Basics of Path Analysis

Wright's Rules of Tracing and the Fundamentals of Path Analysis

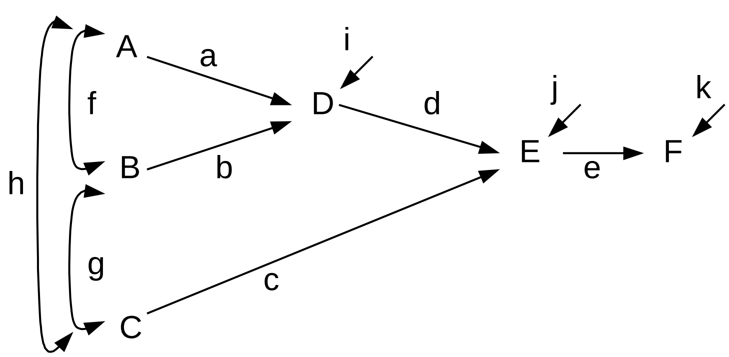
Sewall Wright (1918, 1934) developed a method of estimating causal path coefficients by decomposing the correlations among a set of variables. He articulated a set of rules for examining a path diagram that would allow for this mathematical decomposition.[1] The correlation of any two variables in a path diagram can be expressed as the sum of coefficients that connect the two variables. The connection between one variable and another variable, then, can often be made through more than one route.

1) No loops are allowed. In tracing from one variable to another, you cannot pass through the same variable twice following a particular route.

2) No going forward and then backward. Once you have traveled along a route forward, you cannot travel backward to get to the variable at the end point. (Only use common causes to account for a correlation between two variables, not common outcomes).

3) Only one curved arrow is allowed in tracing from the first variable to the last variable in any route.

Tracing Example[2]



In the above diagram, lower case letters represent values of the standardized coefficients. For curved arrows, they represent correlations; for straight arrows, they represent regression coefficients; and for the short arrows, they represent the value of the error or disturbance.

By using Wright's tracing rules, one can obtain the value of the correlation between any two variables, by adding up the values of the coefficients that link them--provided the rules are followed. Every route between the two variables (acceptable according to the rules) is traced. The coefficients that make up each route are multiplied. If there are multiple routes that link the two variables, products for each route are added together.

- To find the correlation between *B* and *D* (or r_{BD}), two routes are possible: *b* and *fa*. So, the value of r_{BD} is equal to $b + fa$. If $b = .3$, $f = .2$, and $a = .4$,

[1] Note that these rules are often worded differently by different authors and I provide a paraphrased and reorganized version. Also, the RAM style of depicting path models used by the Kline text in most instances notation may be confusing at times (e.g., variables with circular arrows), but we will discuss the differences from the above path model depiction at a future date.
[2] This example is adapted from Loehlin, J.C. (1992). *Latent variable models* (2nd Ed.). Hillsdale, NJ: Erlbaum.

r_{BD} = .3 + (.2)(.4) = .38.

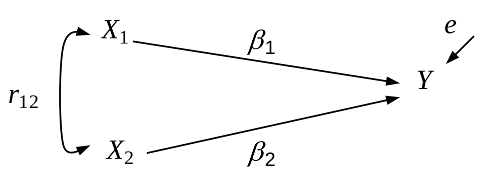
- r_{CD} = *gb* + *ha*.
- r_{AB} is simply equal to f, because the curved arrow represents a correlation.
- r_{AE} = *ad* + *fbd* + *hc*

Some examples of rule violations:
- $r_{CD} \neq cd$ (no forward and back—no common effects, only common causes)
- $r_{AC} \neq fg$ (no two double head arrows)
- $r_{AB} \neq abf$ (no loops)

Decomposing the Correlation Matrix

Although this seems like just a fun little game, it turns out to be immensely useful. Instead of computing the correlations between two variables, one can work backward from the correlations to derive the path coefficients. Taking a simple path diagram representing a two variable regression model, and assuming some values for the correlations between our three variables, we can derive the path coefficients.

Assume that r_{12} = .50, r_{1Y} =.65, and r_{2Y} = .70.



From our tracing rules, we know $r_{1Y} = \beta_1 + r_{12}\beta_2$ and $r_{2Y} = \beta_2 + r_{12}\beta_1$. We can plug in the known values and solve for β_1 and β_2.

