Chapter 21
Poisson Regression for Outcome Rates (50 Patients)

1 General Purpose

Poisson regression is different from linear en logistic regression, because it uses a log transformed dependent variable. For rates, defined as numbers of events per person per time unit, Poisson regression is very sensitive and probably better than standard regression methods.

2 Schematic Overview of Type of Data File

Outcome	predictor	predictor	predictor	weight
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

T.J. Cleophas, A.H. Zwinderman, *SPSS for Starters and 2nd Levelers*,
DOI 10.1007/978-3-319-20600-4_21

3 Primary Scientific Question

Can a multiple poisson regression be used to estimate the effect of certain predictors on numbers of clinical events.

4 Data Example

Fifty patients were followed for numbers of episodes of paroxysmal atrial fibrillation (PAF), while on treated with two parallel treatment modalities. The data file is below. The scientific question was: do psychological and social factors affect the rates of episodes of paroxysmal atrial fibrillation.

Paf	Treat	Psych	Soc	Weight
4	1	56,99	42,45	73
4	1	37,09	46,82	73
2	0	32,28	43,57	76
3	0	29,06	43,57	74
3	0	6,75	27,25	73
13	0	61,65	48,41	62
11	0	56,99	40,74	66
7	1	10,39	15,36	72
10	1	50,53	52,12	63
9	1	49,47	42,45	68

outcome = numbers of episodes of paroxysmal atrial fibrillation (paf)
treat = treatment modality predictor
psych = psychological score predictor
soc = social score predictor
weight = days of observations

The entire data file is in extras.springer.com, and it is entitled "chapter21poissoncontinuous". First, we will perform a linear regression analysis with paf as outcome variable and the other variables as predictors. Start by opening the data file in SPSS.

5 Multiple Linear Regression

For analysis the statistical model Linear in the module Regression is required.

Command:

Analyze....Regression....Linear....Dependent Variable: episodes of paroxysmal atrial fibrillation....Independent: treatment modality, psychological score, social score, days of observation....click OK.

Coefficients[a]

Model		Unstandardized coefficients		Standardized coefficients		
		B	Std. error	Beta	t	Sig.
1	(Constant)	49,059	5,447		9,006	,000
	Treat	−2,914	1,385	−,204	−2,105	,041
	Psych	,014	,052	,036	,273	,786
	Soc	−,073	,058	−,169	−1,266	,212
	Days	−,557	,074	−,715	−7,535	,000

[a]Dependent Variable: paf

The above table show that treatment modality is weakly significant, and psychological and social score are not. Furthermore, days of observation is very significant. However, it is not entirely appropriate to include this variable if your outcome is the numbers of events per person per time unit. Therefore, we will perform a linear regression, and adjust the outcome variable for the differences in days of observation using weighted least square regression.

6 Weighted Least Squares Analysis

For analysis the statistical model Linear in the module Regression is required.

Command:

Analyze....Regression....Linear....Dependent: episodes of paroxysmal atrial fibrillation....Independent: treatment modality, psychological score, social score ....WLS Weight: days of observation.... click OK.

Coefficients[a,b]

Model		Unstandardized coefficients		Standardized coefficients		
		B	Std. error	Beta	t	Sig.
1	(Constant)	10,033	2,862		3,506	,001
	Treat	−3,502	1,867	−,269	−1,876	,067
	Psych	,033	,069	,093	,472	,639
	Soc	−,093	,078	−,237	−1,194	,238

[a]Dependent Variable: paf
[b]Weighted Least Squares Regression – Weighted by days

The above table shows the results. A largely similar pattern is observed, but treatment modality is no more statistically significant. We will now perform a Poisson regression which is probably more appropriate for rate data.

7 Poisson Regression

For analysis the module Generalized Linear Models is required. It consists of two submodules: Generalized Linear Models and Generalized Estimation Models. The first submodule covers many statistical models like gamma regression (Chap. 30), Tweedie regression (Chap. 31), Poisson regression (the current chapter and Chap. 47), and the analysis of paired outcomes with predictors (Chap. 3). The second submodule is for analyzing binary outcomes (Chap.42). For the current analysis the statistical model Poisson regression in the module Generalized Linear Models is required.

Command:

Analyze….Generalized Linear Models….mark: Custom….Distribution: Poisson …..Link function: Log….Response: Dependent variable: numbers of episodes of PAF….Scale Weight Variable: days of observation….Predictors: Main Effect: treatment modality….Covariates: psychological score, social score…. Model: main effects: treatment modality, psychological score, social score…. Estimation: mark Model-based Estimation….click OK.

Parameter estimates

Parameter	B	Std. error	95 % Wald confidence interval		Hypothesis test		
			Lower	Upper	Wald chi-square	df	Sig.
(Intercept)	1,868	,0206	1,828	1,909	8256,274	1	,000
[Treat = 0]	,667	,0153	,637	,697	1897,429	1	,000
[Treat = 1]	0[a]						
Psych	,006	,0006	,005	,008	120,966	1	,000
Soc	−,019	,0006	−,020	−,017	830,264	1	,000
(Scale)	1[b]						

Dependent Variable: paf
Model: (Intercept), treat, psych, soc
[a]Set to zero because this parameter is redundant
[b]Fixed at the displayed value

The above table gives the results. All of a sudden, all of the predictors including treatment modality, psychological and social score are very significant predictors of the paf rate.

8 Conclusion

Poisson regression is different from linear en logistic regression, because it uses a log transformed dependent variable. For rate analysis Poisson regression is very sensitive and probably better than standard regression methods. The methodology is explained.

9 Note

More background, theoretical and mathematical information about Poisson regression is given in Statistics applied to clinical studies 5th edition, Chap. 23, Springer Heidelberg Germany, 2012, from the same authors.