Partial correlation

An alternative medical example suitable for partial correlation analysis

Family history, measured as a binary variable, is a predictor for depression (DEPRESSION). It may be, however, that a family history of depression is related to a familial tendency towards anxiety, but there is a distinction here between anxiety as a relatively stable personality trait (trait anxiety) and anxiety as a temporary state (state anxiety) in response to stressful internal or external events. The balance between state and trait anxiety, measured as state anxiety minus trait anxiety, may provide a better predictor of depression than the existence of a family history of depression. We also know that people in employment are less prone to depression than people who are unemployed. A partial correlation analysis is carried out to clarify the relationship between family history of depression (FAMILYHIST) and susceptibility to depression, controlling for anxiety (state-trait) pattern (ANXIETY) and employment (EMPLOY). The purpose of this analysis is to determine whether the family history – depression relationship can be construed as direct or indirect (i.e., mediated by anxiety (state-trait) pattern).

Partial correlation: the (fabricated) data

A depression scale and state/trait anxiety scales are administered to eighty adults, and their status on the two binary variables (family history of depression and employment) is recorded. A power analysis using SPSS SamplePower indicated that a sample size of $N = 80$ would be sufficient to provide power = 0.8, assuming that $r = 0.3$ was the smallest correlation we wanted to avoid missing with $\alpha = 0.05$, two-tailed. The first few rows of data are in Table 6.1 (the full dataset can be found as med.partial.sav on

the book website). The data should be entered in an SPSS worksheet in four columns, just as in Table 6.1.

Table 6.1
The first few rows of data on family history of depression, depression, anxiety, and employment status for 80 adults (the full dataset can be found as med.partial.sav on the website)

familyhist[1]	depression	anxiety	employ[2]
1	2.06	-0.15	1
2	3.65	-1.38	1
2	2.06	-0.35	1
2	1.24	-0.50	1
2	2.29	-0.85	1
1	1.24	-0.55	2

[1] familyhist: 1 = no history of depression, 2 = history of depression
[2] employ: 1 = unemployed, 2 = employed

Partial correlation: the relationships tested

We carried out a partial correlation analysis to clarify the relationship between FAMILYHIST and DEPRESSION, controlling for ANXIETY and EMPLOY. This analysis was carried out to determine whether the family history of depression - depression relationship could be construed as direct or indirect (i.e., mediated by anxiety (state - trait) and/or employ). Additional partial correlation analyses, controlling for ANXIETY and EMPLOY one at a time, could also be carried out to further clarify the relationship, but these analyses are not reported here.

Partial correlation: requesting the analysis in SPSS

First, it is a good idea to look at the bivariate (zero-order) correlations among the variables. To do this, choose **Analyze**, then **Correlate**, then **Bivariate**. Select the four variables and use the arrow to put them in the **Variables** box. Accept all of the defaults and click **OK** to get a Pearson correlation matrix with two-tailed *p* levels, with significant values indicated by asterisks.

To do the partial correlation analysis, choose **Analyze**, then **Correlate**, then **Partial**. Use the arrows to move FAMILYHIST and DEPRESSION into the **Variables** box and ANXIETY and EMPLOY into the **Controlling for** box. Accept the defaults and click **OK**.

Partial correlation analysis: understanding the output

The bivariate correlation matrix, which is not reproduced here, confirms that correlations between all pairs of variables, with the exception of that between ANXIETY and EMPLOY, were significant at $p < 0.05$ or better. That suggests that it makes sense to proceed with the partial analysis. The result of the partial correlation analysis is shown in SPSS Output 6.1.

Correlations

Control Variables			familyhist	depression
anxiety & employ	familyhist	Correlation	1.000	.065
		Significance (2-tailed)	.	.572
		df	0	76
	depression	Correlation	.065	1.000
		Significance (2-tailed)	.572	.
		df	76	0

SPSS Output 6.1. Partial correlation between depression and family history of depression with anxiety (state – trait) and employment status controlled

The bivariate correlation between FAMILYHIST and DEPRESSION was $r(df=78) = 0.31$, $p < 0.01$, two-tailed. This indicates that people with a family history of depression had higher depression scores. However, you can see in SPSS Output 6.1 that, when ANXIETY and EMPLOY were controlled, the correlation dropped to $r(df=76) = 0.06$, $p = 0.57$, two-tailed. We can infer that the relationship between family history of depression and depression is mediated by ANXIETY (state - trait) and/or EMPLOY. We have not reproduced the analyses here, but we can tell you that partialling out ANXIETY (state - trait) and EMPLOY one at a time suggested that ANXIETY (state - trait) was the principle mediator.

This result leads to the question of whether ANXIETY (state - trait) is directly related to depression when FAMILYHIST and EMPLOY are controlled. In passing, we suggest that FAMILYHIST is an antecedent variable (prior to ANXIETY) and EMPLOY is probably an intermediate variable (i.e., established subsequent to FAMILYHIST and prior to DEPRESSION, if we assume that ANXIETY (state - trait) influences EMPLOY rather than the reverse). You may note here that it makes no difference to the analysis whether a covariate is antecedent or intermediate, but it may make a difference to your interpretation of the result of the analysis. The procedure for this partial correlation analysis is the same as before, except that ANXIETY and DEPRESSION go in the **Variables** box and FAMILYHIST and EMPLOY go in the **Controlling for** box.

