Chapter 1:
Path Models in Factor, Path, and Structural Equation Analysis

Scientists dealing with behavior, especially those who observe it occurring in its natural settings, rarely have the luxury of the simple bivariate experiment, in which a single independent variable is manipulated and the consequences observed for a single dependent variable. Even those scientists who think they do are often mistaken: The variables they directly manipulate and observe are typically not the ones of real theoretical interest but are merely some convenient variables acting as proxies or indexes for them. A full experimental analysis would again turn out to be multivariate, with a number of alternative experimental manipulations on the one side, and a number of alternative response measures on the other.

Over many years, numerous statistical techniques have been developed for dealing with situations in which multiple variables, some unobserved, are involved. Such techniques often involve large amounts of computation. Until the advent of powerful digital computers and associated software, the use of these methods tended to be restricted to the dedicated few. But in the last few decades it has been feasible for any interested behavioral scientists to take a multivariate approach to their data. Many have done so. The explosive growth in the use of computer software packages such as SPSS, SAS, and R is one evidence of this.

The common features of the methods discussed in this book are that (a) multiple variables—three or more—are involved, and that (b) one or more of these variables is unobserved, or latent. Neither of these criteria provides a decisive boundary. Bivariate methods may often be regarded as special cases of multivariate methods. Some of the methods we discuss can be—and often are—applied in situations where all the variables are, in fact, observed. Nevertheless, the main focus of our interest is on what we call, following Bentler (1980), *latent variable analysis*, a term encompassing such specific methods as factor analysis, path analysis, and structural equation modeling (SEM), all of which share these defining features.

Path Diagrams

An easy and convenient representation of the relationships among a number of variables is the *path diagram*. In such a diagram we use capital letters, A, B, X, Y, and so on, to represent variables. The connections among variables are represented in path diagrams by two kinds of arrows: a straight, one-headed arrow represents a causal relationship between two variables, and a curved two-headed arrow represents a simple correlation between them.

Fig. 1.1 shows an example of a path diagram. Variables A, B, and X all are assumed to have causal effects on variable C. Variables A and B are assumed to be correlated with each other. Variable X is assumed to affect C but to be uncorrelated with either A or B. Variable C might (for example) represent young children's intelligence. Variables A and B could represent father's and mother's intelligence, assumed to have a causal influence on their child's intelligence. (The diagram is silent as to whether this influence is environmental, genetic, or both.) The curved arrow between A and B allows for the likely possibility that father's and mother's intelligence will be correlated. Arrow X represents the fact that there are other variables, independent of mother's and father's intelligence, that can affect a child's intelligence.

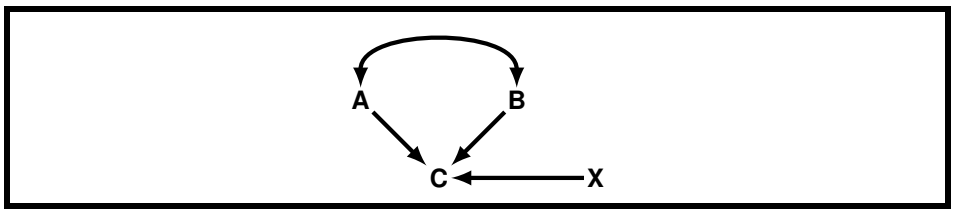


Figure 1.1 Example of a simple path diagram.

Fig. 1.2 shows another example of a path diagram. T is assumed to affect both A and B, and each of the latter variables is also affected by an additional variable; these are labeled U and V, respectively. This path diagram could represent the reliability of a test, as described in classical psychometric test theory. A and B would stand (say) for scores on two alternate forms of a test. T would represent the unobserved true score on the trait being measured, which is assumed to affect the observed scores on both forms of the test. U and V would represent factors specific to each form of the test or to the occasions on which it was administered, which would affect any given

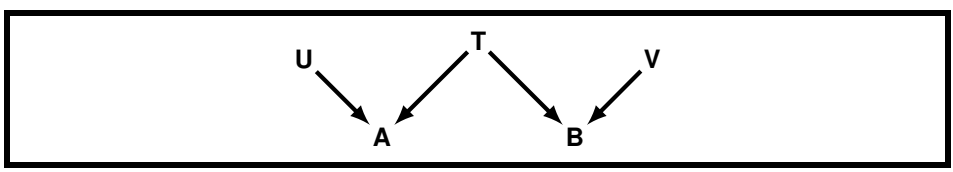


Figure 1.2 Another path diagram: test reliability.

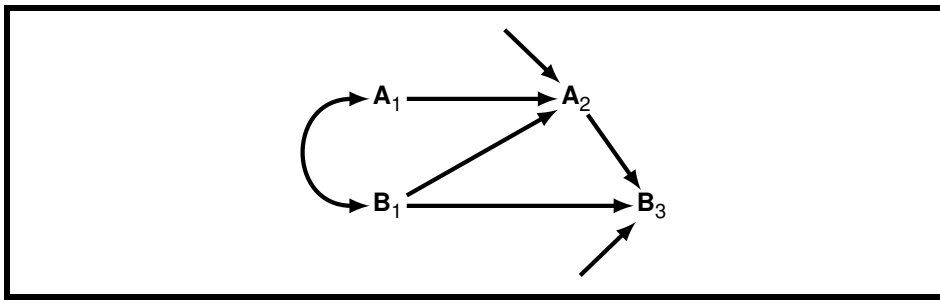


Figure 1.3 A path diagram involving events over time.

performance but be unrelated to the true trait. (In classical psychometric test theory, the variance in A and B resulting from the influence of T is called *true score variance*, and that caused by U or V is called *error variance*. The proportion of the variance of A or B due to T is called the *reliability* of the test.)

Fig. 1.3 shows a path representation of events over time. In this case, the capital letters A and B are used to designate two variables, with subscripts to identify the occasions on which they are measured: Both A and B are measured at time 1, A is measured again at time 2, and B at time 3. In this case, the diagram indicates that both A_1 and B_1 are assumed to affect A_2, but that the effect of A_1 on B at time 3 is wholly via A_2—there is no direct arrow drawn leading from A_1 to B_3. It is assumed that A_1 and B_1 are correlated, and that A_2 and B_3 are subject to additional influences independent of A and B, here represented by short, unlabeled arrows. These additional influences could have been labeled, say, X and Y, but are often left unlabeled in path diagrams, as here, to indicate that they refer to other, unspecified influences on the variable to which they point. Such arrows are called *residual arrows* to indicate that they represent causes residual to those explicitly identified in the diagram.

The meaning of "cause" in a path diagram

Straight arrows in path diagrams are said to represent causal relationships—but in what sense of the sometimes slippery word "cause"? In fact, we do not need to adopt any strict or narrow definition of cause in this book, because path diagrams can be—and are—used to represent causes of various kinds, as the examples we have considered suggest. The essential feature for the use of a causal arrow in a path diagram is the assumption that a change in the variable at the tail of the arrow will result in a change in the variable at the head of the arrow, all else being equal (i.e., with all other variables in the diagram held constant). Note the one-way nature of this process—imposing a change on the variable at the head of the arrow does *not* bring about a change in the tail variable. A variety of common uses of the word "cause" can

be expressed in these terms, and hence can legitimately be represented by a causal arrow in a path diagram.

Completeness of a path diagram

Variables in a path diagram may be grouped in two classes: those that do not receive causal inputs from any other variable in the path diagram, and those that receive one or more such causal inputs. Variables in the first of these two classes are referred to as *exogenous*, *independent*, or *source* variables. Variables in the second class are called *endogenous*, *dependent*, or *downstream* variables. *Exogenous variables* (Greek: "of external origin") are so called because their causal sources lie external to the path diagram; they are causally independent with respect to other variables in the diagram—straight arrows may lead away from them but never toward them. These variables represent causal sources in the diagram. Examples of such source variables in Fig. 1.3 are A_1, B_1, and the two unlabeled residual variables. *Endogenous variables* ("of internal origin") have at least some causal sources that lie within the path diagram; these variables are causally dependent on other variables—one or more straight arrows lead into them. Such variables lie causally *downstream* from source variables. Examples of downstream variables in Fig. 1.3 are A_2 and B_3. In Fig. 1.2, U, T, and V are source variables, and A and B are downstream variables. Look back at Fig. 1.1. Which are the source and downstream variables in this path diagram? (We hope you identified A, B, and X as source variables, and C as downstream.)

In a proper and complete path diagram, all the source variables are interconnected by curved arrows, to indicate that they may be correlated—unless it is explicitly assumed that their correlation is zero, in which case the curved arrow is omitted. Thus the absence of a curved arrow between two source variables in a path diagram, as between X and A in Fig. 1.1, or T and U in Fig. 1.2, is not an expression of ignorance but an explicit statement about assumptions underlying the diagram.

Downstream variables, on the other hand, are never connected by curved arrows in path diagrams. (Actually, some authors use downstream curved arrows as a shorthand to indicate correlations among downstream variables caused by other variables than those included in the diagram: We use correlations between residual arrows for this purpose, which is consistent with our convention because the latter are source variables.) Residual arrows point at downstream variables, never at source variables. Completeness of a path diagram requires that a residual arrow be attached to every downstream variable unless it is explicitly assumed that all the causes of variation of that variable are included among the variables upstream from it in the diagram. (This convention is also not universally adhered to: Occasionally, path diagrams are published with the notation "residual arrows omitted." This is an unfortunate practice

Figure 1.4 Path diagrams illustrating the implication of an omitted residual arrow.

because it leads to ambiguity in interpreting the diagram: Does the author intend that all the variation in a downstream variable is accounted for within the diagram, or not?)

Fig. 1.4 shows an example in which the presence or absence of a residual arrow makes a difference. The source variables G and E refer to the genetic and environmental influences on a trait T. The downstream variable T in Fig. 1.4(a) has no residual arrow. That represents the assumption that the variation of T is completely explained by the genetic and environmental influences upon it. This is a theoretical assumption that one might sometimes wish to make. Fig. 1.4(b), however, represents the assumption that genetic and environmental influences are not sufficient to explain the variation of T—some additional factor or factors, perhaps measurement error or gene-environment interaction—may need to be taken into account in explaining T. Obviously, the assumptions in Figs. 1.4(a) and 1.4(b) are quite different, and one would not want it assumed that (a) was the case when in fact (b) was intended.