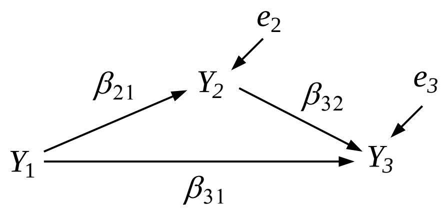
$$.65 = \beta_1 + .50\beta_2$$
$$.70 = \beta_2 + .50\beta_1$$

by rearranging, substituting, and solving for β_1 and then β_2, we get β_1 = .4 and β_2 = .5. These coefficients are the standardized regression coefficients, because we started with a correlation matrix (i.e., standardized variables).

The disturbance term, e, is the amount of unaccounted for variance in Y and is equal to $\sqrt{1-R^2}$. We know from regression analysis that $R^2 = \beta_1^2 + \beta_2^2 + \beta_1\beta_2 r_{12}$. So, $R^2 = .4^2 + .5^2 + .5(.4)(.5) = .61$. The disturbance term then equals $e = \sqrt{1-.61} = .62$ What we have just done is decompose the correlation matrix into unique values for the coefficients that are implied by the model we specified.

Matrices

Let's return to a path diagram, using a simple example with a single indirect path. The figure below uses all Y variables to keep the notation simple and each path in labeled as a standardized path with two subscripts to keep them separate. The first subscript is the "effect," or dependent variable in the regression, and the second subscript is the "cause," or the independent variable in the regression. For example, for coefficient β_{32}, the order of the subscripts is because Y_3 is predicted by Y_2.



We can organize the all of the path coefficients into a (lower triangle) correlation matrix, each decomposed into the paths that make up each coefficient. The rows and the columns for the matrix correspond to each of the three variables, Y_1, Y_2, and Y_3.

$$\begin{array}{c c} & \begin{array}{ccc} Y_1 & \quad Y_2 & \quad Y_3 \end{array} \\ \begin{array}{c} Y_1 \\ Y_2 \\ Y_3 \end{array} & \begin{bmatrix} & & \\ \beta_{21} & & \\ \beta_{31}+\beta_{21}\beta_{32} & \beta_{32} & \end{bmatrix} \end{array}$$

So, the correlation between Y_1 and Y_2 is r_{12} and that is equal to the path β_{21}, a simple direct effect. Correlation r_{13}, however, is decomposed into the two possible paths, β_{31} and $\beta_{21}\beta_{32}$.

Covariances

I present path analysis in terms of correlations and standardized path coefficients, but it is not necessary to use correlations and standardized values, only simpler. In fact, in actual practice, estimating structural models using covariances instead of correlations is standard (and better practice).

The variances need to be taken into account, however, for each time the predictor variable is used. For example, the covariance for Y_1 with Y_3, or $\mathrm{Cov}(Y_1,Y_3)$, for the diagram above, is $\mathrm{Cov}(Y_1,Y_3) = Var(Y_1)\cdot b_{31} + Var(Y_1)\cdot b_{21}\cdot b_{32}$, where b is used for the unstandardized regression coefficients.

References and Further Reading

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40*, 37-47.
Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology*, *17*, 37-69.
Cheong, J., & MacKinnon, D.P. (2012). Mediation/indirect effects in structural equation modeling. In R.H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 417-435). New York: Guilford.
Duncan, O. D. (1966). Path analysis: Sociological examples. *American Journal of Sociology, 72*, 1-16.
Loehlin, J. C., & Beaujean, A. A. (2017). *Latent variable models: An introduction to factor, path, and structural equation analysis (5th ed.)*. New York, NY: Routledge
Mulaik, S. A. (2009). *Linear causal modeling with structural equations*. CRC press.
Wolfle, L. M. (1999). Sewall Wright on the method of path coefficients: An annotated bibliography. *Structural Equation Modeling: An Multidisciplinary Journal, 6*, 280-291.
Wolfle, L.M. (2003) The Introduction of Path Analysis to the Social Sciences, and Some Emergent Themes: An Annotated Bibliography, *Structural Equation Modeling: A Multidisciplinary Journal, 10*, 1-34
Wright, S. (1918). On the Nature of Size Factors. *Genetics, 3*, 367-74.
Wright, S. (1934). The Method of Path Coefficients. *Annals of Mathematical Statistics, 5,* 161-215.