

Chapter 20

Generalized Linear Models for Outcome Prediction with Paired Data (100 Patients and 139 Physicians)

General Purpose

With linear and logistic regression *unpaired* data can be used for outcome prediction. With generalized linear models *paired* data can be used for the purpose.

Specific Scientific Question

Can crossover studies (1) of sleeping pills and (2) of lifestyle treatments be used as training samples to predict hours of sleep and lifestyle treatment in groups and individuals.

Generalized Linear Modeling, the Computer Teaches Itself to Make Predictions

Var 1	Var 2	Var 3	Var 4
6,10	79,00	1,00	1,00
5,20	79,00	1,00	2,00
7,00	55,00	2,00	1,00
7,90	55,00	2,00	2,00
8,20	78,00	3,00	1,00

(continued)

This chapter was previously published in “Machine learning in medicine-cookbook 1” as Chap. 5, 2013.

Var 1	Var 2	Var 3	Var 4
3,90	78,00	3,00	2,00
7,60	53,00	4,00	1,00
4,70	53,00	4,00	2,00
6,50	85,00	5,00	1,00
5,30	85,00	5,00	2,00
8,40	85,00	6,00	1,00
5,40	85,00	6,00	2,00

Var 1 = outcome (hours of sleep after sleeping pill or placebo)

Var 2 = age

Var 3 = patientnumber (patientid)

Var 4 = treatment modality (1 sleeping pill, 2 placebo)

Only the data from first 6 patients are given, the entire data file is entitled “generalizedlm-pairedcontinuous” and is in extras.springer.com. SPSS 19.0 is used for analysis, with the help of an XML (eXtended Markup Language) file. Start by opening the data file.

Command:

Click Transform....click Random Number Generators....click Set Starting Pointclick Fixed Value (2000000)....click OK....click Analyze....Generalized Linear Models....again click Generalized Linear models....click Type of Model....click Linear....click Response....Dependent Variable: enter Outcome....Scale Weight Variable: enter patientid....click Predictors....Factors: enter treatment.... Covariates: enter age....click Model: Model: enter treatment and age....click Save: mark Predicted value of linear predictor....click Export....click Browse....File name: enter "exportpairedcontinuous"....click Save....click Continue....click OK.

Parameter estimates

Parameter	B	Std. Error	95 % Wald confidence interval		Hypothesis test		
			Lower	Upper	Wald Chi- Square	df	Sig.
(Intercept)	6,178	,5171	5,165	7,191	142,763	1	,000
[treatment= 1,00]	2,003	,2089	1,593	2,412	91,895	1	,000
[treatment=2,00]	0 ^a						
age	-,014	,0075	-,029	,001	3,418	1	,064
(Scale)	27,825 ^b	3,9351	21,089	36,713			

Dependent variable: outcome

Model: (Intercept), treatment, age

^aSet to zero because this parameter is redundant

^bMaximum likelihood estimate

The output sheets show that both treatment and age are significant predictors at $p < 0.10$. Returning to the data file we will observe that SPSS has computed predicted values of hours of sleep, and has given them in a novel variable entitled XBPredicted

(predicted values of linear predictor). The saved XML file (entitled "exportpairedcontinuous") will now be used to compute the predicted hours of sleep in five novel patients with the following characteristics. For convenience the XML file is given in extras.springer.com.

Var 2	Var 3	Var 4
79,00	1,00	1,00
55,00	2,00	1,00
78,00	3,00	1,00
53,00	4,00	2,00
85,00	5,00	1,00

Var 2 = age
 Var 3 = patientnumber (patientid)
 Var 4 = treatment modality (1 sleeping pill, 2 placebo)

Enter the above data in a new SPSS data file.

Command:

Utilities....click Scoring Wizard....click Browse....click Select....Folder: enter the exportpairedcontinuous.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 hours of sleep as computed by the linear model with the help of the XML file.

Var 2	Var 3	Var 4	Var 5
79,00	1,00	1,00	7,09
55,00	2,00	1,00	7,42
78,00	3,00	1,00	7,10
53,00	4,00	2,00	5,44
85,00	5,00	1,00	7,00

Var 2 = age
 Var 3 = patientnumber (patientid)
 Var 4 = treatment modality (1 sleeping pill, 2 placebo)
 Var 5 = predicted values of hours of sleep in individual patient

Conclusion

The module generalized linear models can be readily trained to predict hours of sleep in groups, and, with the help of an XML file, in individual future patients.

