

# Count data: Poisson Models

Jan van den Broek

January, 2020



**Universiteit Utrecht**

Count data:  
Poisson  
Models

Jan van den  
Broek

## Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

# Counts

# Count data

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

	City			
Age	City 1	City 2	City 3	City 4
40-54	11	13	4	5
55-59	11	6	8	7
60-64	11	15	7	10
65-69	10	10	11	14
70-74	11	12	9	8
>75	10	2	12	7

The number of lung cancer cases in 4 Danish cities between 1968 and 1971.

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

# The Poisson distribution

# Poisson distribution

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

Call the count variable  $Y$  and the observed outcome  $y$ .

So the first of the 24 counts can be represented by  $Y_1$  and its observation by  $y_1$ . For the  $i^{th}$  count this is  $Y_i$  with observation  $y_i$ .

The **Poisson** probability of observing  $y_i$  for the  $i^{th}$  count is then

$$P(Y_i = y_i) = \frac{e^{-\mu} \mu^{y_i}}{y_i!}$$

The parameter  $\mu$  is the population mean of the counts. This means that if we had a large population of counts the mean of all these count would be  $\mu$ .

# Poisson distribution

Count data:  
Poisson  
Models

Jan van den  
Broek

The variance of all these count would also be  $\mu$ .

So, with a poisson distribution the mean and the variance are the same.

There seems to be some logic in that: if counts have a small mean, the variance can not be large since counts can not be smaller than zero.

For counts with a large mean say 200, some relatively small values or large values for the counts might occur and thus the variance can be large.

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

# Poisson distribution

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

One can now use the poisson distribution to calculate the probability of the outcomes.

If for instance the population mean is known to be 10 then the probability of observing no cases is  $e^{-10} = 4.5 \times 10^{-5}$  and that of observing 5 cases is  $\frac{e^{-10} 10^5}{5!} = 0.04$

So, for every of the 24 observations one can write down the probability of observing that observation.

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

## Poisson likelihood



# Likelihood

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

The product of these probabilities is the likelihood which depends on the population mean  $\mu$ , so

$$L(\mu) = \frac{e^{-\mu} \mu^{y_1}}{y_1!} \cdot \frac{e^{-\mu} \mu^{y_2}}{y_2!} \cdot \dots \cdot \frac{e^{-\mu} \mu^{y_{24}}}{y_{24}!}.$$

and the log-likelihood is

$$l(\mu) = \sum_{i=1}^{24} [-\mu + y_i \ln(\mu) - \ln(y_i!)]$$

That value for  $\mu$  that maximises the log-likelihood and thus the likelihood is taken as an estimate for  $\mu$  the **maximum likelihood estimate**  $m$  which in this case is the mean of the 24 counts.

# Likelihood

Count data:  
Poisson  
Models

Jan van den  
Broek

A measure for the peakedness of the log-likelihood is the second derivative at it's maximum :  $l''(m)$ .

For a flat function the second derivative is small, for a peaked function it is large. The amount of information is found by calculating the negative of the second derivative:

$$Information = -l''(m)$$

From this the standard error can be calculated:

$$standard\ error = \sqrt{\frac{1}{Information}}$$

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

**evidence**

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

evidence

# Evidence

Count data:  
Poisson  
Models

Jan van den  
Broek

Now the fact that the observations com from 4 different cities needs to be modelled.

A linear model for the 4 city groups is

$\beta_0 + \beta_1 city2 + \beta_2 city3 + \beta_3 city4$ , were  $city2$  is a indicator variable having value zero except for city 2.  $city3$  and  $city4$  only have one's for city 3 and city 4.

This is just a linear model for the one way anova case using indicator variables.

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

# Evidence

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

This linear model is used not for the population mean but for the logarithm of the population mean.

$$\ln(\mu) = \beta_0 + \beta_1 \text{city2} + \beta_2 \text{city3} + \beta_3 \text{city4}$$

One can also write the model as  $\mu = e^{\beta_0 + \beta_1 \text{city2} + \beta_2 \text{city3} + \beta_3 \text{city4}}$  and plug this in the likelihood. The first derivatives w.r.t. the  $\beta$ 's give the maximum likelihood estimates  $b_0, b_1, b_2$  and  $b_3$ .

