Basic Concepts of Fit

Fit with Path Models

In mediation analysis, it can be shown that the correlation between a predictor and outcome variable is theoretically equal to the product of the correlation between the predictor variable and the mediator times the correlation between the mediator and the outcome (Baron & Kenny, 1987). That is, assuming the mediational model is true, the correlation between y_1 and y_3, depicted below, should be determined by the correlations between the variables connected by direct paths (Alker, 1965; Blalock, 1962; Simon, 1957; see Asher, 1983, for a review).[1]

$$y_1 \xrightarrow{r_{12}} y_2 \xrightarrow{r_{23}} y_3$$

Putting this statement into a formula, we have:

$$r_{13} = (r_{12})(r_{23})$$

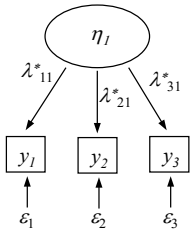
Assume that r_{12} = .3, r_{23} = .4, and r_{13} = .6. Based on the formula above, however, the value of r_{13} that is implied is .12 (i.e., .3 times .4). The data then are not in agreement with the value that for r_{13} that is implied by the model. Because the paths in this model are just simple regressions, we can also think about the correlations on the right side of the equation as standardized path coefficients, where previously we showed that the standardized coefficients could be derived from the correlations among the variables, given sufficient information.

$$r_{13} = (\beta^*_{21})(\beta^*_{32})$$

This logic is the basic idea behind model fit in structural modeling. The correlations that are obtained from the data are compared to those values implied by the model. In essence, then, there is an *implied* correlation matrix. In actual practice, however, we usually use covariances—thus, the obtained and the implied covariance matrices are compared. The hypothesized model is evaluated by "fitting" the covariances obtained to the covariances implied by the model (more on this in a subsequent handout). The basic logic behind the comparative evaluation of causal models can be extended to much more complex models (e.g., Blalock, 1962; Pearl, 1998; Simon, 1957).

Fit with Confirmatory Factor Models

The same general logic can be applied to confirmatory factor models. As we have discussed earlier (see the handout "Latent Variables"), loadings from a factor model can be derived from the correlation matrix (and, hence, variance-covariance matrix), as long as there are as many correlations as unknown loading parameters that need to be estimated. Below, standardized loadings are represented by λ^*_{ij}.



$$r_{12} = \lambda^*_{11}\lambda^*_{21}$$
$$r_{23} = \lambda^*_{21}\lambda^*_{31}$$
$$r_{13} = \lambda^*_{11}\lambda^*_{31}$$

[1] This statement assumes several things including perfectly measured variables, population values, and no other mediating variables.

If there are four or more indicators, then it is possible to compare the correlations among the variables to the correlations implied by the model, because the correlation between any two indicator variables for the same factor should be equal to the product between their respective loadings.

Model Fit and Causality

Back to the mediation model. Assuming just three variables in the mediational model, there are 9 possible theoretical models that can be drawn.[2] By drawing out these models, computing implied correlations, and comparing them to obtained correlations, one discovers several interesting things.

For this example assume that r_{AB} = .3, r_{BC} = .4, and r_{AC} = .12.

Theoretical Model	Implied Correlation between Predictor and Outcome	Obtained Correlation	Does It Fit?
A → B → C	.12	.12	yes
A → C → B	.048	.3	no
B → A → C	.036	.4	no
C → B → A	.12	.12	yes
B → C → A	.036	.3	no
C → A → B	.048	.4	no
A → B, A → C	.036	.4	no
B → A, B → C	.12	.12	yes
C → A, C → B	.048	.3	no

There are several things to notice. First, some models can be rejected as being inconsistent with the model. Second, not all models can be eliminated, because some fit equally well. Third, one can think about some conditions in which models could be ruled out because the model includes variables that are unlikely to be influenced by other factors (e.g., race, gender, experimental variable).

Model fit then has implications for causal inference in so far as model fitting can assist in rejecting alternative causal models (Blalock, 1962), even when those models are all based on cross-sectional data. This is an important advantage of structural equation modeling. In actual practice, however,

[2] This example is based on one by Steve West.

several things must be kept in mind. These examples assume there are no other (fourth) spurious variables omitted that might account for the relationships obtained. With sample data, there is always sampling variability in estimating covariances (or correlations), and, thus, even for correct models, there is likely to be some lack of fit between what is implied and what is obtained. It is important to note that the superior fit of one model over another does not *prove* causation—it is some additional evidence in favor of a causal hypothesis. Prudently cautious language about causation is still needed.

References

Alker, H. R., Jr. (1965). *Mathematics and Politics*. New York: Macmillan.
Asher, H. B. (1983). Causal modeling. 2nd edit. Beverly Hills: Sage.
Blalock, H. M. (1962). Four-variable causal models and partial correlations. *American Journal of Sociology, 68*(2), 182-194.
Pearl, J. (1998). Graphs, causality, and structural equation models. *Sociological Methods & Research, 27*(2), 226-284.
Simon, H. (1957). Spurious correlations: a causal interpretation, In H.A. Simon, *Models of Man* (pp. 37–49). New York: John Wiley.