Poisson Regression

Prerequisite and related topics

Basics of statistics are an advantage for understanding this description:

www.weibull.de/COM/Data_Analysis.pdf

Key words: Poisson-Regression – countable events – Error rate – Generalized Linear Model GLM – Maximum-Likelihood-Estimation MLE

Introduction

The Poisson distribution describes countable events that occur in a defined time interval. This can be, for example the number of error messages within a week. Poisson regression is treated in connection with the so-called generalized linear regression model (GLM).

Purpose and benefit

The goal is to determine a model for countable events based on the Poisson distribution, which is well suited for this purpose.

Basics

The Poisson density function (probability that the number of errors y occurs) is defined as:

$$g(x) = \frac{\lambda^y}{y!}\, e^{-\lambda}$$

y : number of events or errors integer

λ : Poisson- or Error rate

Normally, the number of errors is declared as x. Since in the following x is required for the influence parameters, y should be used here. As with lifetime models, the relationship for the Poisson rate can be best represented logarithmically:

$$\ln(\lambda) = b_o + b_1 x_1 + b_2 x_2 + \cdots + b_z x_z$$

b : coefficients of the model

x : influence parameters

therefore, it is:

$$\lambda = e^{b_o + b_1 x_1 + b_2 x_2 + \cdots + b_z x_z}$$

Assuming that the observed events are y_i Poisson-distributed, the expected value is $E[y_i] = \lambda_i$. Each observation must be independent of each other and random. The best estimator for the coefficients b is given if the product of all probabilities (likelihood ⇨ L) of each i-th observation is a maximum. The following applies to this:

$$L(b_o, b_1, b_2, \dots b_z) = \prod_{i=1}^{n} \frac{\lambda_i^{y_i}}{y_i!}\, e^{-\lambda_i}$$

The determination of the coefficients cannot be done analytically, as in multiple regression, because y is both in the exponent and in the denominator with faculty. This makes it clear that the result must differ from the method of the least square estimation.

However, the differences are small and the coefficients determined herewith can be used as a starting condition of the iterative calculation. The so-called Newton-Raphson method is most often mentioned for this iterative calculation. The logarithmic form can also be used to search for the maximum probabilities, resulting in the so-called log likelihood function LL:

$$LL(b_o, b_1, b_2, \dots b_z) = \sum_{i=1}^{n} y_i \ln \lambda_i - \lambda_i - \ln(y_i!)$$

The goal is to find the coefficients where the sum of the right side is a maximum. Because of this the method is well known as Maximum Likelihood Estimation MLE for short. Depending on the cancellation criterion or the number of selected iteration steps, slightly different results can arise!

The dispersion of the coefficients sb is determined by:

$$s_{b,i} = diag(\sqrt{(X^T W)^{-1}})$$

with the weight

$$w_{j,i} = x_{j,i} \cdot \mu_i \quad \text{and} \quad \mu_i = \exp(b_o + b_1 x_{1,i} + b_2 x_{2,i} + \cdots)$$

To determine whether the coefficients are significant, the p-value is obtained by

$p\text{-}value$ = 2·*DistributionNormal*(-|z|)

with

$$z = \frac{b_i}{s_{b,i}}$$

The parameter is regarded as significant if $p\text{-}value < 0{,}05$, see:

www.weibull.de/COM/Hypothesis_Tests.pdf

To assess the overall model, the sum of the deviation squares is calculated as a so-called deviance. The model scattering is:

$$D_{(b)} = 2 \sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{e^{Xb}}\right) - (y_i - e^{Xb}) \qquad \text{with } Xb = b_o + b_1 x_{1,i} + b_2 x_{2,i} + \cdots$$

and where only the constant b_o is in the model:

$$D_{(bo)} = 2 \sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{e^{bo}}\right) - (y_i - e^{bo})$$

which results in the so-called coefficient of determination with:

$$R^2 = 1 - \frac{D_{(b)}}{D_{(bo)}}$$

Instead of the adjusted R², like at the multiple regression, here a corrected R^2_{cor} is used:

$$R^2_{kor} = R^2 - \frac{z}{D_{(bo)}}$$

Poisson-Regression with Intercept

If there are no events for certain combinations of the influencing parameters, there is a problem that ln(yi=0) is not possible. Therefore, an offset should be used, which is called intercept here and which can be set individually for each observation. It is recommended to set this intercept only for the case of yi=0, e.B. to 0.01.

