Path analysis using regression

A simple medical example of an input path model

Data for the 20 cases in our first demonstration are shown in Table 7.1. As part of a larger study of variables implicated in susceptibility to suicidal tendencies, a path model linking family history of mental health problems (FAMILYHIST), positive mental health (POSMENTAL) and susceptibility to depression following adverse life events (SUSCEPTIBILITY) with suicidal thoughts (THOUGHTS) is tested using a series of regression analyses in SPSS. Twenty adults between the ages of 25 and 45 are recruited and they complete self report inventories designed to yield scores on the four variables.

Table 7.1
Fabricated data for a path analysis (med.path1.sav)

case	familyhist	posmental	susceptibility	thoughts
1	48	8	8	4
2	87	4	4	4
3	52	9	8	3
4	37	11	4	4
5	39	8	6	4
6	46	6	10	6
7	54	8	5	5
8	36	7	12	5
9	37	8	8	6
10	58	9	4	3
11	65	6	8	3
12	77	7	4	2
13	51	10	6	3
14	20	12	9	5
15	59	6	8	4
16	63	6	8	4
17	55	7	6	5
18	42	10	6	5
19	63	7	8	4
20	68	8	6	3

An input path diagram representing a proposed causal model involving the four variables in Table 7.1 is shown in Figure 7.2.

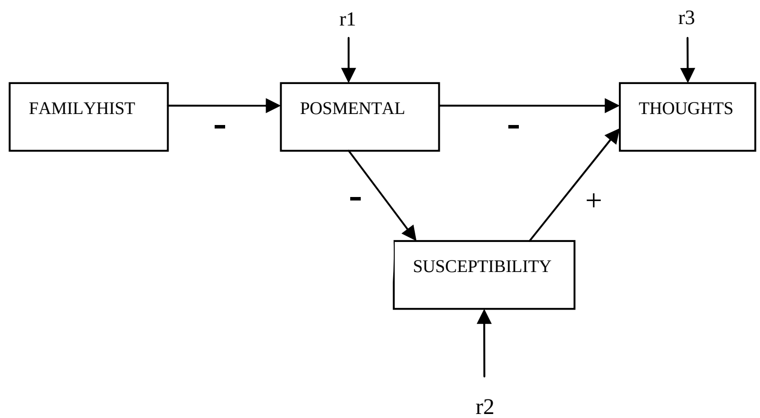

Figure 7.2. Input path diagram representing a proposed causal model

The causal model in Figure 7.2 proposes that a family history of mental problems decreases the likelihood of positive mental health (a negative effect: high on FAMILYHIST → low on POSMENTAL). It is also proposed that good mental health (POSMENTAL) results in fewer and less serious suicidal thoughts (THOUGHTS) and lower SUSCEPTIBILITY to depression following adverse life events (both negative effects). Finally, it is proposed that high SUSCEPTIBILITY to depression following adverse life events results in more suicidal THOUGHTS (a positive effect). There are 10 data points and 7 parameters to be estimated, so the model is under-identified, with 10-7-1 = 2 *df*s.

Path analysis: requesting the regression analyses in SPSS

So, we need to conduct a series of regression analyses. They can be specified as follows:

THOUGHTS is the DV, regressed on POSMENTAL and SUSCEPTIBILITY as IVs

SUSCEPTIBILITY is the DV, regressed on POSMENTAL as the sole IV

POSMENTAL is the DV, regressed on FAMILYHIST as the sole IV

First, enter the data for the four variables into an SPSS datasheet (20 cases in 4 columns). For the first regression, select **Analyze** from the menu bar, then **Regression**, then **Linear**, to get a dialog box like SPSS Dialog Box 4.1 shown in the chapter on Regression. Move THOUGHTS into the **Dependent** box and POSMENTAL and SUSCEPTIBILITY into the **Independent(s)** box. Check that **Enter**, the default **Method**, is selected. Click the **Statistics** button and select **Descriptives** in order to get the correlations. Click **Continue** and **OK** to get the analysis.

Path analysis: understanding the output

The relevant output is in the tables labelled Correlations, Model Summary and Coefficients. These tables are shown in SPSS Output 7.1. The first table in the output (not reproduced here) gives descriptive statistics such as the means and standard deviations. Next comes the Correlations table, and we see that the correlations that our model specified as causal effects are all statistically significant. This is encouraging. The next table (not reproduced here) is Variables Entered/Removed, and this just tells us that POSMENTAL and SUSCEPTIBILITY were entered and that THOUGHTS was the DV. Next is the Model Summary, in which we see that *R* Square = 0.37. Then comes an ANOVA table (not reproduced here) that just confirms that the regression equation is significant. Finally, we have the coefficients table, in which we see that the standardized beta coefficients are -0.45 for POSMENTAL to THOUGHTS, which is just significant (p = 0.05) and 0.25 for SUSCEPTIBILITY to THOUGHTS, which is not significant (p = 0.25).

