

Chapter 21

Generalized Linear Models Event-Rates (50 Patients)

General Purpose

To assess whether in a longitudinal study event rates, defined as numbers of events per person per period, can be analyzed with the generalized linear model module.

Specific Scientific Question

Can generalized linear modeling be trained to predict rates of episodes of paroxysmal atrial fibrillation both in groups and in individual future patients.

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.

This chapter was previously published in “Machine learning in medicine-cookbook 1” as Chap. 6, 2013.

Var 1	Var 2	Var 3	Var 4	Var 5
1	56,99	42,45	73	4
1	37,09	46,82	73	4
0	32,28	43,57	76	2
0	29,06	43,57	74	3
0	6,75	27,25	73	3
0	61,65	48,41	62	13
0	56,99	40,74	66	11
1	10,39	15,36	72	7
1	50,53	52,12	63	10
1	49,47	42,45	68	9
0	39,56	36,45	72	4
1	33,74	13,13	74	5

Var 1 =treatment modality

Var 2=psychological score

Var 3=social score

Var 4=days of observation

Var 5=number of episodes of paroxysmal atrial fibrillation (PAF)

The first 12 patients are shown only, the entire data file is entitled “generalizedlmeventrates” and is in extras.springer.com.

The Computer Teaches Itself to Make Predictions

SPSS 19.0 is used for training and outcome prediction. It uses XML (eXtended Markup Language) files to store data. We will perform the analysis with a linear regression analysis of variable 5 as outcome variable and the other 4 variables as predictors. Start by opening the data file.

Command:

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

Coefficients ^a						
Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	49,059	5,447		9,006	,000
	Treat	-2,914	1,385	-,204	-2,105	,041
	Psych	,014	,052	,036	,273	,786

(continued)

Coefficients ^a						
Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. Error	Beta		
	Soc	-,073	,058	-,169	-1,266	,212
	Days	-,557	,074	-,715	-7,535	,000

^aDependent variable: paf

The above table shows that treatment modality is weakly significant, and psychological and social scores 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.

Coefficients ^{a,b}						
Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. Error	Beta		
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

^aDependent variable: paf

^bWeighted least squares regression -Weighted by days

Command:

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

The above table shows the results. A largely similar pattern is observed, but treatment modality is no more statistically significant. We will use the generalized linear modeling module to perform a Poisson regression which is more appropriate for rate data. The model applied will also be stored and reapplied for making predictions about event rates in individual future patients.

Command:

Click Transform....click Random Number Generators....click Set Starting Point.... click Fixed Value (2000000)....click OK....click Generalized Linear Models click again 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 Estimationclick Save....mark Predicted value of mean of response....click Export....mark Export model in XML....click Browse.... in File name enter "exportrate"....in Look in: enter the appropriate map in your computer for storage....click Save....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

^aSet to zero because this parameter is redundant

^bFixed at the displayed value

The outcome sheets give 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. When minimizing the output sheets the data file returns and now shows a novel variable entitled “PredictedValues” with the mean rates of PAF episodes per patient (per day). The saved XML file will now be used to compute the predicted PAF rate in 5 novel patients with the following characteristics. For convenience the XML file is given in extras.springer.com.

Var 1	Var 2	Var 3	Var 4	Var 5
1,00	56,99	42,45	73,00	4,00
1,00	30,09	46,82	34,00	4,00
,00	32,28	32,00	76,00	2,00
,00	29,06	40,00	36,00	3,00
,00	6,75	27,25	73,00	3,00

Var 1 = treatment modality

Var 2 = psychological score

Var 3 = social score

Var 4 = days of observation

Var 5 = number of episodes of paroxysmal atrial fibrillation (PAF)

Enter the above data in a new SPSS data file.

Command:

Utilities....click Scoring Wizard....click Browse....click Select....Folder: enter the exportrate.xml file....click Select....in Scoring Wizard click Next....click Use value substitution....click Next....click Finish.

The above data file now gives individually predicted rates of PAF as computed by the linear model with the help of the XML file. Enter the above data in a new SPSS data file.

Var 1	Var 2	Var 3	Var 4	Var 5	Var 6
1,00	56,99	42,45	73,00	4,00	4,23
1,00	30,09	46,82	34,00	4,00	3,27
,00	32,28	32,00	76,00	2,00	8,54
,00	29,06	40,00	36,00	3,00	7,20
,00	6,75	27,25	73,00	3,00	7,92

- Var 1 =treatment modality
- Var 2=psychological score
- Var 3=social score
- Var 4=days of observation
- Var 5=number of episodes of paroxysmal atrial fibrillation (PAF)
- Var 6=individually predicted mean rates of PAF (per day)

Conclusion

The module generalized linear models can be readily trained to predict event rate of PAF episodes both in groups, and, with the help of an XML file, in individual patients.

Note

More background, theoretical and mathematical information of generalized linear modeling is available in SPSS for Starters part two, Chap. 10, entitled “Poisson regression”, pp 43–48, Springer Heidelberg Germany 2012, from the same authors.