Normalized Model

As with multiple regression, it is recommended that the influence parameters x are normalized:

$$x_{norm} = \frac{(x-\bar{x})}{x_{max}-x_{min}}$$

This makes the range of values between -1 .. +1 and the determination of the model becomes more stable, the significance becomes clearer. In contrast to Visual-XSel, the many statistical programs use the original value ranges as the default setting, so the coefficients differ and the models may be different.

Further characteristics

Another very frequently mentioned characteristics for likelihood-based models and thus also for Poisson regression is the Akaike Information Criterion, AIC for short, named after the Japanese Hirotsugu Akaike AIC:

$$AIC = -2\,LL + 2(z+1)$$

The smaller AIC is, the better, but AIC alone cannot be interpreted, because there is no universal limit value for a best value. AIC is therefore more used for comparisons between different models. As with the R², models with more terms, such as interaction terms, usually provide better AIC values, although z increases and is received with a factor of 2. However, these terms may only remain in the model if they are also significant (see p-value)

Another criterion is the so-called Bayesian Information Criterion, BIC for short (named after the English statistician Thomas Bayes). This key figure is very similar to the AIC and also takes into account the number of measures n:

$$BIC = -2\,LL + (z+1)\cdot\ln(n)$$

From n > 7, BIC becomes larger than AIC and punishes more complex models more severely. However, the influence of n should not be understood in such a way that the smallest possible data sets should be used. More information is basically an advantage for modeling.
Although AIC and BIC are often mentioned as important characteristics, the practical benefits seem limited due to the disadvantages mentioned.

Confidence intervals

The χ^2 distribution can be used for the confidence intervals of the model predictions. With a usual confidence range of 90%, the probability of error is α = 10%. The two-sided trust scope for the number (error) events y is:

$$\frac{1}{2}\,\chi^2_{\frac{\alpha}{2},2y} \leq y \leq \frac{1}{2}\,\chi^2_{1-\frac{\alpha}{2},2(y+1)}$$

Example wastage in a manufacturing process

In an experimental design, the wastage of components was counted as a function of temperature, pressure, time and particle grain size in the material.

The model yields the following coefficients

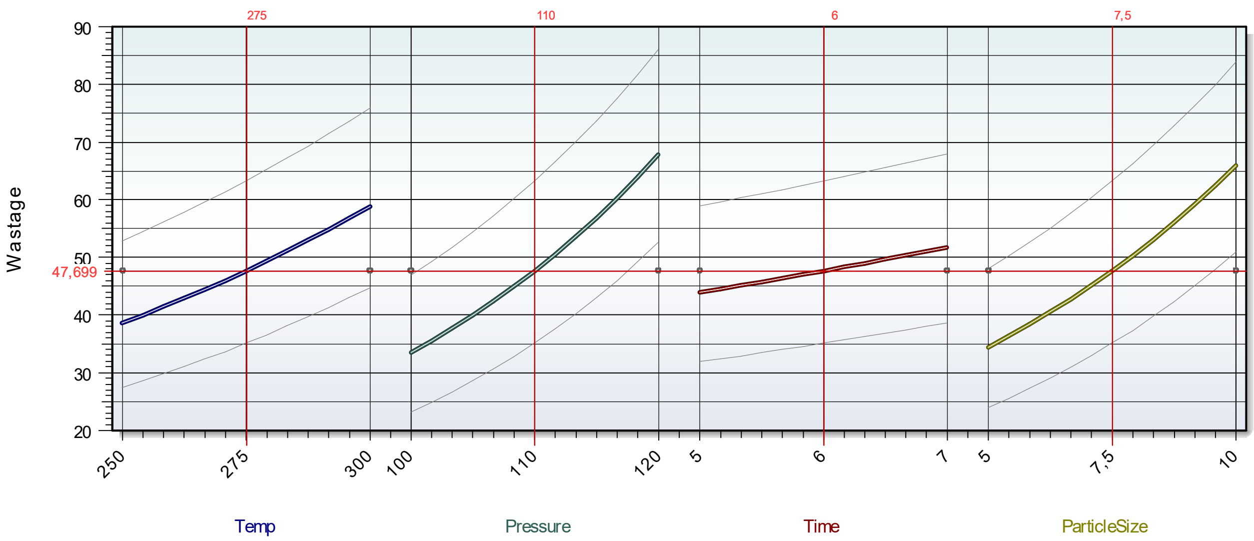
	Coefficient	p-value
Constant	-3,79657	
Temp	0,008435	0.000
Pressure	0,035221	0.000
Time	0,082689	0,015
ParticleSize	0,129531	0.000