Finally, all significant direct causal connections between source and downstream variables, or between one downstream variable and another, should be included as straight arrows in the diagram. Omission of an arrow between A_1 and B_3 in Fig. 1.3 is a positive statement: that A_1 is assumed to affect B_3 only by way of A_2.

The notion of completeness in path diagrams should not be taken to mean that the ideal path diagram is one containing as many variables as possible connected by as many arrows as possible. Exactly the opposite is true. The smallest number of variables connected by the smallest number of arrows that can do the job is the path diagram to be sought for, because it represents the most parsimonious explanation of the phenomenon under consideration. Big, messy path diagrams are likely to give trouble in many ways. Nevertheless, often the simplest explanation of an interesting behavioral or biological phenomenon does involve causal relationships among a number of variables, not all observable. A path diagram provides a way of representing in a clear and straightforward fashion what is assumed to be going on in such a case.

Notice that most path diagrams could in principle be extended indefinitely back past their source variables: These could be taken as downstream variables in an extended path diagram, and the correlations among them explained by the linkages among their own causes. Thus, the parents in Fig. 1.1 could be taken as children in

their own families, and the correlation between them explained by a model of the psychological and sociological mechanisms that result in mates having similar IQs. Or in Fig. 1.3, one could have measured A and B at a preceding time zero, resulting in a diagram in which the correlation between A_1 and B_1 is replaced by a superstructure of causal arrows from A_0 and B_0, themselves probably correlated. There is no hard-and-fast rule in such cases, other than the general maxim that simpler is better, which usually means that if going back entails multiplying variables, do not do it unless you have to. Sometimes, of course, you have to, when some key variable lies back upstream.

Other assumptions in path diagrams

It is assumed in path diagrams that causes are unitary, that is, in a case such as Fig. 1.2, that it is meaningful to think of a single variable T that is the cause of A and B, and not (say) two separate and distinct aspects of a phenomenon T, one of which causes A and one B. In the latter case, a better representation would be to replace T by two different (possibly correlated) variables.

An exception to the rule of unitary causes is residual variables, which typically represent multiple causes of a variable that are external to the path diagram. Perhaps for this reason, path analysts do not always solve for the path coefficients associated with the residual arrows in their diagrams. It is, however, good practice to solve at least for the proportion of variance associated with such residual causes (more on this later). It is nearly always useful to know what proportion of the variation of each downstream variable is accounted for by the causes explicitly included within the path diagram, and what proportion is not.

Another assumption made in path diagrams is that the causal relationships represented by straight arrows are linear. This is usually not terribly restricting—mild departures from linearity are often reasonably approximated by linear relationships, and if not, it may be possible to transform variables so as to linearize their relationships with other variables. The use of log income, rather than income, or reciprocals of latency measures, or arcsine transformations of proportions would be examples of transformations often used by behavioral scientists for this purpose. In drawing a path diagram, one ordinarily does not have to worry about such details—one can always make the blanket assumption that one's variables are measured on scales for which relationships are reasonably linear. But in evaluating the strength of causal effects with real data, the issue of nonlinearity may arise. If variable A has a positive effect on variable B in part of its range and a negative effect in another, it is hard to assign a single number to represent the effect of A on B. However, if A is suitably redefined, perhaps as an absolute deviation from some optimum value,

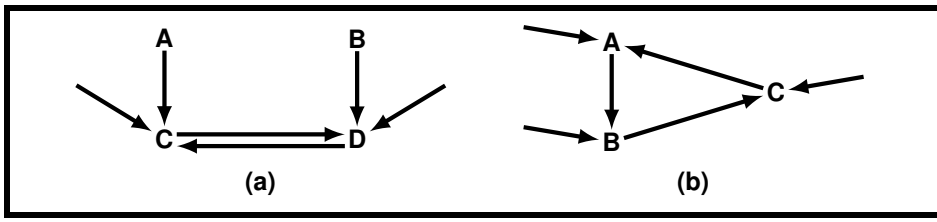


Figure 1.5 Path diagrams with: (a) mutual influences and (b) a feedback loop.

this may be possible. In Chapter 3 we consider some approaches to dealing with nonlinear relationships of latent variables.

Feedbacks and mutual influences

In our examples so far we have restricted ourselves to path diagrams in which, after the source variables, there was a simple downstream flow of causation—no paths that loop back on themselves or the like. Most of the cases we consider in this book have this one-way causal flow, but path representations can be used to deal with more complex situations involving causal loops, as we see in a later chapter. Examples of two such non-one-way cases are shown in Fig. 1.5. In Fig. 1.5(a) there is a mutual causal influence between variables C and D: each affects the other. A causal sequence could go from A to C to D to C to D again and so on. In Fig. 1.5(b) there is an extended feedback loop: A affects B which affects C which in turn affects A.

Direct and indirect causal paths

Sometimes it is useful to distinguish between direct and indirect causal effects in path diagrams. A direct effect is represented by a single causal arrow between the two variables concerned. In Fig. 1.5(b) variable B has a direct effect on variable C. There is a causal arrow leading from B to C. If B is changed, we expect to observe a change in C. Variable A, however, has only an indirect effect on C because there is no direct arrow from A to C. There is, however, an indirect causal effect transmitted via variable B. If A changes, B will change, and B's change will affect C, other things being equal. Thus, A can be said to have a causal effect on C, although an indirect one. In Fig. 1.5(a) variable B has a direct effect on variable D, an indirect effect on variable C, and no causal effect at all on variable A.

Path Analysis

Path diagrams are useful enough as simple descriptive devices, but they can be much more than that. Starting from empirical data, one can solve for a numerical value of each curved and straight arrow in a diagram to indicate the relative strength of that

correlation or causal influence. Numerical values, of course, imply scales on which they are measured. For most of this chapter we assume that all variables in the path diagram are expressed in standard score form, that is, with a mean of zero and a standard deviation of one. Covariances and correlations are thus identical. This simplifies matters of presentation, and is a useful way of proceeding in many practical situations. Later, we see how the procedures can be applied to data in original raw-score units, and consider some of the situations in which this approach is preferred. We also assume for the present that we are dealing with unlooped path diagrams.

The steps of constructing and solving path diagrams are referred to collectively as *path analysis*, a method originally developed by the American geneticist Sewall Wright as early as 1920, but only extensively applied in the social and behavioral sciences during the last few decades.

Wright's tracing rules

Briefly, Wright showed that if a situation can be presented as a proper path diagram, then the correlation between any two variables in the diagram can be expressed as the sum of the compound paths connecting these two points, where a compound path is a path along arrows that follows three rules:

(a) no loops;
(b) no going forward then backward;
(c) a maximum of one curved arrow per path.

The first rule means that a compound path must not go twice through the same variable. In Fig. 1.6(a) the compound path ACF would be a legitimate path between A

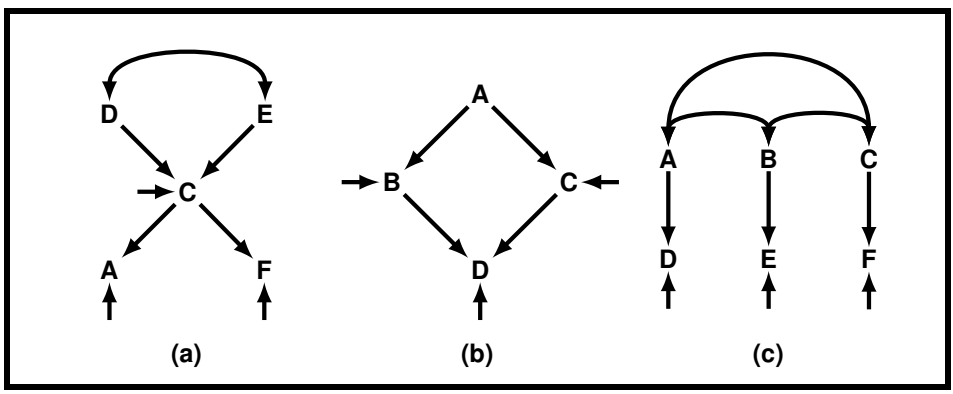


Figure 1.6 Illustrations of Wright's tracing rules.

and F, but the path ACDECF would not be because it involves going twice through variable C.

The second rule means that on a particular path, after one has once gone forward along one or more arrows, it is not legitimate to proceed backwards along others. (Going backward first and then forward is, however, quite proper.) In Fig. 1.6(b) the compound path BAC is a legitimate way to go from B to C; the path BDC is not. In the former, one goes backward along an arrow (B to A) and then forward (A to C), which is allowable, but path BDC would require going forward then backward, which is not. This asymmetry may seem a bit less arbitrary if one realizes that it serves to permit events in the diagram to be connected by common causes (A), but not by common consequences (D). The third rule is illustrated in Fig. 1.6(c). DACF is a legitimate compound path between D and F; DABCF is not, because it would require traversing two curved arrows. Likewise, DABE is a legitimate path between D and E, but DACBE is not.

Fig. 1.7 serves to provide examples of tracing paths in a path diagram according to Wright's tracing rules. This figure incorporates three source variables, A, B, and C, and three downstream variables, D, E, and F. We have designated each arrow by a lower case letter for convenience in representing compound paths. Each lower case letter stands for the value or magnitude of the particular causal effect or correlation. A simple rule indicates how these values are combined: *The numerical value of a compound path is equal to the product of the values of its constituent arrows*. Therefore, simply writing the lower case letters of a path in sequence is at the same time writing an expression for the numerical value of that path.

For example, what is the correlation between variables A and D in Fig. 1.7? Two paths are legal: *a* and *fb*. A path like *hgb* would be excluded by the rule about only one curved arrow, and paths going further down the diagram like *adcgb* would violate both the rules about no forward then backward and no loops. So the numerical value

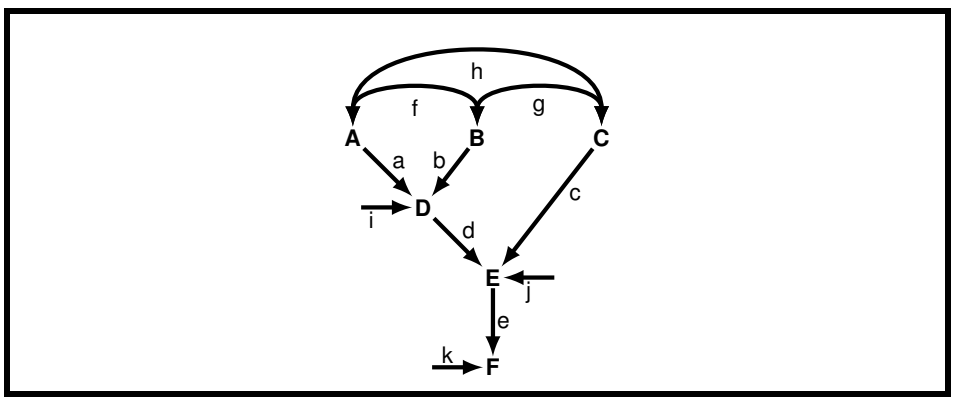


Figure 1.7 Examples of tracing paths in a path diagram.

of r_{AD} can be expressed as $a + fb$. We hope that readers can see that $r_{BD} = b + fa$, and that $r_{CD} = gb + ha$.

