

Which regression model is better suited for analyzing count data: Negative Binomial or Poisson?

Authored by
stats writer

April 27, 2024

RECOMMENDED CITATION

stats writer (2024). *Which regression model is better suited for analyzing count data: Negative Binomial or Poisson?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=139944>

The Negative Binomial and Poisson regression models are both commonly used for analyzing count data. However, the choice of which model is better suited depends on the specific characteristics of the data being analyzed. In general, the Negative Binomial model is more appropriate for count data that is overdispersed, meaning that the variance is greater than the mean. On the other hand, the Poisson model is better suited for count data that is not overdispersed and has a mean that is close to the variance. Therefore, it is important to carefully examine the characteristics of the count data before deciding which regression model to use for analysis.

Negative Binomial vs. Poisson: Choose a Regression Model

Negative binomial regression and Poisson regression are two types of regression models that are appropriate to use when the is represented by discrete count outcomes.

Here are a few examples of response variables that represent discrete count outcomes:

**The number of students who graduate from a certain program
The number of traffic accidents at a certain intersection
The number of participants who finish a marathon
The number of returns in a given month at a retail store**

If the variance is roughly equal to the mean, then a Poisson regression model typically fits a dataset well.

However, if the variance is significantly greater than the mean, then a negative binomial regression model is typically able to fit the data better.

There are two techniques we can use to determine if Poisson regression or negative binomial regression is more appropriate to use for a given dataset:

1. Residual Plots

We can create a residual plot of the standardized residuals vs. predicted values from a regression model.

If the majority of the standardized residuals fall within the range of -2 and 2 then a Poisson regression model is likely appropriate.

However, if many residuals fall outside of this range then a negative binomial regression model will likely provide a better fit.

2. Likelihood Ratio Test

We can fit a Poisson regression model and a negative binomial regression model to the same dataset and then perform a Likelihood Ratio Test.

If the p-value of the test is less than some significance level (e.g. 0.05) then we can conclude that the negative binomial regression model offers a significantly better fit.

The following example shows how to use both of these techniques in R to determine whether a Poisson regression or negative binomial regression model is better to use for a given dataset.

Example: Negative Binomial vs. Poisson Regression

Suppose we want to know how many scholarship offers a high school baseball player in a given county receives based on their school division ("A", "B", or "C") and their college entrance exam score (measured from 0 to 100).

Use the following steps to determine if a negative binomial regression model or Poisson regression model offers a better fit to the data.

The following code creates the dataset we will work with, which includes data on 1,000 baseball players:

```
#make this example reproducible  
set.seed(1)
```

```
#create dataset
```

```
data <- data.frame(offers = c(rep(0, 700), rep(1, 100),  
rep(2, 100),  
rep(3, 70), rep(4, 30)),  
division = sample(c('A', 'B', 'C'), 100, replace = TRUE),  
exam = c(runif(700, 60, 90), runif(100, 65, 95),  
runif(200, 75, 95)))
```

```
#view first six rows of dataset
```

```
head(data)
```

```
offers division exam
```

```
1 0 A 66.22635
```

```
2 0 C 66.85974
```

```
3 0 A 77.87136
```

```
4 0 B 77.24617
```

```
5 0 A 62.31193
```

```
6 0 C 61.06622
```

Step 2: Fit a Poisson Regression Model & Negative Binomial Regression Model

The following code shows how to fit both a Poisson regression model and negative binomial regression model to the data:

#fit Poisson regression model

```
p_model <- glm(offers ~ division + exam, family =  
'poisson', data = data)
```

#fit negative binomial regression model

```
library(MASS)
```

```
nb_model <- glm.nb(offers ~ division + exam, data =  
data)
```

Step 3: Create Residual Plots

The following code shows how to produce residual plots for both models.

