# Notes on Latent Variable Modeling

compiled by D.Gueorguiev 3/31/2024

## Introductory Notes

The Ideal experiment: a single independent variable is manipulated, and the consequences are observed in a single dependent variable.

Experiments in reality : the variables which are observed are typically not the one of real theoretical interest but are merely some convenient variables acting as proxies. A full analysis will turn out to be multivariate, with a number of alternative experimental manipulators on the one side, and a number of alternative response measures on the other.

There is a variety of statistical techniques for dealing with situations in which multiple variables, some of which unobserved, are involved. In [1] are discussed a variety of methods with the following common features:

( a ) multiple variables – three or more - are involved

( b ) one or more of these variables are unobserved i.e. *latent*

*Latent variable analysis*, discussed in [1], encompasses specific methods such as factor analysis, path analysis and structural equation modeling applied to (a) and (b).

### Path Models in Factor, Path, and Structural Equation Analysis

#### Path Diagrams

*Path diagram* is a representation of the relationships among a number of variables. We use capital letters to denote variables in such diagram. The connection among variables are represented in path diagrams by two kinds of arrows : a straight, directed arrow represents *a causal relationship* between two variables, while a curved bidirectional arrow represents *a correlation* between the variables which it connects.

Figure 1: example of a path diagram, Modeling intelligence in single child family

Variables , , and are all assumed to have causal effects on . Variables and are assumed to be correlated with each other. Variable is assumed to affect but to be uncorrelated with either or .

What the diagram on Figure 1 might model?

Possible representation of the diagram on Figure 1: Modeling intelligence in single child family

would represent the child intelligence, and would represent father’s and mother’s intelligence which are assumed to have causal influence on child intelligence. The intelligence of the mother and father is correlated. represents the other variables, independent on the father and mother intelligence, which influence child’s intelligence.

Figure 2 shows another path diagram. Here is assumed to affect both and , and each of the latter variables is also affected by an additional variable – U and V respectively.

Possible representation of the diagram on Figure 2: reliability of psychometric test

and would represent scores on two alternate forms of a test. 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 a) to each form of the test or b) to the occasions on which it was administered. These additional factors would affect any given performance, but it would be unrelated to the true trait.

*Note*: the variance in and resulting from the influence of would be called *true score variance* and that caused by or would be called *error variance*. The proportion of the variance of or due to would be called the *reliability of the test*.

Figure 2: Reliability of psychometric test

Figure 3: path diagram involving events over time

## References

[1] [Latent Variable Models: Introduction to Factor Analysis and Structural Equation Analysis, John C. Loehlin, 2004](https://github.com/dimitarpg13/information_theory_and_statistical_mechanics/blob/main/literature/books/Latent_Variable_Models_Loehlin_2004.pdf)