What about r_{AB}? Just f. Path hg would violate the third rule, and paths like ab or $adcg$ would violate the second. It is, of course, quite reasonable that r_{AB} should equal f, because that is just what the curved arrow between A and B means. Likewise, $r_{BC} = g$ and $r_{AC} = h$.

Let us consider a slightly more complicated case: r_{AE}. There are three paths: ad, fbd, and hc. Note that although variable D is passed through twice, this is perfectly legal, because it is only passed through once on any given path. You might wish to pause at this point to work out r_{BE} and r_{CE} for yourself.

(We hope you got: $bd + fad + gc$ and $c + gbd + had$.)

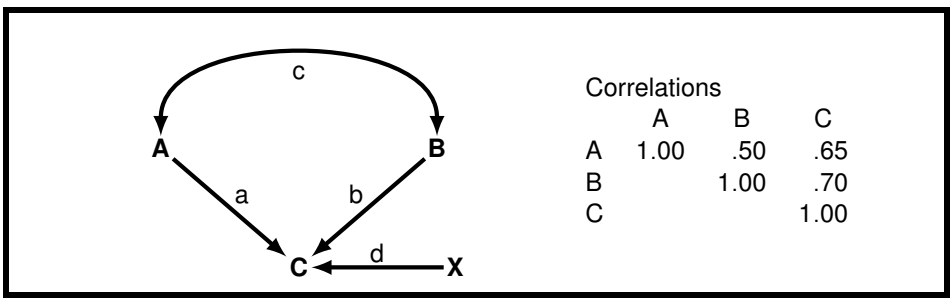
Now you might try your hand at some or all of the six remaining correlations in Fig. 1.7: r_{DE}, r_{EF}, r_{BF}, r_{CF}, r_{DF}, and r_{AF}. (The answers are not given until later in the chapter, to minimize the temptation of peeking at them first.)

Numerical solution of a path diagram

Given that we can express each of the correlations among a set of observed variables in a path diagram as a sum of compound paths, can we reverse this process and solve for the values of the causal paths given the correlations? The answer is that often we can.

Consider the example of Fig. 1.1, redrawn as Fig. 1.8. Recall that variables A and B were fathers' and mothers' intelligence, and C was children's intelligence. X is a residual variable, representing other unmeasured influences on child's intelligence that are independent of the parents' intelligence.

Suppose that in some suitable population of families we were to observe the correlations shown on the right in Fig. 1.8. We can now, using our newfound knowledge of path analysis (and ignoring X for the moment), write the following three



Correlations

	A	B	C
A	1.00	.50	.65
B		1.00	.70
C			1.00

Figure 1.8 Example of Fig. 1.1, with observed correlations of A, B, and C.

equations:

$$r_{AB} = c$$
$$r_{AC} = a + cb$$
$$r_{BC} = b + ca.$$

Because we know the observed values r_{AB}, r_{AC}, and r_{BC}, we have three simultaneous equations in three unknowns:

$$c = .50$$
$$a + cb = .65$$
$$b + ca = .70.$$

Substitution for c in the second and third equations yields two equations in two unknowns:

$$a + .50b = .65$$
$$.50a + b = .70.$$

These equations are readily solved to yield $a = .40$ and $b = .50$. Thus, if we were to observe the set of correlations given in Fig. 1.8 , *and if our causal model is correct*, we could conclude that the causal influences of fathers' and mothers' intelligence on child's intelligence could be represented by values of .40 and .50, respectively, for the causal paths a and b.

What do these numbers mean? They are, in fact, standardized partial regression coefficients—we call them *path coefficients* for short. Because they are *regression coefficients*, they tell us to what extent a change on the variable at the tail of the arrow is transmitted to the variable at the head of the arrow. Because they are *partial* regression coefficients, this is the change that occurs with all other variables in the diagram held constant. Because they are *standardized* partial regression coefficients, we are talking about changes measured in standard deviation units. Specifically, the value of .40 for a means that if we were to select fathers who were one standard deviation above the mean for intelligence—but keeping mothers at the mean—their offspring would average four tenths of a standard deviation above the population mean. (Unless otherwise specified, we are assuming in this chapter that the numbers we deal with are population values, so that issues of statistical inference do not complicate the picture.)

Because paths a and b are standardized partial regression coefficients—also known in multiple regression problems as *beta weights*—one might wonder if we can solve for them as such, by treating the path analysis as a sort of multiple regression

problem. The answer is: Yes, we can—at least in cases where all variables are measured. In the present example, A, B, and C are assumed known, so we can solve for *a* and *b* by considering this as a multiple regression problem in predicting C from A and B.

Using standard formulas (e.g., McNemar, 1969, p. 192):

$$\beta_1 = \frac{.65 - (.70 \times .50)}{1 - .50^2} = .40$$
$$\beta_2 = \frac{.70 - (.65 \times .50)}{1 - .50^2} = .50,$$

or exactly the same results as before.

Viewing the problem in this way, we can also interpret the squared multiple correlation between C and A and B as the proportion of the variance of C that is accounted for by A and B jointly. In this case

$$R^2_{C \cdot AB} = \beta_1 r_{AC} + \beta_2 r_{BC} = .40 \times .65 + .50 \times .70 = .61.$$

Another way in which we can arrive at the same figure from the path diagram is by following a path-tracing procedure. We can think of the predicted variance of C as that part of its correlation with itself that occurs via the predictors. In this case, this would be *the sum of the compound paths from C to itself via A or B or both*. There is the path to A and back, with value a^2, the path to B and back, with value b^2, and the two paths *acb* and *bca*: $.40^2 + .50^2 + 2 \times .40 \times .50 \times .50 = .16 + .25 + .20 = .61$.

We can then easily solve for the value of the path *d* which leads from the unmeasured residual X. The variance that A and B jointly account for is R^2, or .61. The variance that X accounts for is thus 1 - R^2; that is, $1 - .61$, or .39. The correlation of C with itself via X is the variance accounted for by X, and this is just *dd*. So the value of *d* is $\sqrt{.39}$, or .62.

So long as all variables are measured one can proceed to solve for the causal paths in a path diagram as beta weights in a series of multiple regression analyses. Thus, in Fig. 1.7 one could solve for *a* and *b* from the correlations among A, B, and D; for *d* and *c* from the correlations among D, C, and E; and for *e* as the simple correlation between E and F. The residuals *i*, *j*, and *k* can then be obtained as $\sqrt{1 - R^2}$ in the various multiple regressions.

In general, however, we must deal with path diagrams involving unmeasured, latent variables. We cannot directly calculate the correlations of these with observed variables, so a simple multiple regression approach does not work. We need, instead, to carry out some variant of the first approach—that is, to solve a set of simultaneous equations with at least as many equations as there are unknown values to be obtained.

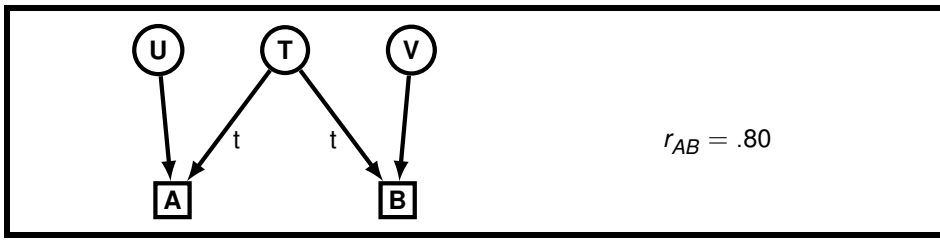


Figure 1.9 The example of Fig. 1.2, with observed correlation of .80 between alternate forms A and B of a test.

Consider the example of Fig. 1.2, test reliability, repeated for convenience as Fig. 1.9. Because this diagram involves both latent variables and observed variables, we have followed a common practice of latent variable modelers by putting the letters representing observed variables in squares (or rectangles), and variables representing latent variables in circles (or ovals).

We wish to solve for the values of the causal paths between the true score T and the observed scores A and B. But T is an unobserved, latent variable; all we have is the observed correlation .80 between forms A and B of the test. How can we proceed? If we are willing to assume that A and B have the same relation to T, which they should have if they are really parallel alternate forms of a test, we can write from the path diagram the equation

$$r_{AB} = t^2 = .80,$$

from which it follows that $t = \sqrt{.80} = .89$. It further follows that t^2 or 80% of the variance of each of the alternate test forms is attributable to the true score on the trait, that 20% is due to error, and that the values of the residual paths from U and V are $\sqrt{.20}$ or .45.

Fig. 1.10 presents another case of a path diagram containing a latent variable. It is assumed that A, C, and D are measured, as shown by the squares. Their correlations are given to the right of the figure. B, as indicated by the circle, is not measured, so we do not know its correlations with A, C, and D. We can, however, write equations for the

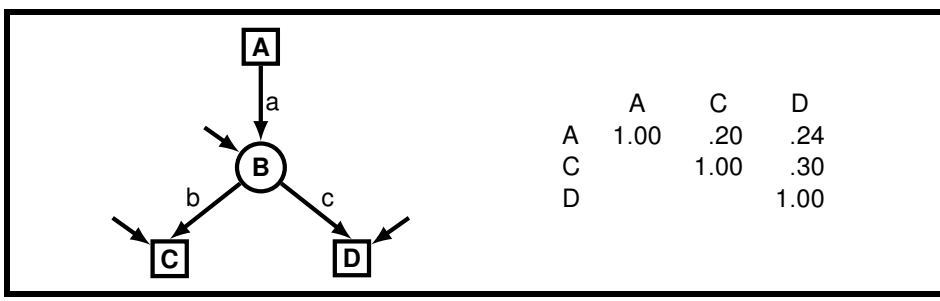


Figure 1.10 Another simple path diagram with a latent variable.

three known correlations in terms of the three paths a, b, and c, and (as it turns out) these three equations can be solved for the values of the three causal paths.