Correlations

		thoughts	posmental	susceptibility
Pearson Correlation	thoughts	1.000	-.567	.457
	posmental	-.567	1.000	-.448
	susceptibility	.457	-.448	1.000
Sig. (1-tailed)	thoughts	.	.005	.022
	posmental	.005	.	.024
	susceptibility	.022	.024	.
N	thoughts	20	20	20
	posmental	20	20	20
	susceptibility	20	20	20

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.611[a]	.373	.300	.896

a. Predictors: (Constant), susceptibility, posmental

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.881	1.298		3.761	.002
	posmental	-.031	.015	-.454	-2.114	.050
	susceptibility	.125	.106	.253	1.180	.254

a. Dependent Variable: thoughts

SPSS Output 7.1. Output from the first regression analysis for path analysis of the model in Figure 7.2

The remaining two regression analyses required by the path model are carried out in the same way, using regression dialog boxes. SUSCEPTIBILITY is entered as the DV and POSMENTAL as the sole IV for the second regression, and POSMENTAL is entered as the DV and FAMILYHIST as the sole IV for the third regression. We will not reproduce the output tables again, but the relevant information from them is as follows: the correlations specified in the model are both significant (POSMENTAL/SUSCEPTIBILITY = -0.45, $p < 0.05$; FAMILYHIST/POSMENTAL = -0.70, $p < 0.0001$), the IV in each regression was entered and the ANOVAs confirmed that both were significant. The *R* Square value for the regression with SUSCEPTIBILITY as the DV is 0.20 and that for the regression with POSMENTAL as the DV is 0.49. The standardized beta coefficients are -0.45 for POSMENTAL to SUSCEPTIBILITY and -0.70 for FAMILYHIST to POSMENTAL (both significant, $p < 0.05$). The partial regression coefficients are displayed in our *output* *path diagram* in Figure 7.3, where we also display the r1, r2, and r3 values of R^2 (proportion of variance accounted for). Sometimes, instead of the R^2 values, the values entered in an output path diagram are $\sqrt{1-R^2}$ (the *residuals*) or $1 - R^2$ (proportion of variance not accounted for).

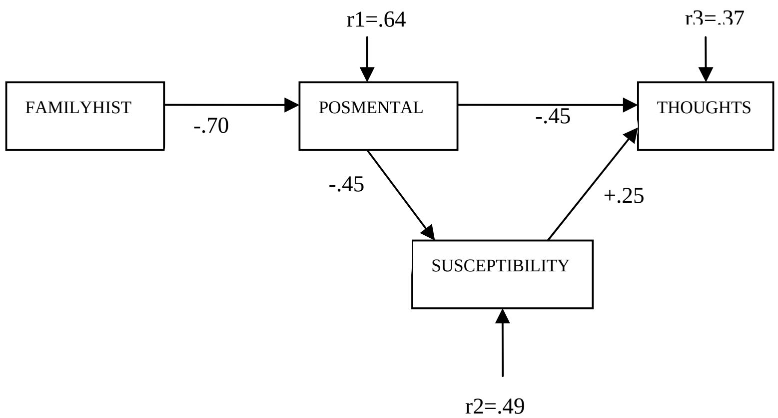

Figure 7.3. Output path diagram for data in Table 7.1

Goodness of fit

We need some further information to tell us how well the data fit our proposed models. A problem with using SPSS Regression to do the analysis is that no estimate of fit is provided. We note that one beta value, for the path from SUSCEPTIBILITY to THOUGHTS, was not significant (beta = 0.25, $p = 0.25$, from SPSS Output 7.1), which is not encouraging. However, we will delay discussion of goodness of fit indicators until we analyse some real data using AMOS, a relatively new program marketed by SPSS that has impressed us with its excellent graphic interface and ease of use. Then, we will be able to discuss the goodness of fit indices computed by the program. In the meantime, we just note that the value of Chi Square (obtained from AMOS), which should be non-significant if the model is a good fit, is highly significant; $\chi^2 = 22.24$, with 2 *df*s, given by number of data points (10) minus number of parameters estimated (7) minus the constant (1).

For the moment, we just note that, as our model turned out to be a bad fit, we might choose to modify it by adding additional variables, by adding more paths between the existing variables or rethinking the directions of influences within our model. For example, a case might be made for expecting SUSCEPTIBILITY to cause changes in POSMENTAL, rather than the reverse. Of course, it would have been better if we had generated some alternative models at the outset, so that we could compare their fit with our preferred model now. Incidentally, Chi Square tends to reach significance rather readily as sample size increases, so it is not generally the best indicator of fit when the sample size is high. It could still be useful though, even if it were significant for all of the proposed models, because it would tell us something if its value was much smaller for one model than for another. If we make changes to our model at this stage, we really need new data to test the new models, though it would still be alright to explore new models with our existing data in order to get an idea of whether it would be worth collecting new data to carry out valid tests.

Direct and indirect effects

Before leaving this example, which we analysed with regression analyses in order to reveal the logic that underlies the computations in dedicated packages, we note that sometimes researchers want to know the overall impact of one variable on another; the total of direct and indirect effects. For example, we might want to know the overall effect of POSMENTAL on THOUGHTS. To answer that question we need to take the direct effect of POSMENTAL (-0.45) and add to it the indirect effect via SUSCEPTIBILITY. Indirect effects are obtained by multiplying the effects along each indirect path. In this example, there is only one indirect path from POSMENTAL to THOUGHTS and the coefficients along that path are -0.45 and 0.25, so the indirect effect of POSMENTAL on THOUGHTS is -0.45 x 0.25 = -0.113. So the total effect of POSMENTAL on THOUGHTS is (-0.45) + (-0.113) = -0.56.