# Evidence

Count data:  
Poisson  
Models

Jan van den  
Broek

the interpretation of these  $b$ 's can be seen as follows.

For city 1 all the indicator variables are zero so the model for city 1 is  $\ln(m_1) = b_0$ . So  $e^{b_0}$  is the estimated mean count for city 1. For city 4 for instance only the indicator for city 4 is one so the model for city 4 is:  $\ln(m_4) = b_0 + b_3$ . Subtracting gives  $\ln(m_4) - \ln(m_1) = \ln\left(\frac{m_4}{m_1}\right) = b_0 + b_3 - b_0 = b_3$ .

Thus  $e^{b_3}$  shows how much larger  $m_4$  is as compared to  $m_1$ . It is a **ratio**. This means that  $b_3$  is the **log-ratio** comparing the mean of city 4 with that of city 1.

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

# Evidence

Count data:  
Poisson  
Models

Jan van den  
Broek

In order to see whether or not the cities differ in the mean number of cases one can fit a model with cities in it (Model 1) and one without the cities (Model 0) and then compare these models.

**Model 0:** The model is  $\ln(\mu) = \beta_0$  or  $\mu = e^{\beta_0}$

This model for  $\mu$  is used in the log-likelihood. The derivative w.r.t.  $\beta_0$  gives the maximum likelihood estimator  $b_0$  for  $\beta_0$ , the second derivative gives the standard error.

The maximum of the log-likelihood,  $l_0$ , is obtained by plugging in the value of the above maximum likelihood estimate. So take for  $\mu$  in the log-likelihood  $e^{b_0}$  the fitted value for  $\mu$ .

The maximum for the likelihood then is  $L_0 = e^{l_0}$ .  $L_0$  gives the probability of the data if the estimated model 0 generated the data

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

# Evidence

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

**Model 1:** The model is

$\ln(\mu) = \beta_0 + \beta_1 \text{city2} + \beta_2 \text{city3} + \beta_3 \text{city4}$  or  
 $\mu = e^{\beta_0 + \beta_1 \text{city2} + \beta_2 \text{city3} + \beta_3 \text{city4}}$ . Plug in this value for  $\mu$  in the log-likelihood.

The first derivatives w.r.t the  $\beta$ 's gives the maximum likelihood estimators for the  $\beta$ 's and the second derivatives gives the standard errors for these estimates.

The maximum of the log-likelihood,  $l_1$ , is obtained by plugging in the maximum likelihood estimates so by plugging in the fitted values. The maximum for the likelihood then is  $L_1 = e^{l_1}$ .  $L_1$  gives the probability of the data if the estimated model 1 generated the data.



# Evidence

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

If  $L_1$  is larger than  $L_0$  then the observed data is more likely using model 1, this model makes the data more probable. In that case under model 1, one better understands why the data is observed as it is. Thus one calculates the likelihood ratio:  $\frac{L_1}{L_0}$ .

One can use **Akaike's information criterion (AIC)**:

$AIC = -2 \cdot (\log\text{-likelihood}) + 2 \cdot (\text{number of parameters in the model}) = -2 \cdot l + 2 \cdot p$  where  $p$  is the number of parameters in the model.

That model is best that has the lowest AIC as compared to the others.