The equations are:

$$r_{AC} = ab$$
$$r_{AD} = ac$$
$$r_{CD} = bc.$$

A solution is:

$$\frac{r_{AC} \times r_{CD}}{r_{AD}} = \frac{ab \times bc}{ac} = b^2 = \frac{.20 \times .30}{.24} = .25 \implies b = .50$$
$$a = \frac{r_{AC}}{b} = \frac{.20}{.50} = .40$$
$$c = \frac{r_{AD}}{a} = \frac{.24}{.40} = .60.$$

Note that another possible solution would be numerically the same, but with all paths negative, because b^2 also has a negative square root. This would amount to a model in which B, the latent variable, is scored in the opposite direction, thus reversing its relationships with the manifest variables.

(By the way, to keep the reader in suspense no longer about the correlations in Fig. 1.7: $r_{DE} = d + ahc + bgc$, $r_{EF} = e$, $r_{BF} = bde + fade + gce$, $r_{CF} = ce + gbde + hade$, $r_{DF} = de + ahce + bgce$, and $r_{AF} = ade + fbde + hce$.)

Underdetermined, overdetermined, and just-determined path diagrams

Fig. 1.11(a) shows another simple path diagram. It is somewhat like Fig. 1.10 upside down: Instead of one cause of the latent variable and two effects, there are now two causes and one effect.

However, this change has made a critical difference. There are still just three correlations among the three observed variables A, B, and D, yielding three equations. But now there are four unknown values to be estimated: a, b, c, and d. One observed correlation, r_{AB}, estimates d directly. But that leaves only two equations, $r_{AD} = ac + dbc$ and $r_{BD} = bc + dac$, to estimate the three unknowns, a, b, and c, and no unique solution is possible. The path diagram is said to be *underdetermined* (or *unidentified*).

In the preceding problem of Fig. 1.10, there were three equations in three unknowns, and an exact solution was possible. Such a case is described as *just determined* (or *just identified*). Fig. 1.11(b) shows a third case, of an *overdetermined* (or *overidentified*) path diagram. As in Fig. 1.11(a), C is a latent variable and A and B are source variables, but an additional measured downstream variable E has been

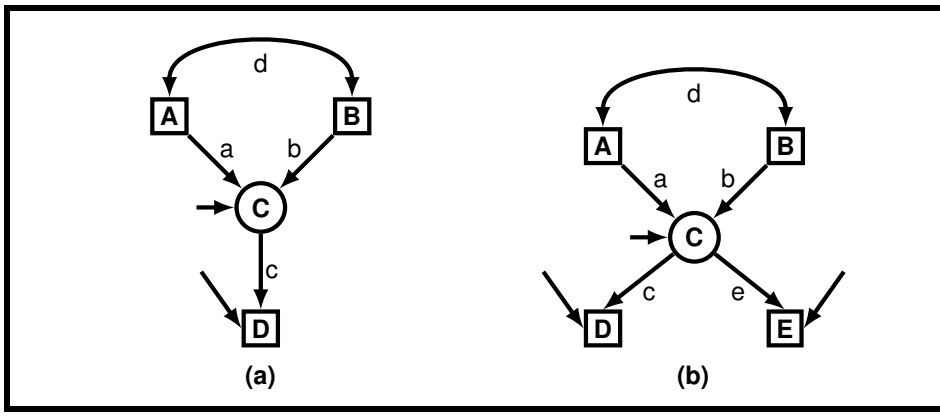


Figure 1.11 Path diagrams that are: underdetermined (a) and overdetermined (b).

added. Now there are six observed correlations among the observed variables A, B, D, and E, yielding six equations, whereas we have only added one unknown, giving five unknowns to be solved for. More equations than unknowns does not guarantee overdetermination, but in this case for most observed sets of correlations there will be no single solution for the unknowns that will satisfy all six equations simultaneously. What is ordinarily done in such cases is to seek values for the unknowns that come as close as possible to accounting for the observed correlations (we defer until the next chapter a consideration of what "as close as possible" means).

It might be thought that just-determined path diagrams, because they permit exact solutions, would be the ideal to be sought for. But in fact, for the behavioral scientist, overdetermined path diagrams are usually much to be preferred. The reason is that the data of the behavioral scientist typically contain sampling and measurement error, and an exact fit to these data is an exact fit to the error as well as to the truth they contain. Whereas—if we assume that errors occur at random—a best overall fit to the redundant data of an overdetermined path diagram will usually provide a better approximation to the underlying true population values. Moreover, as we see later, overdetermined path diagrams permit statistical tests of goodness of fit, which just-determined diagrams do not.

A computer-oriented symbolism for path diagrams—RAM

A way of drawing path diagrams which has advantages for translating them into computer representations has been developed by John McArdle. He called his general approach to path modeling *Reticular Action Modeling*—RAM for short.

Fig. 1.12(a) (next page) is a reproduction of Fig. 1.10, while Fig. 1.12(b) shows the same model in a RAM representation. The following points may be noted:

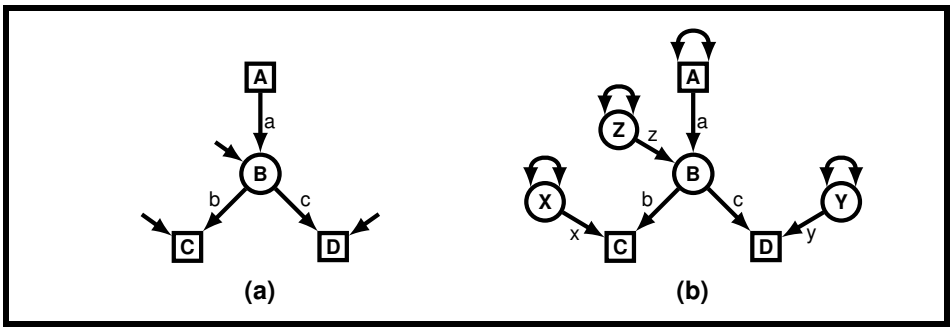


Figure 1.12 The path model of Fig. 1.10 (a) shown in RAM symbolism (b).

1. Latent variables are designated by placing them in circles, observed variables by placing them in squares, as usual in latent variable modeling. In addition to squares and circles for variables, RAM uses triangles to designate constants—not involved in the present example, but important in models involving means, discussed later in this book.
2. Residual variables are represented explicitly as latent variables (X, Y, Z).
3. Two-headed curved arrows leaving and re-entering the same variable are used to represent the variance of source variables. When they are unlabeled, as here, they are assumed to have a value of 1.0—thus these are standardized variables. Curved arrows connecting two different source variables represent their covariance or correlation, in the usual manner of path diagrams.

Although a little cumbersome in some respects—which is why we do not use it routinely in this book—RAM symbolism, by rendering explicitly a number of things often left implicit in path diagrams, facilitates a direct translation into computer representations. We will see examples in Chapter 2.

Factor Models

An important subdivision of latent variable analysis is traditionally known as factor analysis. In recent discussions of factor analysis, a distinction is often drawn between *exploratory* and *confirmatory* varieties. In exploratory factor analysis, which is what is usually thought of as “factor analysis” if no qualification is attached, one seeks under rather general assumptions for a simple latent variable structure, one with no causal arrows from one latent variable to another, that could account for the correlations of an observed set of variables. In confirmatory factor analysis, on the other hand, one takes a specific hypothesized structure of this kind and sees how well it accounts for the observed relationships in the data.

Traditionally, textbooks on factor analysis discuss the topic of exploratory factor analysis at length and in detail, and then they put in something about confirmatory factor analysis in the later chapters. We, however, find it instructive to proceed in the opposite direction, to consider first confirmatory factor analysis and structural equation modeling more broadly, and to defer an extended treatment of exploratory factor analysis until later (Chapters 5 and 6).

From this perspective, exploratory factor analysis is a preliminary step that one might sometimes wish to take to locate latent variables to be studied via structural modeling. It is by no means a necessary step. Theory and hypothesis may lead directly to confirmatory factor analysis or other forms of structural models, and path diagrams provide a natural and convenient way of representing the hypothesized structures of latent and manifest variables that the analyst wishes to compare to real-world data.

The origins of factor analysis: Charles Spearman and the two-factor theory of intelligence

As it happens, the original form of factor analysis, invented by the British psychologist Charles Spearman in the early 1900s, was more confirmatory than exploratory, in the sense that Spearman had an explicit theory of intellectual performance that he wished to test against data. Spearman did not use a path representation, Wright not yet having invented it, but Fig. 1.13 represents the essentials of Spearman's (1904) theory in the form of a path diagram.

Spearman hypothesized that performance on each of a number of intellectual tasks shared something in common with performance on all other intellectual tasks, a factor of general intellectual ability that Spearman called *g*. (*g* originally stood for general intelligence, but later he referred to the factor as just *g*). In addition to *g*, he believed that performance on each task also involved a factor of skills specific to that task, hence the designation *two-factor theory*. In Spearman's words: "All branches of intellectual activity have in common one fundamental function (or group of functions),

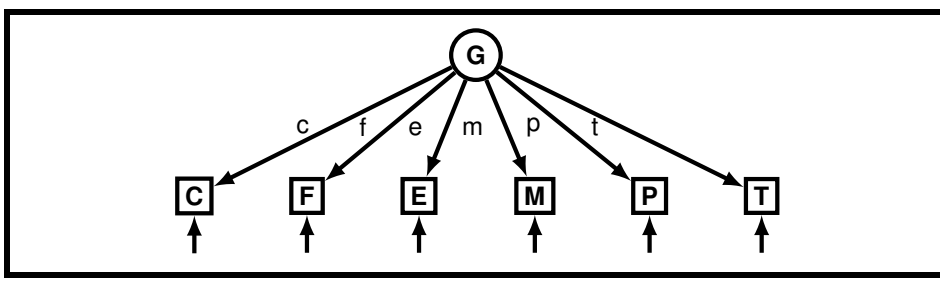


Figure 1.13 Path representation of Spearman's two-factor theory.

whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others" (p. 284).