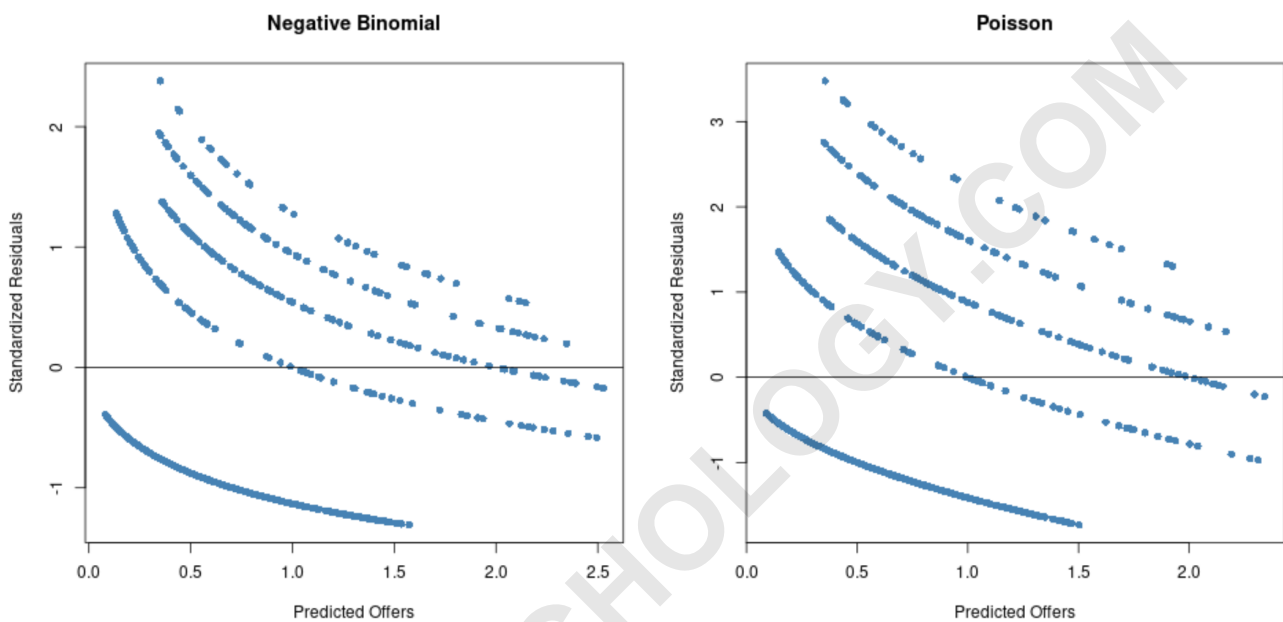
#Residual plot for Poisson regression

```
p_res <- resid(p_model)  
plot(fitted(p_model), p_res, col='steelblue', pch=16,  
xlab='Predicted Offers', ylab='Standardized Residuals',  
main='Poisson')  
abline(0,0)
```

#Residual plot for negative binomial regression

```
nb_res <- resid(nb_model)  
plot(fitted(nb_model), nb_res, col='steelblue', pch=16,
```

```
xlab='Predicted Offers', ylab='Standardized Residuals',  
main='Negative Binomial')  
abline(0,0)
```



From the plots we can see that the residuals are more spread out for the Poisson regression model (notice that some residuals extend beyond 3) compared to the negative binomial regression model.

This is a sign that a negative binomial regression model is likely more appropriate since the residuals of that model are smaller.

Step 4: Perform a Likelihood Ratio Test

Lastly, we can perform a likelihood ratio test to determine if there is a statistically significant difference in the fit of the two regression models:

```
pchisq(2 * (logLik(nb_model) - logLik(p_model)), df = 1,  
lower.tail = FALSE)
```

```
'log Lik.' 3.508072e-29 (df=5)
```

The p-value of the test turns out to be 3.508072e-29, which is significantly less than 0.05.

Thus, we would conclude that the negative binomial regression model offers a significantly better fit to the data compared to the Poisson regression model.