC)	373.505		
Finite Sample Corrected AIC (AICC)	373.710		
Bayesian Information Criterion (BIC)	386.698		
Consistent AIC (CAIC)	390.698		

Dependent Variable: num_awards
Model: (Intercept), prog, math

a. The full log likelihood function is displayed and used in computing information criteria.
b. Information criteria are in small-is-better form.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
98.223	3	.000

Dependent Variable: num_awards
Model: (Intercept), prog, math

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	60.952	1	.000
prog	14.838	2	.001
math	45.195	1	.000

Dependent Variable: num_awards
Model: (Intercept), prog, math

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-4.877	.6297	-6.112	-3.643	59.984	1	.000
[prog=1]	-.370	.4004	-1.155	.415	.853	1	.356
[prog=2]	.714	.2986	.129	1.299	5.717	1	.017
[prog=3]	0[a]	.	.	.	.	.	.
math	.070	.0104	.050	.091	45.195	1	.000
(Scale)	1[b]						

Dependent Variable: num_awards
Model: (Intercept), prog, math

a. Set to zero because this parameter is redundant.
b. Fixed at the displayed value.

Sometimes, we might want to present the regression results as incident rate ratios. These IRR values are equal to our coefficients from the output above exponentiated and we can ask SPSS to printsolution(exponentiated).

GENLIN num_awards BY prog WITH math
/MODEL prog math INTERCEPT=YES DISTRIBUTION=POISSON LINK=LOG
/CRITERIA METHOD=FISHER(1) SCALE=1 COVB=ROBUST
/PRINT SOLUTION (EXPONENTIATED).

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-4.877	.6297	-6.112	-3.643	59.984	1	.000	.008	.002	.026
[prog=1]	-.370	.4004	-1.155	.415	.853	1	.356	.691	.315	1.514
[prog=2]	.714	.2986	.129	1.299	5.717	1	.017	2.042	1.137	3.667
[prog=3]	0[a]	.	.	.	.	.	.	1	.	.
math	.070	.0104	.050	.091	45.195	1	.000	1.073	1.051	1.095
(Scale)	1[b]									

Dependent Variable: num_awards
Model: (Intercept), prog, math

a. Set to zero because this parameter is redundant.
b. Fixed at the displayed value.

The output above indicates that the incident rate for is 2.042 times the incident rate for the reference group, . Likewise, the incident rate for is 0.691 times the incident rate for the reference group holding the other variables at constant. The percent change in the incident rate of num_awards is an increase of 7% for every unit increase in math.

Recall the form of our model equation:

log(num_awards) = Intercept + b1(prog=1) + b2(prog=2) + b3math.

This implies:

num_awards = exp(Intercept + b1(prog=1) + b2(prog=2)+
b3math) = exp(Intercept) * exp(b1(prog=1)) * exp(b2(prog=2))
*** exp(b3math)**

The coefficients have an additive effect in the log(y) scale and the
IRR have a multiplicative effect in the y scale.

For additional information on the various metrics in which the results can be
presented, and the interpretation of such, please see Regression Models for
Categorical Dependent Variables Using Stata, Second Edition by J. Scott Long
and Jeremy Freese (2006).

To understand the model better, we can use the
emmeans command to calculate the predicted counts at each
level of prog, holding all other variables (in this example,
math)

in the model at their means.

GENLIN num_awards BY prog WITH math
/MODEL prog math INTERCEPT=YES
DISTRIBUTION=POISSON LINK=LOG
/CRITERIA METHOD=FISHER(1) SCALE=1
COVB=ROBUST
/PRINT NONE
/EMMEANS TABLES=prog SCALE=ORIGINAL.

Estimates

type of program	Mean	Std. Error	95% Wald Confidence Interval Lower	95% Wald Confidence Interval Upper
general	.21	.063	.09	.33
academic	.62	.088	.45	.80
vocation	.31	.083	.14	.47

Covariates appearing in the model are fixed at the following values: math=52.65

In the output above, we see that the predicted number of events for level 1 of prog is about .21, holding math at its mean. The predicted number of events for level 2 of prog is higher at .62, and the predicted number of events for level 3 of prog is about .31. Note that the

predicted count of level 2 of prog is (.62/.31) = 2.0 times higher than the predicted count for level 3 of prog. This matches what we saw in the IRR output table.

Below we will obtain the predicted counts for each value of prog at two set values of math: 35 and 75.

```
GENLIN num_awards BY prog WITH math
/MODEL prog math INTERCEPT=YES
DISTRIBUTION=POISSON LINK=LOG
/PRINT NONE
/EMMEANS TABLES=prog CONTROL =math(35) SCALE=ORIGINAL.
```

Estimates

type of program	Mean	Std. Error	95% Wald Confidence Interval	
			Lower	Upper
general	.06	.024	.01	.11
academic	.18	.055	.07	.29
vocation	.09	.031	.03	.15

Covariates appearing in the model are fixed at the following values: math=35.00

```
GENLIN num_awards BY prog WITH math
/MODEL prog math INTERCEPT=YES
DISTRIBUTION=POISSON LINK=LOG
/PRINT NONE
```

/EMMEANS TABLES=prog CONTROL =math(75) SCALE=ORIGINAL.

Estimates

type of program	Mean	Std. Error	95% Wald Confidence Interval	
			Lower	Upper
general	1.01	.409	.21	1.82
academic	3.00	.510	2.00	4.00
vocation	1.47	.548	.39	2.54

Covariates appearing in the model are fixed at the following values: math=75.00

The table above shows that with prog=1 and math held at 35, the average predicted count (or average number of awards) is about .06; when math = 75, the average predicted count for prog=1 is about 1.01. If we look at these predicted counts at math = 35 and math = 75, we can see that the ratio is (1.01/0.06) = 16.8. This matches (within rounding error) the IRR of 1.0727 for a 40 unit change: 1.0727^40 = 16.1.

Things to consider

See also

Regression Models for Categorical and Limited Dependent Variables.

References