Spearman obtained several measures on a small group of boys at an English preparatory school: a measure of pitch discrimination, a ranking of musical talent, and examination grades in several academic areas—Classics, French, English studies, and Mathematics. Fig. 1.13 applies his two-factor theory to these data. The letter G at the top of the figure represents the latent variable of *g*; C, F, E, and M at the bottom represent observed performances in the academic subjects, P stands for pitch discrimination and T for musical talent. General intellectual ability is assumed to contribute to all these performances. Each also involves specific abilities, represented by the residual arrows.

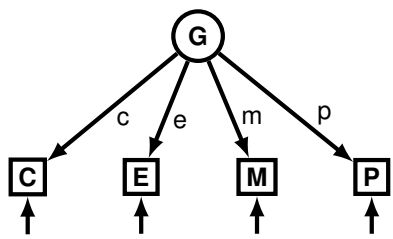
If Spearman's theory provides an adequate explanation of these data, the path diagram implies that the correlation between any two tasks should be equal to the product of the paths connecting them to the general factor: the correlation between Classics and Mathematics should be *cm*, that between English and French should be *ef*, between French and musical talent *ft*, and so on. Because we are attempting to explain $6 \times 5/2 = 15$ different observed correlations by means of 6 inferred values—the path coefficients *c, f, e, m, p*, and *t*—a good fit to the data is by no means guaranteed. If one is obtained, it is evidence that the theory under consideration has some explanatory power.

Fig. 1.14 gives the correlations for part of Spearman's data: Classics, English, Mathematics, and pitch discrimination.

If the single general-factor model fit the data exactly, we could take the correlations among any three variables and solve for the values of the three respective path coefficients, since they would provide three equations in three unknowns. For example:

$$\frac{r_{CE} \times r_{CM}}{r_{EM}} = \frac{cecm}{em} = c^2 = \frac{.78 \times .70}{.64} = .853 \implies c = .92$$

$$\frac{r_{EM} \times r_{CE}}{r_{CM}} = \frac{emce}{cm} = e^2 = \frac{.64 \times .78}{.70} = .713 \implies e = .84$$



Correlations

	C	E	M	P
C	1.00	.78	.70	.66
E		1.00	.64	.54
M			1.00	.45
P				1.00

Figure 1.14 Data to illustrate the method of triads.

$$\frac{r_{CM} \times r_{EM}}{r_{CE}} = \frac{cmem}{ce} = m^2 = \frac{.70 \times .64}{.78} = .574 \implies m = .76.$$

This procedure has been given a name; it is called the *method of triads*. (A *triad* is the ratio of correlations among three variables.) For any three variables, there will always be three different triads (e.g., $r_{12} \times r_{13}/r_{23}$, $r_{12} \times r_{23}/r_{13}$, $r_{13} \times r_{23}/r_{12}$). If the data, as here, only approximately fit a model with a single general factor, one will get slightly different values for a particular path coefficient depending on which triads one uses. For example, we may solve for m in two other ways from these data:

$$\frac{r_{CM} \times r_{MP}}{r_{CP}} = \frac{cmmp}{cp} = m^2 = \frac{.70 \times .45}{.66} = .477 \implies m = .69$$
$$\frac{r_{EM} \times r_{MP}}{r_{EP}} = \frac{emmp}{ep} = m^2 = \frac{.64 \times .45}{.54} = .533 \implies m = .73.$$

These three values of m are not very different. One might consider simply averaging them to obtain a compromise value. A slightly preferable method, because it is less vulnerable to individual aberrant values, adds together the numerators and denominators of the preceding expressions, and then divides:

$$m^2 = \frac{.70 \times .64 + .70 \times .45 + .64 \times .45}{.78 + .66 + .54} = .531 \implies m = .73.$$

You may wish to check your understanding of the method by confirming that it yields .97 for *c*, .84 for *e*, and .65 for *p* for the data of Fig. 1.14.

We may get some sense of how accurately our solution accounts for the observed correlations among the four variables by producing the correlation matrix implied by the paths (i.e., *ce*, *cm*, *cp*, *em*, *ep*, *mp*):

$$\begin{array}{ccc} .81 & .71 & .63 \\ & .61 & .55 \\ & & .47 \end{array}$$

As is evident, the implied correlations under the model do not differ much from the observed correlations—the maximum absolute difference is .03. The assumption of a single general factor plus a residual factor for each measure does a reasonable job of accounting for the data.

We may as well go on and estimate the variance accounted for by each of the specific (residual) factors. Following the path model, the proportion of the variance of each test accounted for by a factor equals the correlation of that test with itself by way of the factor (the sum of the paths to itself via the factor). In this case, these have the value c^2, e^2, etc. The variances due to the general factor are thus .93, .70, .53, and .42 for Classics, English, Mathematics, and pitch discrimination, respectively, and the

corresponding residual variances due to specific factors are .07, .30, .47, and .58. In traditional factor analytic terminology, the variance a test shares with other tests in the battery is called its *communality*, symbolized h^2, and the variance not shared is called its *uniqueness*, symbolized u^2. The h^2 values of the four measures are thus .93, .70, .53, and .42, and their uniqueness values are .07, .30, .47, and .58. Pitch discrimination has the least in common with the other three measures; Classics has the most.

The observant reader will notice that the communality and uniqueness of a variable are just expressions in the factor analytic domain of the general notion of the predicted (R^2) and residual variance of a downstream variable in a path diagram, as discussed earlier in the chapter.

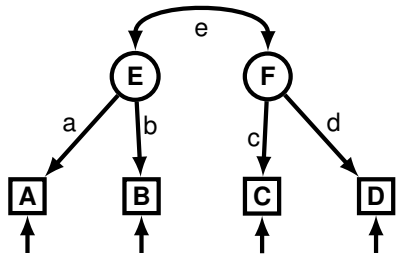
The path coefficients *c, e, m*, etc. are in factor-analytic writing called the *factor pattern coefficients* (or more simply, the *factor loadings*). The correlations between the tests and the factors, here numerically the same as the pattern coefficients, are collectively known as the *factor structure*.

More than one common factor

As soon became evident to Spearman's followers and critics, not all observed sets of correlations are well explained by a model containing only one general factor; factor analysts soon moved to models in which more than one latent variable was postulated to account for the observed correlations among measures. Such latent variables came to be called *common* factors, rather than *general* factors because, although they were common to several of the variables under consideration, they were not general to all. There remained, of course, *specific* factors unique to each measure, although they are usually combined with error into a single residual term for each variable.

Fig. 1.15 gives an example of a path diagram in which there are two latent variables, E and F, and four observed variables, A, B, C, and D. E is hypothesized as influencing A and B, and F as influencing C and D. In the path diagram there are five unknowns, the paths *a, b, c*, and *d*, and the correlation *e* between the two latent variables. There are six equations, shown to the right of the diagram, based on the six correlations between pairs of observed variables. Hypothetical values of the observed correlations are given (e.g., .60 for r_{AB}). Because there are more equations than unknowns, one might expect that a single exact solution would not be available, and indeed this is the case. An iterative least squares solution, carried out in a way discussed in the next chapter, yielded the values shown to the far right of Fig. 1.15.

Table 1-1 reports a typical factor analysis solution based on Fig. 1.15. The factor pattern represents the values of the paths from factors to variables—the paths *a* and *b* and two zero paths from E to A, B, C, and D, and similarly for F. The factor structure presents the correlations of the variables with the factors: for factor E these have the



$r_{AB} = ab = .60$ $a = .658$
$r_{AC} = aec = .15$ $b = .912$
$r_{AD} = aed = .10$ $c = .833$
$r_{BC} = bec = .20$ $d = .600$
$r_{BD} = bed = .15$ $e = .267$
$r_{CD} = cd = .50$

Figure 1.15 A simple factor model with two correlated factors (E and F).

values *a, b, ec*, and *ed*, respectively, and for factor F, *ea, eb, c*, and *d*. The communalities (h^2) are, in this case, simply a^2, b^2, c^2, and d^2 because each variable is influenced by only one factor. Finally, the correlation between E and F is just *e*.

The reproduced correlations (i.e., those implied by the path values) and the residual correlations (the differences between observed and implied correlations) are shown at the bottom of Table 1-1, with the reproduced correlations in the upper right and the residual correlations in the lower left. The reproduced correlations are obtained by inserting the solved values of *a, b, c*, etc. into the equations of Fig. 1.15: $r_{AB} = .658 \times .912$, $r_{AC} = .658 \times .267 \times .833$, and so on. The residual correlations are

Table 1-1. Factor solution for the two-factor problem of Fig. 1.15

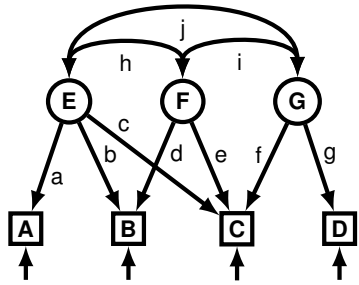
	Factor pattern		Factor structure		
Variable	E	F	E	F	h^2
A	.66	.00	.66	.18	.43
B	.91	.00	.91	.24	.83
C	.00	.83	.22	.83	.69
D	.00	.60	.16	.60	.36

Factor correlations

	E	F
E	1.00	.27
F	.27	1.00

Reproduced (upper) and residual (lower) correlations

	A	B	C	D
A		.600	.146	.105
B	.000		.203	.146
C	.004	−.003		.500
D	−.005	.004	.000	



$$r_{AB} = ab + ahd$$
$$r_{AC} = ac + ahe + ajf$$
$$r_{AD} = ajg$$
$$r_{BC} = bc + de + bhe + dif + dhc + bjf$$
$$r_{BD} = dig + bjg$$
$$r_{CD} = fg + cjg + eig$$

Figure 1.16 A more complex three-factor model.

obtained by subtracting the reproduced correlations from the observed ones. Thus the residual r_{AC} is .15 − .146, or .004.

A more complex model with three factors is shown in Fig. 1.16. Because this model has 10 unknowns and only 6 equations, it is underdetermined and cannot be solved as it stands. However, if one were to fix sufficient values by *a priori* knowledge or assumption, one could solve for the remaining values.

The factor solution in symbolic form is given in Table 1-2. By inserting known and solved-for values in place of the unknowns, one could obtain numerical values for the factor pattern, the factor structure, the communalities, and the factor correlations. Also, one could use the path equations of Fig. 1.16 to obtain the implied correlations and thence the residuals. Notice that the factor pattern is quite simple in terms of the paths, but that the factor structure (the correlations of factors with variables) and the communalities are more complex functions of the paths and factor correlations.