The result of this analysis is shown in SPSS Output 6.2. The bivariate correlation between ANXIETY and DEPRESSION was $r(df=78) = -0.45$, $p < 0.001$, two-tailed. Recall that ANXIETY was computed as 'state' minus 'trait' scores, so that high scores mean higher state anxiety, so the negative correlation tells us that people with higher trait than state anxiety scores have higher DEPRESSION SCORES. You can see in SPSS Output 6.2 that the relationship remained strong ($r(df=76) = -0.41$, $p < 0.001$, two-tailed), when FAMILYHIST and EMPLOY were controlled. We conclude that a higher trait than state anxiety score predicts a higher depression score regardless of family history of depression and employment status.

Correlations

Control Variables			depression	anxiety
employ & familyhist	depression	Correlation	1.000	-.406
		Significance (2-tailed)	.	.000
		df	0	76
	anxiety	Correlation	-.406	1.000
		Significance (2-tailed)	.000	.
		df	76	0

SPSS Output 6.2. Partial correlation between ANXIETY and DEPRESSION with FAMILYHIST and EMPLOY controlled

Semipartial (or part) correlations

Recall that, in partial correlation analysis, common variance is removed from both the IV and DV. It is possible, however, to remove the contribution of the antecedent or intermediate variable from only the IV or only the DV. When this is done, the correlation between the IV and DV, with the common variance of the other variable removed from just one of them, is called the *semipartial* (or *part*) correlation. It has been suggested (e.g., Diekhoff, 1992, p. 266) that a semipartial correlation may be the correlation of choice when it seems highly likely that the covariate that you are interested in may influence only one of the two principal variables (IV or DV). An example might be a study looking at the relationship between age (the IV) and income (the DV), where we may want to take account of amount of education (a covariate), but this variable would not be expected to influence age. Therefore, according to the criterion expressed by Diekhoff, it may be reasonable to remove its effect only from the DV (income). However, we are not clear that this is necessarily the best course of action. Although amount of education cannot influence age, it is plausible that age will influence education. In other words, amount of education should be construed as an intermediate (mediating) variable rather than an antecedent variable. That does not seem to us a compelling reason for using a semipartial correlation. If it were, it would apply to all cases of intermediate covariates. Our view is that a partial correlation

should be the default and, when the decision is to use a semipartial correlation, a clear justification should be provided.

In our first partial correlation analysis, neither of the covariates (ANXIETY (state - trait) nor EMPLOYment status) could influence family history of depression, so should we have used a semipartial analysis? We do not think so, but perhaps it would be instructive to re-do that analysis using semipartial correlations. SPSS does not offer semipartial correlations in the **Correlate** procedure, but they can be requested in the **Regression** procedure. Incidentally, this is a good illustration of the interrelatedness of these analysis methods (R^2 change when a covariate is added is the square of the semipartial correlation). Furthermore, looking ahead to Chapter 7, it will become apparent that the flow diagrams in Figure 6.2 could be the start point for a path analysis.

Semipartial correlation: requesting the analysis in SPSS

To get the semipartial correlation of FAMILYHIST with DEPRESSION, with ANXIETY and EMPLOY controlled, choose **Analyze**, then **Regression**, then **Linear**. Select DEPRESSION and move it into the **Dependent box**, then select FAMILYHIST and move it into the **Independent(s)** box and click **Next** to obtain **Block 2 of 2**. Then select the IVs to be controlled (ANXIETY and EMPLOY) and move them into the **Independent(s)** box. The dialog box will now look like SPSS Dialog Box 4.6 in the chapter on regression, except that different variables will be displayed and **Block 2 of 2** will be displayed instead of **Block 4 of 4**. Click the **Statistics** button at the bottom and select **Part and Partial Correlations** (the statistics dialog box can be seen in SPSS Dialog Box 4.2). Click **Continue** and **OK** to get the analysis.

Semipartial correlation analysis: understanding the output

We have effectively done a hierarchical regression analysis with FAMILYHIST entered first, followed by the IVs to be controlled. The Coefficients output table (shown in SPSS Output 6.3) contains the partial and semipartial (part) correlations. You can see that the partial correlation between FAMILYHIST and DEPRESSION (.065) is as we found using the partial correlation procedure in SPSS (see SPSS Output 6.2), and the semipartial (part) correlation between these two variables is 0.055. So, in this case it makes very little difference whether we compute a partial or semipartial correlation.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	1.101	.335		3.289	.002			
	familyhist	.541	.188	.310	2.879	.005	.310	.310	.310
2	(Constant)	2.421	.455		5.325	.000			
	familyhist	.110	.193	.063	.567	.572	.310	.065	.055
	anxiety	-.330	.085	-.415	-3.868	.000	-.452	-.406	-.373
	employ	-.538	.195	-.276	-2.758	.007	-.307	-.302	-.266

a. Dependent Variable: depression

SPSS Output 6.3. Partial and semipartial (part) correlation between FAMILYHIST and DEPRESSION