R^2 = 0.926	DF = 11	AIC = 119.3
R^2_{kor} = 0.910		BIC = 123.1

No	Temp[C]	Pressure[t	Time[min]	ParticleSiz	Wastage
1	250	100	5	5	7
2	250	100	5	10	36
3	250	100	7	5	14
4	250	100	7	10	45
5	250	120	5	5	39
6	250	120	5	10	74
7	250	120	7	5	48
8	250	120	7	10	84
9	300	100	5	5	26
10	300	100	5	10	59
11	300	100	7	5	34
12	300	100	7	10	69
13	300	120	5	5	62
14	300	120	5	10	99
15	300	120	7	5	71
16	300	120	7	10	109

All parameters are significant (*p-value* < 0,05).

The following curve diagram shows the relationships as a continuous function, the results of which are to be rounded up or down.



The confidence ranges are relatively strongly asymmetrical, see chapter Confidence intervals.

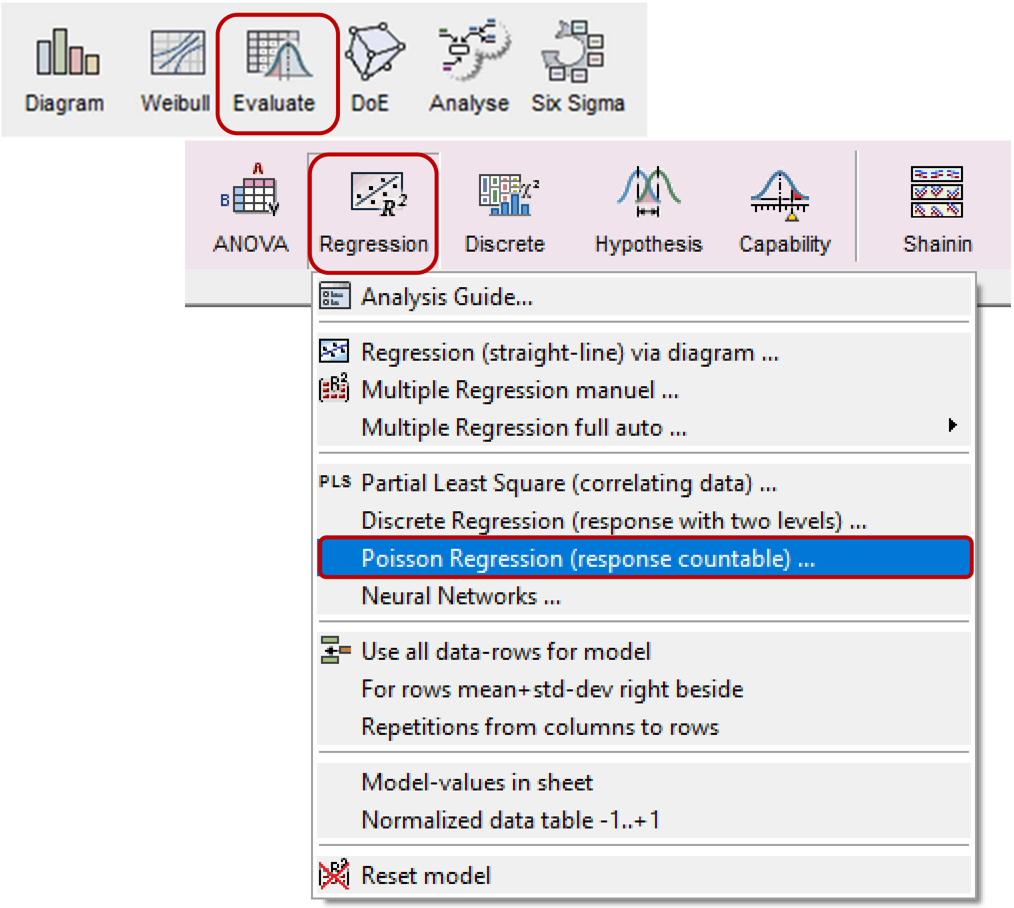
Literature

Generalized Linear Model, 2nd Edition, McCullagh, J.A. Nelder, Chapman and Hall

Using Visual-XSel 17.0

www.crgraph.com

In Visual-XSel the Poisson-Regression is available at version 17.0001



The next steps follow the same procedure as in the program description

www.weibull.de/COM/Data_Analysis.pdf