Table 1-2. Factor solution of Fig. 1.16 in symbolic form

	Factor pattern			Factor structure			h^2
Variable	E	F	G	E	F	G	
A	a	0	0	a	ha	ja	a^2
B	b	d	0	$b + hd$	$d + hb$	$id + jb$	$b^2 + d^2 + 2bhd$
C	c	e	f	$c + he + jf$	$e + hc + if$	$f + ie + jc$	$c^2 + e^2 + f^2 + 2che + 2eif + 2cjf$
D	0	0	g	jg	ig	g	g^2

Factor correlations

	E	F	G
E	1.0	h	j
F	h	1.0	i
G	j	i	1.0

Structural Equations

An alternative way of representing a path diagram is as a set of *structural equations*. Each equation expresses a downstream variable as a function of the causal paths leading into it. There will be as many equations as there are downstream variables.

Fig. 1.17 shows one of the path diagrams considered earlier. It has one downstream variable, hence one structural equation: The score of individuals on variable C is an additive function of their scores on A, B, and X. If the variables are standardized, the values of the weights *a, b*, and *d* required to give a best fit to the data in a least squares sense turn out to be just the standardized partial regression coefficients, or path coefficients, discussed earlier.

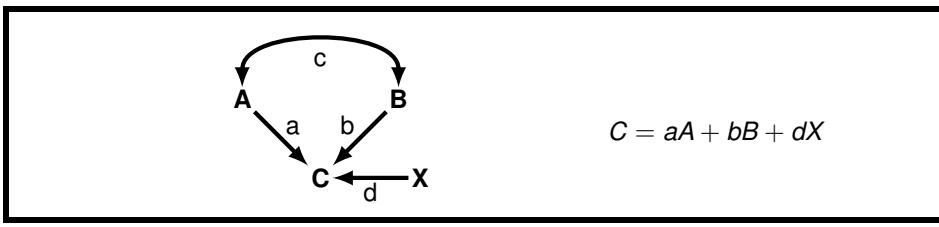


Figure 1.17 A structural equation based on a path diagram.

Fig. 1.18 gives a slightly more complex example, based on the earlier Fig. 1.13. Now there are two downstream variables, A_2 and B_3. A_2 can be expressed as a weighted additive function of the three source variables: A_1, B_1, and X (this is shown in the first equation). B_3 can be expressed in terms of A_2, B_1, and Y. Note that to construct a structural equation one simply includes a term for every straight arrow leading into the downstream variable. The term consists of the variable at the tail of the arrow times the path coefficient associated with it.

For a final example, consider the factor analysis model of Fig. 1.16 in the preceding section. The structural equations are as follows (X_A through X_D represent

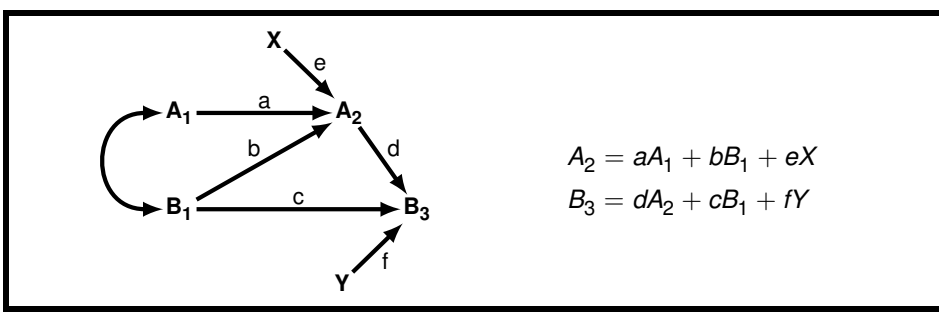


Figure 1.18 Structural equations based on the path diagram of Fig. 1.3.

the terms involving the residual arrows):

$$A = aE + X_A$$
$$B = bE + dF + X_B$$
$$C = cE + eF + fG + X_C$$
$$D = gG + X_D.$$

Notice that the equations are closely related to the rows of the factor pattern matrix (Table 1-2) with residual terms added. The solution of the set of structural equations corresponds essentially to the solution for the paths in the path diagram and would be similarly underdetermined in this instance. Again, by previously defining a sufficient number of the unknowns, the equations could be solved for those remaining.

The structural equation approach to causal models originated in economics, and the path approach in biology. For many purposes, the two may be regarded simply as alternative representations. Note, however, one difference. Path diagrams explicitly represent the correlations among source variables, whereas structural equations do not. If using the latter, supplementary specifications or assumptions must be made concerning the variances and covariances of the source variables in the model.

Original and Standardized Variables

So far, we have assumed that all the variables in a model were standardized. This has simplified the presentation, but is not a necessary restriction. Path, factor, and structural equation analyses can be carried out with variables in their original scale units as well as with standardized variables. In practice, structural equation analysis is usually done in raw-score units, path analysis is done both ways, and factor analysis is usually done with standardized variables. But this is often simply a matter of tradition or (what amounts to much the same thing) of the particular computer program used. There are occasions on which the standardized and raw-score approach each has definite advantages, so it is important to know that one can convert the results of one to the other form and be able to do so when the occasion arises.

Another way of making the distinction between analyses based on standardized and raw units is to say that in the first case one is analyzing correlations, and in the second, covariances. In the first case, one decomposes a correlation matrix among observed variables into additive components; in the second case one decomposes a variance-covariance matrix. The curved arrows in a path diagram are correlations in the first case, covariances in the second. In the first case a straight arrow in a path diagram stands for a standardized partial regression coefficient, in the second case for a raw-score partial regression coefficient. In the first case, a .5 beside a straight arrow leading from years of education to annual income means that, other things equal,

people in this particular population who are one standard deviation above the mean in education tend to be half a standard deviation above the mean in income. In the second case, if education is measured in years and income in dollars, a 5000 alongside the straight arrow between them means that, other things equal, an increase of 1 year in education represents an increase of $5000 in annual income (in this case, .5 would mean 50 cents!). In each case, the arrow between A and B refers to how much change in B results from a given change in A, but in the first case change is measured in standard deviation units of the two variables, and in the second case, in the ratio of their raw-score units (dollars of income per year of education).

Standardized regression coefficients are particularly useful when comparisons are to be made across different variables within a model, unstandardized regression coefficients when comparisons are to be made across different populations.

When comparing across variables, it can be difficult to judge their relative importance using raw-score coefficients. For example, consider the role of education and occupational status in influencing income. Suppose that income is expressed in dollars and the raw-score coefficients are 5000 for years of education and 300 for occupational status measured on a 100-point scale. Which has a stronger relation to income? From the raw coefficients it is hard to say. However, if the standardized regression coefficients for education and occupational status are .5 and .7, respectively, the greater relative influence of occupational status is more evident.

In comparing across populations, raw-score regression coefficients have the merit of independence of the particular ranges of the two variables involved in any particular study. If one study happens to have sampled twice as great a range of education as another, a difference in years of education that is, say, one-half a standard deviation in the first study would be a full standard deviation in the second. A standardized regression coefficient of .3 in one study would then describe exactly the same effect of education on income as a standardized regression coefficient of .6 in the other. This is a confusing state of affairs at best, and could be seriously misleading if the reader is unaware of the sampling difference between the studies. A raw-score regression coefficient of $2000 income per added year of education would, however, have the same meaning across the two studies. If the relevant standard deviations are known, correlations can readily be transformed into covariances, or vice versa, or a raw-score into a standardized regression coefficient and back. This allows the freedom to report results in either or both ways, or to carry out calculations in one mode and report them in the other, if desired. (We qualify this statement later—model fitting may be sensitive to the scale on which variables are expressed, especially if different paths or variances are constrained to be numerically equal—but it will do for now.)

The algebraic relationships between covariances and correlations are simple:

$$cov_{12} = r_{12} s_1 s_2$$

$$r_{12} = \frac{cov_{12}}{s_1 s_2},$$

where cov_{12} stands for the covariance between variables 1 and 2, r_{12} for their correlation, and s_1 and s_2 for their respective standard deviations.

The relationships between raw-score and standardized path coefficients are equally simple. To convert a standardized path coefficient to its raw-score form, *multiply it by the ratio of the standard deviations of its head to its tail variable*. To convert a raw-score path coefficient to standardized form, invert the process: Multiply by the ratio of the standard deviations of its tail to its head variable. These rules generalize to a series of path coefficients, as illustrated by Fig. 1.19 and Table 1-3.

The first line in Table 1-3 shows, via a process of substituting definitions and canceling, that the series of raw-score path coefficients ($a^* b^* c^*$) is equal to the series of standardized path coefficients (abc) multiplied by the ratio of standard deviations of its head and tail variables. The second line demonstrates the converse transformation from raw-score to standardized coefficients.

The rule for expressing the value of a compound path between two variables in terms of concrete path coefficients (stated for a vertically oriented path diagram) is: *The value of a compound path between two variables is equal to the product of the raw-score path coefficients and the topmost variance or covariance in the path*.

The tracing of compound paths according to the tracing rules, and adding compound paths together to yield the overall covariance, proceed in just the same way with raw-score as with standardized coefficients. The covariance between two variables in the diagram is equal to the sum of the compound paths between them. If

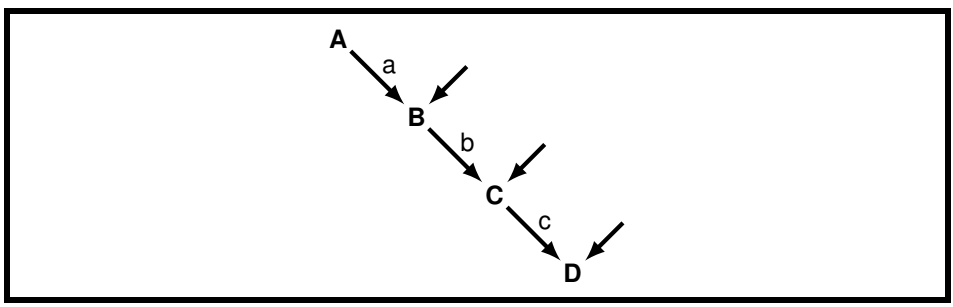


Figure 1.19 Path diagram to illustrate raw-score and standardized path coefficients.

