Chapter 42
Paired Binary Data with Predictor (139 General Practitioners)

1 General Purpose

Paired proportions have to be assessed when e.g. different diagnostic procedures are performed in one subject. McNemar's chi-square test is appropriate for analysis. Mc Nemar's test can not include predictor variables. The analysis of paired outcome proportions including predictor variables requires the module generalized estimating equations. The difference between the two outcomes and the independent effects of the predictors variables on the outcomes are simultaneously tested.

2 Schematic Overview of Type of Data File

Outcome-1 binary	outcome-2 binary	predictor
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.

3   Primary Scientific Questions

Fist, is the numbers of yes-responders of outcome-1 significantly different from that of outcome-2. Second, are the predictor variables significant predictors of the outcomes.

4   Data Example

In a study of 139 general practitioners the primary scientific question was: is there a significant difference between the numbers of practitioners who give lifestyle advise in the periods before and after (postgraduate) education. The second question was, is age an independent predictor of the outcomes.

Lifestyle advise-1	Lifestyle advise-2	Age (years)
,00	,00	89,00
,00	,00	78,00
,00	,00	79,00
,00	,00	76,00
,00	,00	87,00
,00	,00	84,00
,00	,00	84,00
,00	,00	69,00
,00	,00	77,00
,00	,00	79,00

0 = no, 1 = yes

The first ten patients of the data file is given above. We will use the data of the Chap. 41 once more. The entire data file is in extras.springer.com, and is entitled "chapter41paired binary".

5   2×2 Contingency Table of the Effect of Postgraduate Education

			Lifestyleadvise after education	
			No	Yes
			0	1
Lifestyleadvise	No	0	65	28
Before education	Yes	1	12	34

The above table summarizes the numbers of practitioners giving lifestyle advise in the periods prior to and after postgraduate education. Obviously, before education $65 + 28 = 93$ did not give lifestyle, while after education this number fell to 77. It looks as though the education was somewhat successful. According to the McNemar's test this effect was statistically significant (Chap. 41). In this chapter we will assess, if the effect still exists after adjustment for doctors' ages.

Start by opening the data file in SPSS. Prior to a generalized estimation equation analysis which includes additional predictors to a model with paired binary outcomes, the data will have to be restructured. For that purpose the Restructure Data Wizard will be used. The procedure is also applied in the Chap. 12.

6   Restructure Data Wizard

Command:
click Data....click Restructure....mark Restructure selected variables into cases....
   click Next....mark One (for example, w1, w2, and w3)....click Next....Name: id
   (the patient id variable is already provided)....Target Variable: enter
   "lifestyleadvise 1, lifestyleadvise 2 "....Fixed Variable(s): enter age....click
   Next.... How many index variables do you want to create?....mark One....click
   Next....click Next again....click Next again....click Finish....Sets from the origi-
   nal data will still be in use. . .click OK.

Return to the main screen and observe that there are now 278 rows instead of 139 in the data file. The first 10 rows are given underneath.

Id	Age	Index 1	Trans 1
1	89,00	1	,00
1	89,00	2	,00
2	78,00	1	,00
2	78,00	2	,00
3	79,00	1	,00
3	79,00	2	,00
4	76,00	1	,00
4	76,00	2	,00
5	87,00	1	,00
5	87,00	2	,00

id: patient identity number
age: age in years
Index 1: 1 = before postgraduate education, 2 = after postgraduate education
trans 1: lifestyleadvise no = 1, lifestyle advise yes = 2

The above data file is adequate to perform a generalized estimation equation analysis. Save the data file. For convenience of the readers it is given in extras. springer.com, and is entitled "chapter42pairedbinaryrestructured".

7   Generalized Estimation Equation Analysis

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 (Chaps. 21 and 47), and the analysis of paired outcomes with predictors (Chap. 3). The second is for analyzing binary outcomes (current chapter).

Command:

Analyze....Generalized Linear Models....Generalized Estimation Equations....click Repeated....transfer id to Subject variables....transfer Index 1 to Within-subject variables....in Structure enter Unstructured....click Type of Model....mark Binary logistic....click Response....in Dependent Variable enter lifestyleadvise....click Reference Category....click Predictors....in Factors enter Index 1....in Covariates enter age....click Model....in Model enter lifestyleadvise and age....click OK.

Tests of model effects

	Type III		
Source	Wald chi-square	df	Sig.
(Intercept)	8,079	1	,004
Index1	6,585	1	,010
age	10,743	1	,001

Dependent Variable: lifestyleadvise before
Model: (Intercept), Index1, age

Parameter estimates

Parameter	B	Stri. Error	95% Wald confidence interval		Hypothesis test		
			Lower	Upper	Wald chi-square	df	Sig.
(Intercept)	−2,508	,8017	−4,079	−,936	9,783	1	,002
[Indexl = 1]	,522	,2036	,123	,921	6,585	1	,010
[Indexl = 2]	0[a]						
Age	,043	,0131	,017	,069	10,743	1	,001
(Scale)	1						

Dependent Variable: lifestyleadvise before
Model: (Intercept), Index1, age
[a]Set to zero because this parameter is redundant

In the output sheets the above tables are observed. They show that both the index 1 (postgraduate education) and age are significant predictors of lifestyleadvise. The interpretations of the two significant effects are slightly different from one another. The effect of postgraduate education is compared with no postgraduate education at all, while the effect of age is an independent effect of age on lifestyleadvise, the

older the doctors the better lifestyle advise given irrespective of the effect of the postgraduate education.

8 Conclusion

Paired proportions have to be assessed when e.g. different diagnostic procedures are performed in one subject. McNemar's chi-square test is appropriate for analysis. Mc Nemar's test can not include predictor variables, and is not feasible for more than two outcomes. For that purpose Cochran's tests are required (Chap. 43). The analysis of paired outcome proportions including predictor variables requires the module generalized estimating equations as reviewed in the current chapter.

9 Note

More background, theoretical and mathematical information of paired binary outcomes are given in Statistics applied to clinical studies 5th edition, Chap. 3, Springer Heidelberg Germany, 2012, from the same authors. More information of generalized linear models for paired outcome data is given in Machine learning in medicine a complete overview, Chap. 20, Springer Heidelberg Germany, 2015, from the same authors.