

Chapter 2

Multistage Regression (35 Patients)

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

Example

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

	Var 1	Var 2	Var 3 (var = variable)
	Outcome	Pt instrumental variable	Problematic predictor
	Frequency counseling	Pills not used	Efficacy estimator of new laxative (stools/month)
1.	24	8	25
2.	30	13	30
3.	25	15	25
4.	35	14	31
5.	39	9	36
6.	30	10	33
7.	27	8	22
8.	14	5	18
9.	39	13	14
10.	42	15	30
11.	41	11	36
12.	38	11	30
13.	39	12	27
14.	37	10	38
15.	47	15	40

(continued)

(continued)

	Var 1	Var 2	Var 3 (var = variable)
16.	30	13	31
17.	36	12	25
18.	12	4	24
19.	26	10	27
20.	20	8	20
21.	43	16	35
22.	31	15	29
23.	40	14	32
24.	31	7	30
25.	36	12	40
26.	21	6	31
27.	44	19	41
28.	11	5	26
29.	27	8	24
30.	24	9	30
31.	40	15	20
32.	32	7	31
33.	10	6	29
34.	37	14	43
35.	19	7	30

Coefficients^a

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.
	B	Std. error	Beta		
1 (Constant)	2.270	4.823		0.471	0.641
Counseling	1.876	0.290	0.721	6.469	0.000
Non-compliance	0.285	0.167	0.190	1.705	0.098

^a Dependent Variable: ther eff**Coefficients^a**

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.
	B	Std. error	Beta		
1 (Constant)	4.228	2.800		1.510	0.141
Non-compliance	0.220	0.093	0.382	2.373	0.024

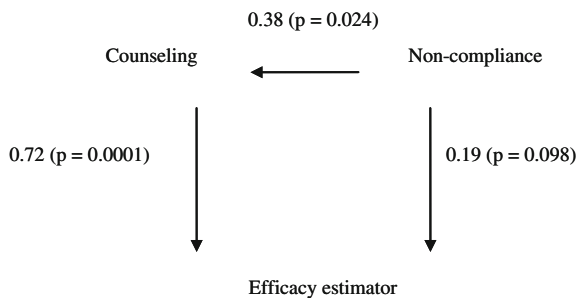
^a Dependent Variable: counseling

The above tables (commands as given in [Chap. 5](#) of the first part of this title) show the results of two linear regressions assessing (1) the effects of counseling and non-compliance on therapeutic efficacy, and (2) the effect of

non-compliance on counseling. With $p = 0.10$ as cut-off p-value for statistical significance all of the effects are statistically significant. Non-compliance is a significant predictor of counseling, and at the same time a significant predictor of therapeutic efficacy. 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.

Path Statistics

Path analysis 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 can not 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 %.

Two Stage Least Square 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.

Command: Analyze....Regression....2 Stage Least Squares....Dependent: therapeutic efficacy....Explanatory: non-compliance....Instrumental: counselingOK.

Model description			Type of variable
Equation 1	VAR00001		dependent
	VAR00003		predictor
	VAR00002		instrumental

MOD_1

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.
	B	Std. error	Beta		
Equation 1 (Constant)	-61.095	37.210		-1.642	0.110
VAR00003	3.113	1.256	2.078	2.478	0.019

The above tables show the results of the 2LS method. As expected the final p-value is smaller than the simple linear regression p-value of the effect of non-compliance on therapeutic efficacy.

Conclusion

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

SPSS for Starters, Part 2

Cleophas, T.J.; Zwinderman, A.H.

2012, IX, 104 p. 20 illus., Softcover

ISBN: 978-94-007-4803-3