Table 1-3. Transformation of a sequence of paths from raw-score to standardized form (example of Fig. 1.19)

$$a^* b^* c^* = a(s_B/s_A)b(s_C/s_B)c(s_D/s_C) = abc(s_D/s_A)$$
$$abc = a^*(s_A/s_B)b^*(s_B/s_C)c^*(s_C/s_D) = a^* b^* c^*(s_A/s_D)$$

Note. Asterisks designate raw-score path coefficients

there is just a single path between two variables, the covariance is equal to the value of that path. The two path diagrams in Fig. 1.20 illustrate the rule for compound paths headed by a variance and a covariance, respectively. A few examples are given in Table 1-4.

Notice that the rule for evaluating compound paths when using raw-score path coefficients is different from that for standardized coefficients only by the inclusion of one variance or covariance in each path product. Indeed, one can think of the standardized rule as a special case of the raw-score rule, because the variance of a standardized variable is 1.0, and the covariance between standardized variables is just the correlation coefficient.

If we are starting from raw data, standard deviations can always be calculated for observed variables, allowing us to express them in either raw score or standard score units, as we choose. What about the scales of latent variables, for which raw scores do not exist? There are two common options. One is simply to solve for them in standard score form and leave them that way. An alternative approach, fairly common among those who prefer to work with covariances and raw-score coefficients, is to assign the latent variable the same units as one of the observed variables. The is done by setting the path linking the latent variable to the observed variable to an arbitrary value, usually 1.0. Several examples of this procedure appear in later chapters.

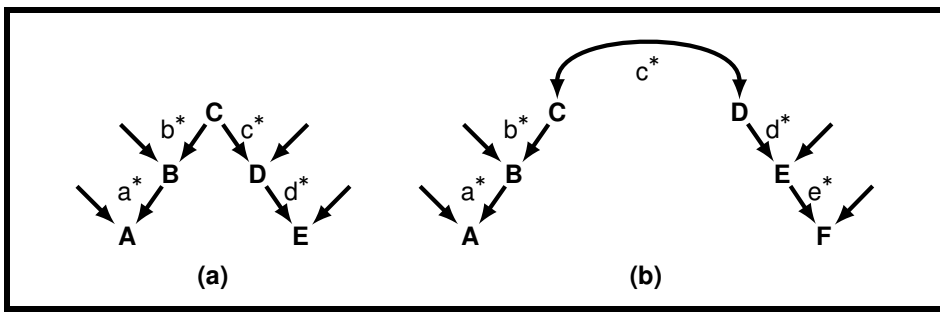


Figure 1.20 Raw-score paths with: a variance (a) and a covariance (b). (Paths a^*, b^*, c^*, etc. represent raw-score coefficients.)

Table 1-4. Illustrations of raw-score compound path rules, for path diagrams of Fig. 1.20

(a)	**(b)**
$cov_{AE} = a^* b^* s_C^2 c^* d^*$	$cov_{AF} = a^* b^* cov_{CD} d^* e^*$
$cov_{BD} = b^* s_C^2 c^*$	$cov_{CF} = cov_{CD} d^* e^*$
$cov_{CE} = s_C^2 c^* d^*$	$cov_{DF} = s_D^2 d^* e^*$

Manifest Versus Latent Variable Models

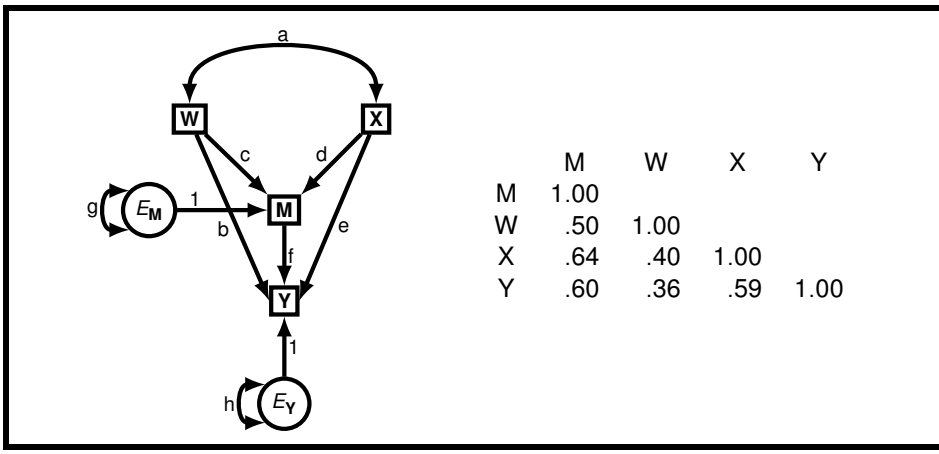
Many multivariate statistical methods, including some of those most familiar to social and behavioral scientists, do not explicitly involve latent variables. Instead, they deal with linear composites of *observed* variables. In ordinary multiple regression, for example, one seeks for an optimally weighted composite of measured independent variables to predict an observed dependent or criterion variable. In discriminant analysis, one seeks composites of measured variables that will optimally distinguish among members of specified groups. In canonical analysis one seeks composites that will maximize correlation across two sets of measured variables.

Path and structural equation analysis come in both forms: all variables measured or some not. Many of the earlier applications of such methods in economics and sociology were confined to manifest variables. The effort was to fit causal models in situations where all the variables involved were observed. Biology and psychology, having to deal with events within the organism, tended to place an earlier emphasis on the latent variable versions of path analysis. As researchers in all the social sciences become increasingly aware of the distorting effects of measurement errors on causal inferences, latent variable methods have increased in popularity, especially in theoretical contexts. In applied situations, where the practitioner must work with existing measures, errors and all, the manifest variable methods retain much of their preeminence.

Factor analysis is usually *defined* as a latent variable method—the factors are unobserved hypothetical variables that underlie and explain the observed correlations. The corresponding manifest variable method is called *component analysis*—or, in its most common form, the method of *principal components*. Principal components are linear composites of observed variables; the factors of factor analysis are always inferred entities, whose nature is at best consistent with a given set of observations, never entirely determined by them.

Chapter 1 Extended Example

For this extended example, we fit the model in Fig. 1.21. It is a mediation model—W and X are thought to influence M, which in turn influences Y (i.e., paths *cf* and *df*). The mediation is only partial, as W and X also have direct effects on Y (paths *b* and *e*). Furthermore, M has causes other than W and X, and Y has causes other than W, M, and X, as shown by the E variables (residuals). One might think, for example, of W and X as representing Kindness and Generosity, with their influence on Helping Behavior (Y) being partially due to an intervening variable that we might call Inclusiveness (i.e., the width of the group of people seen as like oneself). The model hypothesizes: (a) the broader that group, the more occasions an individual will have to



	M	W	X	Y
M	1.00			
W	.50	1.00		
X	.64	.40	1.00	
Y	.60	.36	.59	1.00

Figure 1.21 Path diagram for extended example.

engage in helping behavior; (b) that kindness and generosity are (partial) causes of Inclusiveness; and (c) that kindness and generosity may contribute directly to helping behavior as well. We will use the set of correlations among W, X, M, and Y to find standardized values for all the path coefficients.

The first step is to find the set of paths in the model that correspond to each correlation using Wright's tracing rules. Start with the curved arrows representing correlations, as those are simplest to match with values. In Fig. 1.21, there is only one such path: *a*. As it represents to correlation between W and X, its value is .40. Next, specify the direct paths. In mediation models, this can be tricky as there are typically multiple indirect paths. For example, r_{XM} is not only made up of the direct path, *d*, but also the indirect path through W: *ac*. As another example, r_{XY} is comprised of the direct path, *e*, as well as three indirect compound paths. The path through W is *ab*, the path through M is *df*, and the path through W and M is *acf*. The rest of the paths are:

$$r_{WM} = c + ad \quad r_{WY} = b + ae + cf + adf \quad r_{MY} = f + cb + de + cae + dab.$$

Now that that there are a set of paths for each correlation, the next step is to substitute in the correlation values and then solve for path values.

$$\begin{aligned} r_{MW} &= .50 = c + ad \\ r_{MX} &= .64 = d + ac & r_{WX} &= .40 = a \\ r_{MY} &= .60 = f + cb & r_{WY} &= .36 = b + ae & r_{XY} &= .59 = e + ab \\ &\quad + de + cae + dab & &\quad + cf + adf & &\quad + df + acf. \end{aligned}$$

Since we know *a* is .40, substitute this value in all the equations. The revised equations for r_{MW} and r_{MX} now only have two unknown values—*c* and *d*—so they can

be solved.

$$r_{MW} = .50 = c + .40d \implies c = .50 - .40d$$
$$r_{MX} = .64 = d + .40c = d + .40(.50 - .40d) = d + .20 - .16d = .84d + .20.$$

Solving for d shows that $d = .44/.84$ or approximately .524. Likewise, solving for c shows that $c = .50 - .40(.524)$ or .290. After substituting these values into the three remaining equations, there are only three unknown values: b, e, and f. The new (simplified) equations are

$$r_{MY} = .60 = f + .50b + .64e$$
$$r_{WY} = .36 = b + .40e + .50f$$
$$r_{XY} = .59 = e + .40b + .64f.$$

Solving these equations shows that $b = .043$, $e = .343$, and $f = .359$.

The only values left to find are for the residual variances of M and Y, paths g and h, respectively. Residual variance paths are more complex than other paths as they represent the amount of variance not explained by the variables' predictors (i.e., $1 - R^2$). For M, only W and X are upstream from it. The amount of variance these two variables explain of M is the sum of the direct variables' paths squared plus the paths that include their correlation, a. The direct paths are c and d, while the paths that include a are cad and dac. The cad and dac paths, of course, are the same values. Thus, the amount of M's variance explained by the model is

$$R^2_M = c^2 + d^2 + 2cad.$$

This makes

$$g = 1 - R^2_M = 1 - (c^2 + d^2 + 2cad).$$

The same logic produces the following equation for h:

$$h = 1 - R^2_Y = 1 - (a^2 + e^2 + f^2 + 2fcb + 2fdb + 2bae + 2badf + 2eacf).$$

Both g and h are now in terms of known values. So, all that is left to do is substitute values and do the arithmetic

$$g = 1 - R^2_M = 1 - .480 = .520$$
$$h = 1 - R^2_Y = 1 - .433 = .567.$$

Now that there are values for all the paths, the relations between any two variables can be decomposed into total, direct, and indirect effects. For example, the direct

effect from X to Y is .343, while the indirect effects are $ab + df + acf$ or $.17 + .188 + .042 = .247$. The sum of these two values is .59, which is the total effect.