Generalized Estimation Equations, the Computer Teaches Itself to Make Predictions

Var 1	Var 2	Var 3	Var 4
.00	89,00	1,00	1,00
.00	89,00	1,00	2,00
.00	78,00	2,00	1,00
.00	78,00	2,00	2,00
.00	79,00	3,00	1,00
.00	79,00	3,00	2,00
.00	76,00	4,00	1,00
.00	76,00	4,00	2,00
.00	87,00	5,00	1,00
.00	87,00	5,00	2,00
.00	84,00	6,00	1,00
.00	84,00	6,00	2,00
.00	84,00	7,00	1,00
.00	84,00	7,00	2,00
.00	69,00	8,00	1,00
.00	69,00	8,00	2,00
.00	77,00	9,00	1,00
.00	77,00	9,00	2,00
.00	79,00	10,00	1,00
.00	79,00	10,00	2,00

Var 1 outcome (lifestyle advise given

0=no, 1=yes)

Var 2 physicians' age

Var 3 physicians' id

Var 4 prior postgraduate education regarding

lifestyle advise (1=no, 2=yes)

Only the first 10 physicians are given, the entire data file is entitled “generalized-pairedbinary” and is in extras.springer.com. All physicians are assessed twice, once before lifestyle education and once after. The effect of lifestyle education on the willingness to provide lifestyle advise was the main objective of the study.

SPSS 19.0 is used for analysis, with the help of an XML (eXtended Markup Language) file. Start by opening the data file.

Command:

Click Transform....click Random Number Generators....click Set Starting Pointclick Fixed Value (2000000)....click OK....click Analyze....Generalized Linear Models....Generalized Estimating Equations....click Repeated....in Subjects

variables enter physicianid....in Within-subject variables enter lifestyle advise....in Structure enter Unstructured....click Type of Model....mark Binary logistic....click Response....in Dependent Variable enter outcome....click Reference Category....mark First....click Continue....click Predictors....in Factors enter lifestyleadvise....in Covariates enter age....click Model....in Model enter lifestyle and age....click Save....mark Predicted value of mean of response....click Exportmark Export model in XML....click Browse.... In File name: enter "exportpairedbinary"....in Look in: enter the appropriate map in your computer for storage....click Save....click Continue....click OK.

Parameter estimates							
Parameter	B	Std. Error	95 % Wald confidence interval		Hypothesis test		
			Lower	Upper	Wald Chi- Square	df	Sig.
(Intercept)	2,469	,7936	,913	4,024	9,677	1	,002
[lifestyleadvise= 1,00]	-,522	,2026	-,919	-,124	6,624	1	,010
[lifestyleadvise= 2,00]	0*						
age	-,042	,0130	-,068	-,017	10,563	1	,001
(Scale)	1						

Dependent variable: outcome

Model: (Intercept), lifestyleadvise, age

*Set to zero because this parameter is redundant

The output sheets show that both prior lifestyle education and physicians’ age are very significant predictors at $p < 0.01$. Returning to the data file we will observe that SPSS has computed predicted probabilities of lifestyle advise given or not by each physician in the data file, and a novel variable is added to the data file for the purpose. It is given the name MeanPredicted. The saved XML file entitled “export-pairedbinary” will now be used to compute the predicted probability of receiving lifestyle advise based on physicians’ age and the physicians’ prior lifestyle education in twelve novel physicians. For convenience the XML file is given in extras.springer.com.

Var 2	Var 3	Var 4
64,00	1,00	2,00
64,00	2,00	1,00
65,00	3,00	1,00
65,00	3,00	2,00
52,00	4,00	1,00
66,00	5,00	1,00
79,00	6,00	1,00
79,00	6,00	2,00

53,00	7,00	1,00
53,00	7,00	2,00
55,00	8,00	1,00
46,00	9,00	1,00

Var 2 age
 Var 3 physicianid
 Var 4 lifestyleadvise (prior postgraduate education regarding lifestyle advise (1 = no, 2=yes))

Enter the above data in a new SPSS data file.

Command:

Utilities....click Scoring Wizard....click Browse....click Select....Folder: enter the exportpairedbinary.xml file....click Select....in Scoring Wizard click Next....mark Probability of Predicted Category....click Next....click Finish.

The above data file now gives individually predicted probabilities of receiving lifestyle advise as computed by the logistic model with the help of the XML file.

Var 2	Var 3	Var 4	Var 5
64,00	1,00	2,00	,56
64,00	2,00	1,00	,68
65,00	3,00	1,00	,69
65,00	3,00	2,00	,57
52,00	4,00	1,00	,56
66,00	5,00	1,00	,70
79,00	6,00	1,00	,80
79,00	6,00	2,00	,70
53,00	7,00	1,00	,57
53,00	7,00	2,00	,56
55,00	8,00	1,00	,59
46,00	9,00	1,00	,50

Var 2 age
 Var 3 physicianid
 Var 4 lifestyleadvise
 Var 5 probability of predicted category (between 0.0 and 1.0)

Conclusion

The module generalized estimating equations can be readily trained to predict with paired data the probability of physicians giving lifestyle advice as groups and, with the help of an XML file, as individual physicians.

Note

More background, theoretical and mathematical information of paired analysis of binary data is given in SPSS for starters part one, Chap. 13, entitled “Paired binary (McNemar test)”, pp 47–49, Springer Heidelberg Germany, 2010, from the same authors.