If the difference in AIC between models is small then the model with the smallest number of parameters is chosen (Occam's Razor) AIC is a measure of how good the model fits to the data as compared to the other models that are fitted

# Evidence

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

One can calculate the likelihood ratio statistic :  
 $2 \cdot \ln \left( \frac{L_1}{L_0} \right) = 2(l_1 - l_0)$ . This statistic has approximately a chi-squared distribution with  $df$  degrees of freedom where  $df$  is the difference of the number of parameters between the models. The AIC values and the likelihood ratio test are in the output below. City can be left out of the model.

```
## Single term deletions
##
## Model:
## cases ~ factor(city)
##           Df Deviance    AIC    LRT Pr(>Chi)
## <none>           25.621 130.01
## factor(city)  3   27.704 126.09 2.0833  0.5553
```

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

**Profile  
log-likelihood  
interval**

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

# Profile log-likelihood interval

# Profile log-likelihood interval

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

A way to look at the (log-)likelihood as a function of one parameter, e.g. a log-ratio  $\beta$  (i.e. the parameter for city 2), is to focus on this parameter by holding it constant.

Then for each such a constant value of this parameter the log likelihood is maximized over all other parameters.

For every value of the fixed parameter, a maximum value for the log-likelihood is obtained.

To plot such a (log-)likelihood just put the different values of the parameter ( $\beta$ ) on the horizontal axis and the value of the maximized (log-)likelihood on the vertical axis.

This (log-)likelihood is called **the profile (log-)likelihood**.

# Profile log-likelihood interval

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

This profile likelihood is used to determine confidence intervals that are based on the likelihood ratio test.

The likelihood ratio test statistic was  $2 \cdot \ln \left( \frac{L_1}{L_0} \right) = 2(l_1 - l_0)$ .  $l_1$  is the maximum value of the log-likelihood if the alternative hypothesis is true.

It is the maximum value when the log likelihood is maximized over all the parameters in the model.

This maximum value can be denoted as  $l_1$ .

# Profile log-likelihood interval

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

$l_0$  is the maximum value of the log-likelihood if the null-hypothesis is true, that is it is the maximum value for the log-likelihood for a specific value for  $\beta$ .

This maximum value of the log-likelihood can be denoted as  $l_0$  and can thus be seen as the maximum value of the log-likelihood under the null-hypothesis.

The likelihood ratio test statistic has approximately a chi-squared distribution with one degree of freedom.

So the likelihood ratio test with a significance level of 0.05 rejects the null-hypothesis if the likelihood ratio test statistic  $(2(l_1 - l_0))$  is larger than  $\chi^2_{0.95}$  which is in the case of 1 degree of freedom equal to 3.84.

# Profile log-likelihood interval

Count data:  
Poisson  
Models

Jan van den  
Broek

For all those values for  $\beta$  for which the likelihood ratio test statistic is smaller than 3.84, the conclusion is that the null-hypothesis is not rejected.

This is an interpretation of a 95% confidence interval: all those values for the parameter of interest that, when put in the null-hypothesis would lead to not rejecting the null-hypothesis.

So look for which values for the likelihood ratio test one does not reject the null-hypothesis or equivalent for which values of the profile log-likelihood ( $l_0$ ) the null hypothesis is not rejected.

This determines the confidence interval.

An other interpretation for a 95% confidence interval is: The probability that this interval contains the population value of the parameter is 0.95.

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

**Model  
checking**

Modeling  
risks:  
population  
size

Modeling  
rates: periods

# Model checking



# Model checking

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

Every observation has its own fitted value . This fitted value for individual count  $i$  can be calculated by using

$m_i = e^{b_0 + b_1 \text{city}2 + b_2 \text{city}3 + b_3 \text{city}4}$  and plugging in the values for the city indicators for that individual count.

The contribution for this observation to the maximum value of the likelihood is:  $\frac{e^{-m_i} m_i^{y_i}}{y_i!}$ .

In order to see how far the fitted value for this individual count is from the observation of this individual count one could calculate the likelihood contribution when the fitted value is replaced by the observation:  $\frac{e^{-y_i} y_i^{y_i}}{y_i!}$ .

# Model checking

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

The contribution for this individual count to the deviance is

$$D(y_i) = 2 \cdot \left[ \ln \left( \frac{e^{-y_i} y_i^{y_i}}{y_i!} \right) - \ln \left( \frac{e^{-m_i} m_i^{y_i}}{y_i!} \right) \right]$$

.