Chapter 1 Notes

Data analysis software popularity. R. A. Muenchen provides updated data on the popularity of a variety of data analysis software programs at: r4stats.com/articles/popularity/

Latent variables. Bollen (2002) discusses a number of ways in which latent variables have been defined and distinguished from observed variables.

Cause. Mulaik (1994), Sobel (1987), and Bullock et al. (1995) discuss how this concept is used in causal modeling. An effort to put the notion of cause in SEM on a well-defined and scientifically intelligible basis is represented by the work of Judea Pearl (1998, 2000), discussed in Chapter 7. See also Spirtes et al. (1993, 1998) and Shipley (2000).

Path analysis. An introductory account, somewhat oriented toward genetics, is Li (1975). The statement of Wright's tracing rules in this chapter is adapted from Li's. Kenny (1979) provides another introductory presentation with a slightly different version of the path-tracing rules: A single rule—a variable entered via an arrowhead cannot be left via an arrowhead—covers rules 2 and 3. The sociologist O. D. Duncan (1966) is usually credited with rediscovering path analysis for social scientists; Werts and Linn (1970) wrote a paper calling psychologists' attention to the method. For an annotated bibliography on the history of path analysis, see Wolfle (2003).

Factor analysis. Wolfle (1940) wrote one of the first histories of the field. Maxwell (1977) has a brief account of some of the early history, which Mulaik (1986) updates. Carroll (1993) provides a history as it relates to the study of cognitive ability. See also Hägglund (2001) and many of the chapters in Cudeck and MacCullum (2012). Bartholomew (1995) discusses Spearman's contributions to the development of factor analysis. Individuals at the Thurstone Psychometric Lab created a timeline showing significant publications and events in factor analysis: fa100.info/timeline050504.pdf

The Notes to Chapter 5 list some books on factor analysis and Cudeck (2000) provides an overview. For an explicit distinction between the exploratory and confirmatory varieties, see Jöreskog and Lawley (1968), and for a discussion of some of the differences, McArdle (1996), Nesselroade and Baltes (1984), and Thompson (2004).

Structural equations. These come from econometrics—for some relationships between econometrics and psychometrics, see Goldberger (1971) and a special issue of the journal *Econometrics* edited by de Leeuw et al. (1983). A historical perspective is given by Bentler (1986).

Direct and indirect effects. For a discussion of such effects, and the development of matrix methods for their systematic calculation, see Fox (1980, 1985). See also Sobel (1988). Finch et al. (1997) discuss how sample size and nonnormality affect the estimation of indirect effects.

Under and overdetermination in path diagrams. Often discussed in the structural equation literature as *identification*. More in Chapter 2.

"Recursive" and "nonrecursive." In the technical literature, path models with loops are described as "nonrecursive," and path models without loops as "recursive." Beginning students find this terminology confusing, to say the least. It may help to know that "recursive" refers to the corresponding sets of equations and how they can be solved, rather than describing path diagrams.

Original and standardized variables. Their relative merits are debated by Tukey (1954) and Wright (1960), also see Kim and Ferree (1981) and Alwin (1988). See Bielby (1986), Williams and Thomson (1986), and several commentators for a discussion of some of the hazards involved in scaling latent variables. Yuan and Bentler (2000a) discuss the use of correlation versus covariance matrices in exploratory factor analysis. Again, more on this topic in Chapter 2.

Related topics. Several examples of manifest-variable path and structural analysis may be found in Marsden (1981), especially Part II. Principal component analysis is treated in most factor analysis texts (see Chapter 5); for discussions of relationships between factor analysis and principal component analysis, see an issue of *Multivariate Behavioral Research* (Vol. 25, No. 1, 1990), and Widaman (1993, 2007).

For a broad treatment of structural models that covers both quantitative and qualitative variables see Kiiveri and Speed (1982); for related discussions, see Bartholomew (2002, 2013), and Molenaar and von Eye (1994).

Journal sources. Some journals that frequently publish articles on developments in the area of latent variable models include *The British Journal of Mathematical and Statistical Psychology*, *Educational and Psychological Measurement*, *Journal of Marketing Research*, *Multivariate Behavioral Research*, *Organizational Research Methods*, *Psychological Methods*, *Psychometrika*, *Sociological Methods and Research*, and *Structural Equation Modeling*. See also the annual series *Sociological Methodology*.

Books. Some books dealing with path and structural equation modeling include those written or edited by Arminger et al. (1995), Asher (1983), Bartholomew (2013), Bartholomew et al. (2011), Berkane (1997), Bollen (1989b), Bollen and Long (1993), Brown (2015), Bryne (1998, 2006, 2011, 2016), Cudeck et al. (2001), Cuttance and Ecob (1988), Duncan (1975), Everitt (1984), Hayduk (1987, 1996), Heise (1975), Hoyle (1995, 2012), Keith (2014), Kenny (1979), James et al. (1982), Kaplan (2009), Kline (2015), Long (1983a, 1983b, 1988), Marcoulides and Moustaki (2002),

Marcoulides and Schumacker (1996, 2001), Maruyama (1998), Mueller (1996), Pugasek et al. (2003), Raykov and Marcoulides (2006), Saris and Stronkhorst (1984), Schumacker and Lomax (2015), and von Eye and Clogg (1994).

Annotated bibliography. An extensive annotated bibliography of books, chapters, and articles in the area of structural equation modeling, by J. T. Austin and R. F. Calderón, appeared in the journal *Structural Equation Modeling* (1996, Vol. 3, No. 2, pp. 105-175).

Internet resources. There are many. One good place to start is with a web page called SEMFAQ (*S*tructural *E*quation *M*odeling: *F*requently *A*sked *Q*uestions). It contains brief discussions of SEM issues that often give students difficulty, as well as lists of books and journals, plus links to a variety of other relevant web pages. SEMFAQ's address (at the time of writing) is gsu.edu/~mkteer/semfaq.html. Other useful listings of internet resources for SEM can be found at smallwaters.com, statmodel.com, and davidakenny.net/cm/causalm.htm. A bibliography on SEM is at upa.pdx.edu/IOA/newsom/semrefs.htm.

There is an SEM listserv called SEMNET. Information on how to join this network is given by E. E. Rigdon in the journal *Structural Equation Modeling* (1994, Vol. 1, No. 2, pp. 190-192), or may be obtained via the SEMFAQ page mentioned above. Searchable archives of SEMNET discussions exist.

Chapter 1 Exercises

Note: Answers to most exercises are given at the back of the book, preceding the References. Correlation or covariance matrices required for computer-based exercises are included on the text's web site. There are none in this chapter.

1. Draw a path diagram of the relationships among impulsivity and hostility at one time and delinquency at a later time, assuming that the first two influence the third but not vice versa.

2. Draw a path diagram of the relationships among ability, motivation, and performance, each measured on two occasions.

3. Consider the path diagram of Fig. 1.10 (p. 13). Think of some actual variables A, B, C, and D that might be related in the same way as the hypothetical variables in that figure. (Don't worry about the exact sizes of the correlations.)

Problems 4–10 all refer to Fig. 1.22 (next page).

4. Identify the source and downstream variables.

5. What assumption is made about the causation of variable D?

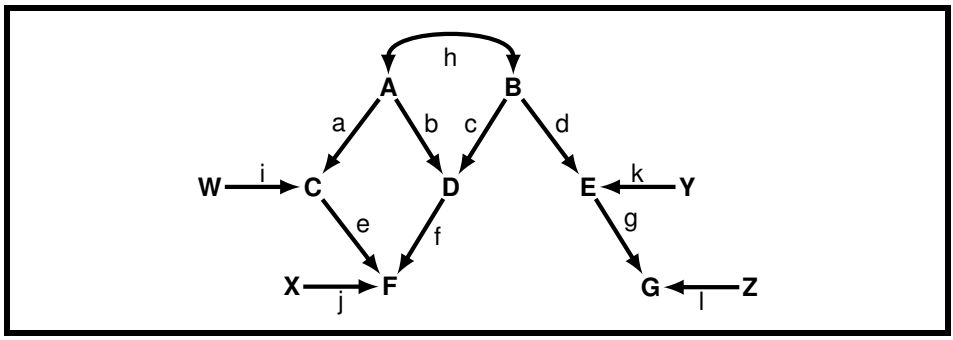


Figure 1.22 Path diagram for problems 4 to 10 (all variables standardized unless otherwise specified).

6. Write path equations for the correlations r_{AF}, r_{DG}, r_{CE}, and r_{EF}.

7. Write path equations for the variances of C, D, and F.

8. If variables A, B, F, and G are measured, and the others latent, would you expect the path diagram to be solvable? (Explain why or why not.)

9. Now, assume that the variables are *not* standardized. Write path equations, using raw-score coefficients, for the covariances cov_{CD}, cov_{FG}, cov_{AG} and the variances s^2_G and s^2_D.

10. Write structural equations for the variables D, E, and F.

11. Redraw Fig. 1.23 as a RAM path diagram. (E and F are latent variables, A through D are observed.)

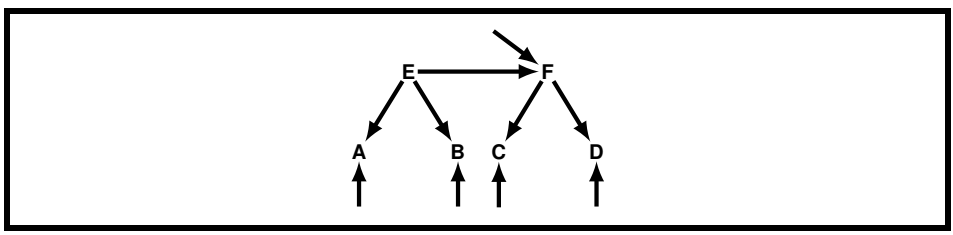
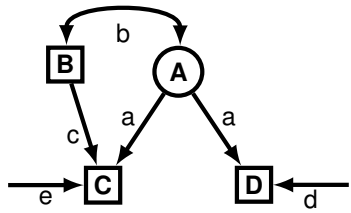


Figure 1.23 Path diagram for problem 11.

12. Given the path diagram shown in Fig. 1.24 and the observed correlations given to the right, solve for *a, b, c, d*, and *e*.



	B	C	D
B	1.00	.70	.30
C		1.00	.48
D			1.00

Figure 1.24 Path diagram for problem 12.

13. The following correlations among three variables are observed:

	A	B	C
A	1.00	.42	.12
B		1.00	.14
C			1.00

Solve for the loadings on a single common factor using the method of triads (see pp. 19–20).