

GENERAL PRINCIPLES OF PATH ANALYSIS

Structural Equation Modeling (SEM) combines factor analysis and regression. Path analysis is part of the larger concept of SEM, its specificity lies in the absence of latent variables. Our analysis did not involve latent variables so we chose to use path analysis in this article. This is a multivariate method that is associated with a graphical display that necessitates a model *a priori* that is subsequently refined. Thanks to this technique it is possible to study mediating factors and to distinguish direct, indirect and total effects. This approach is more mechanistic than the more classical approach that uses regression and appears particularly adequate when one studies biological mechanisms. (Loehlin, 2004)

I. Keys to read a path model

1.1. Causality and correlation

In **figure 1** causal pathways are depicted using unidirectional arrows. Bidirectional (curved) arrows also exist in path diagrams. They represent the correlation between two variables without any causal hypothesis. Error terms are associated with all the variables that have at least one arrow pointing toward them and represent the part of the variance that is not explained by these variables (i.e. measurement error along with the effects of variables not measured in the study). In order to simplify the diagram, these residual variances are often not displayed but they are included in all the computations.

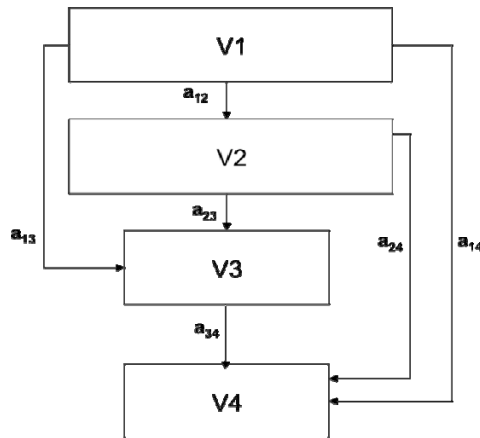


Figure 1. Example of a path model

1.2. Direct, indirect and total effects

The **total effect** of variable 1 on variable 4 corresponds to all the valid paths that link variable 1 to variable 4. The validity of a path is determined by rules edicted by Sewall Wright at the beginning of the last century.

Wright Rules (Loehlin, 2004)-

- 1) No loops are allowed. In tracing from one variable to another, you cannot pass through the same variable twice.
- 2) No going forward and then backward. Once you have travelled along a path forward, you cannot travel backward across the path. However, going backward and then forward is possible.
- 3) Only one bidirectional curved arrow is allowed in tracing from the first variable to the last variable.

The **direct effect** of one variable on another is depicted by the arrow that links the two variables.

The **indirect effect** of a variable on another is mediated by one or more variables.

SUPPLEMENTARY DATA

2. Path coefficients, test statistic and standardization

In **figure 1**, coefficients a_{12} to a_{34} represent the coefficients attached to each path. The indirect effect of variable 1 on variable 4 is the multiplication of the coefficients of the mediating paths. The total effect of variable 1 on variable 4 corresponds to the sum of all valid paths between variable 1 and variable 4 that is the direct and the indirect paths.

In addition to the value itself, the statistical significance of the coefficients can be tested by comparing these parameters to the null thanks to Wald tests.

Coefficients can be standardized or non standardized. When standardized, they can be interpreted as partial correlation coefficients (all other variables being held constant) and the relative importance of the correlations can be compared (similar unit for all coefficients)

2.3. Conditions for a path analysis

- Sample size should be at least 100 to 200 subjects.
- The distribution of the variables must be normal as all variables in the model may be both dependent and independent variables.
- The path model must be « identified », that is the number of estimated parameters should not exceed $n(n+1)/2$, number of observed parameters of the variance-covariance matrix of the n variables of the model.

III. USING PATH ANALYSIS

The first step is to specify the model. In this step, previous findings coming from the literature (epidemiologic or experimental) and/or from previous analyses are transposed in a series of structural equations and in a figure that represents all the variables and their potential interrelations. The fit of this model to the data is then tested. To assess model fit, the hierarchical chi-square test, the goodness of fit index (GFI) and the normal fitted index (NFI) can be used. A significant chi-square test indicates a difference between the observed and the implied correlation matrix from the set of linear equations that is too great to be likely to result from sampling error. A GFI and NFI close to 1 (in practice, higher than 0.95) indicate a model that fits the data. If the adequation of the model is good enough it is important to compare it to other possible models. When two models are nested it is possible to compare the values of the Hierarchical Chi² test. Once the most adequate model is selected, its adequation or parcimony can be improved by adding more paths or suppressing non significant paths respectively. A coefficient, a standard deviation and the accompanying statistical test is attached to each path. This final model provides direct, indirect and total effects of the variables of interest.

In SAS 9.2, it is also possible to carry out a multigroup analysis. The aim is then to test if a single model can be used in two groups, for example in boys and girls. If not, there is an interaction and the path coefficients must be computed separately in the two different groups. If a single model is applicable in the two groups, it is still possible to test potential significant differences between the coefficients of a given (a combination of) path(s) in the two groups.

Loehlin, J. C. (2004). Latent variable models: An introduction to factor, path, and structural analysis. Hillsdale, NJ, Erlbaum.