Chapter 16
Multistage Regression (35 Patients)

1 General Purpose

The multistage regression assumes that an independent variable (x-variable) is problematic, meaning that it is somewhat uncertain. An additional variable can be argued to provide relevant information about the problematic variable, and is, therefore, called instrumental variable, and included in the analysis.

2 Schematic Overview of Type of Data

Outcome	problematic predictor	instrumental predictor
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.

3 Primary Scientific Question

Is multistage regression better for analyzing outcome studies with multiple predictors than multiple linear regression.

T.J. Cleophas, A.H. Zwinderman, *SPSS for Starters and 2nd Levelers*,
DOI 10.1007/978-3-319-20600-4_16

4 Data Example

The effects of counseling frequencies and non-compliance (pills not used) on the efficacy of a novel laxative drug is studied in 35 patients. The first 10 patients of the data file is given below.

Pat no	Efficacy of new laxative (stools/month)	Pills not used (n)	Counseling (n)
1	24	25	8
2	30	30	13
3	25	25	15
4	35	31	14
5	39	36	9
6	30	33	10
7	27	22	8
8	14	18	5
9	39	14	13
10	42	30	15

The entire data file is in extras.springer.com, and is entitled "chapter16multistageregression". Start by opening the data file in SPSS. We will first perform a multiple regression, and then a multistep regression.

5 Traditional Multiple Linear Regression

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

Command:
Analyze....Regression....Linear....Dependent: ther eff....Independent(s): counseling, non-compliance....click OK.

Coefficients[a]

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	2,270	4,823		,471	,641
	Counseling	1,876	,290	,721	6,469	,000
	Non-compliance	,285	,167	,190	1,705	,098

[a]Dependent Variable: ther eff

The above table shows the results of a linear regression assessing (1) the effects of counseling and non-compliance on therapeutic efficacy.

Command:
Analyze....Regression....Linear....Dependent: counseling. . .Independent(s): non-compliance....click OK.

Coefficients[a]

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	4,228	2,800		1,510	,141
	Non-compliance	,220	,093	,382	2,373	,024

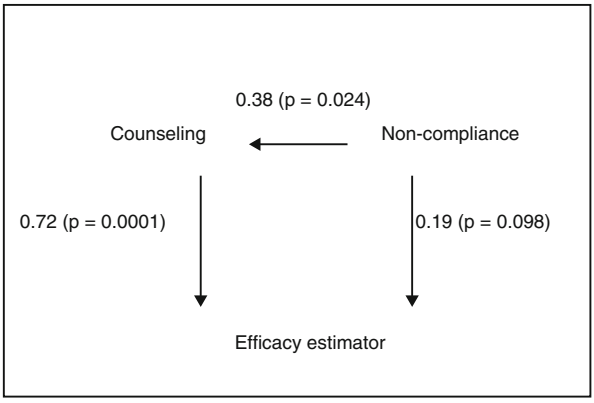
[a]Dependent Variable: counseling

The above table give the effect of non-compliance on counseling.

With p = 0,10 as cut-off p-value for statistical significance all the effects above are statistically significant. Non-compliance is a significant predictor of counseling, and at the same time a significant predictor of therapeutic efficacy at p = 0,024. This would mean that non-compliance works two ways: it predicts therapeutic efficacy *directly* and *indirectly* through counseling. However, the indirect way is not taken into account in the usual one step linear regression. An adequate approach for assessing both ways simultaneously is path statistics.

6 Multistage Regression

Multistage regression, otherwise called path analysis or path statistics, uses add-up sums of regression coefficients for better estimation of multiple step relationships. Because regression coefficients have the same unit as their variable, they cannot be added up unless they are standardized by dividing them by their own variances. SPSS routinely provides the standardized regression coefficients, otherwise called path statistics, in its regression tables as shown above. The underneath figure gives a path diagram of the data.



The standardized regression coefficients are added to the arrows. Single path analysis gives a standardized regression coefficient of 0.19. This underestimates the real effect of non-compliance. Two step path analysis is more realistic and shows that the add-up path statistic is larger and equals

$$0.19 + 0.38 \times 0.72 = 0.46$$

The two-path statistic of 0.46 is a lot better than the single path statistic of 0.19 with an increase of 60 %.

7 Alternative Analysis: Two Stage Least Square (2LS) Method

Instead of path analysis the two stage least square (2LS) method is possible and is available in SPSS. It works as follows. First, a simple regression analysis with counseling as outcome and non-compliance as predictor is performed. Then the outcome values of the regression equation are used as predictor of therapeutic efficacy. For analysis the statistical model 2 Stage Least Squares in the module Regression is required.

Command:
Analyze. . ..Regression. . ..2 Stage Least Squares. . ..Dependent: stool. . .. Explanatory: non-compliance. . ..Instrumental:counseling . . ..mark: include constant in equation....click OK.

Model description

		Type of variable
Equation 1	Stool	Dependent
	Noncompliance	Predictor
	Counseling	Instrumental

MOD_3

ANOVA

		Sum of squares	df	Mean square	F	Sig.
Equation 1	Regression	1408,040	1	1408,040	4,429	,043
	Residual	10490,322	33	317,889		
	Total	11898,362	34			

Coefficients

		Unstandardized coefficients		Beta	t	Sig.
		B	Std. error			
Equation 1	(Constant)	−49,778	37,634		−1,323	,195
	Noncompliance	2,675	1,271	1,753	2,105	,043

The above tables show the results of the 2LS method. As expected the final p-value of the effect of non-compliance on stool is smaller than that of the traditional linear regression with p-values of 0,043 instead 0,098.

8 Conclusion

Multistage regression methods often produce better estimations of multi-step relationships than standard linear regression methods do. Examples are given.

9 Note

More background, theoretical and mathematical information of multistep regression is given in Statistics applied to clinical studies 5th edition, Chap. 20, Springer Heidelberg Germany, 2012, from the same authors.