If this number is large then this individual count contributes a large number to the deviance and, if there are more of these counts, to a bad fit.

If this number is small then this individual count contributes a small value to the deviance implying the model fits this individuals count observation.

Note that the sum of all these individual contributions is the deviance for the model used here.

# Model checking

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

The deviance residuals are defined as the square root from the individual deviance contributions multiplied with a plus or a minus sign depending on whether or not the observation are larger or smaller then the fitted values:

$$res_{dev} = \text{sign}(y_i - m_i) \cdot \sqrt{D(y_i)}$$

A large residual means that this individual contributes a large individual deviance, thus contributing to a bad fit. Or, to put it differently: a large deviance residual means that the difference between the fitted value and the observation in terms of log-likelihoods is large for this individual.

One can now make a plot of the deviance residuals and the fitted values. One can then check whether or not there are large residuals

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

**Modeling  
risks:  
population  
size**

Modeling  
rates: periods

# Modeling risks: population size

# Modeling risks: population size

In the lung cancer research every city has a certain population size in every age group, but this size was considered to be very large as compared to the number of cases. The population sizes of the age groups were:

	City			
Age	City 1	City 2	City 3	City 4
40-54	3059	2879	3142	2520
55-59	800	1083	1050	878
60-64	710	923	895	839
65-69	581	834	702	631
70-74	509	634	535	539
>75	605	782	659	619

# Modeling risks: population size

Count data:  
Poisson  
Models

Jan van den  
Broek

Instead of assuming that the population sizes are large enough, one better tries to get them in the model.

One way to do that is to model the mean number of cases per individual in the population that is model  $\frac{\mu}{\text{population size}}$ .

So now a risk is modelled: the mean number of cases per individual.

The model then becomes

$\ln\left(\frac{\mu}{\text{population size}}\right) = \beta_0 + \beta_1 \text{age} + \beta_2 \text{city2} + \dots$  in the short notation mentioned above.

This is the same as

$\ln(\mu) = \beta_0 + \ln(\text{population size}) + \beta_1 \text{age} + \beta_2 \text{city2} + \dots$ .

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

# Modeling risks: population size

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

That is  $\ln(\text{population size})$  is in the model without a  $\beta$ -parameter.

To put it differently the  $\beta$  for the variable  $\ln(\text{population size})$  is taken to be one.

One calls a variable like this an **offset** variable.

Putting this in the model might be important if the population sizes differ a lot among the different groups.

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

## Modeling rates: periods



# Modeling rates: periods

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

A common feature of count data is a period of observation over which these counts occur.

In the urinary track disease example for each patient the number of urinary track episodes was determined. Besides this the follow-up time for each patient was recorded. An individual with a long follow-up time might have more chance of showing the event of interest as compared to a patient with a shorter follow-up time.

Using  $\ln(\text{followuptime})$  as an **offset** in the model means that the mean number of events per time unit is modeled, which is a rate:

$$\ln(\mu) = \beta_0 + \ln(\text{followup}) + \dots \text{ or}$$

$$\ln\left(\frac{\mu}{\text{followup}}\right) = \beta_0 + \dots$$

# Modeling rates: periods

Count data:  
Poisson  
Models

Jan van den  
Broek

Counts

The Poisson  
distribution

Poisson  
likelihood

evidence

Profile  
log-likelihood  
interval

Model  
checking

Modeling  
risks:  
population  
size

Modeling  
rates: periods

Sometimes the data can be more aggregated.

Suppose that in a research one is interested in the age effect on the number of deaths due to a certain cause. Suppose further that age is a factor with say 5 levels. In the first age group there were 3 subjects of which one died. The first subject was observed 3 weeks, the second for 5 weeks and the third for 7 weeks. The there are 15 person-weeks observed. This is an example of what is called **person-time**. Person-time shows how long several person were at risk.

The death rate then is  $\frac{1}{15} = 0.067$  per person-week. Because this is the number of new cases per person-time this is also called the incidence rate.