Specifying Covariance Models using Path Diagrams

Will Gertsch

April 28, 2021

Introduction

A path diagram is a graphical approach for specifying covariances.

- Commonly used in psychology and other social sciences.
- Subject expert can easily draw relationships between variables.
- Can handle mediation and multiple outcomes.
- Goal: specify a model for the population covariance matrix Σ.

Basic Ideas

Let's start with a simple model.

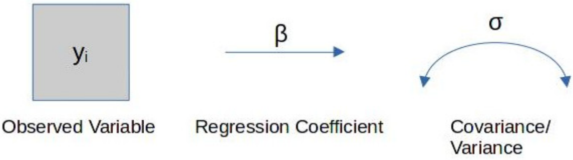
- Suppose we have the linear model

$$y_i = \beta_0 + \beta_1 x_{1i} + e_i$$

  where $e_i \sim N(0, \sigma^2)$ for independent observations $i = 1, \ldots, n$.
- We will recreate this regression with a path diagram to demonstrate the basics of path analysis.

Path Diagram Notation

Path diagrams use the following notation



- Labeled squares are observed variables.
- Symbols attached to arrows are population parameters.
- Single-headed arrows are regression coefficients.
- Regression coefficients are commonly called “paths”.
- Curved double-headed arrows are covariances or variances.

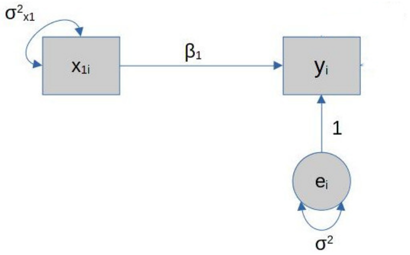
Simple Model

This diagram specifies a covariance model for variables x_1, y on subject i.

$$\Sigma = \mathrm{Var}\begin{pmatrix} x_{1i} \\ y_i \end{pmatrix} = \begin{pmatrix} \sigma^2_{x1} & 0 \\ 0 & \sigma^2_y \end{pmatrix}$$

- σ^2_{x1}, σ^2_y are variances of x_{1i} and y_i.
- There are no paths or covariances between x_{1i} and y_i.
- Therefore the covariance of x_{1i} and y_i is 0.

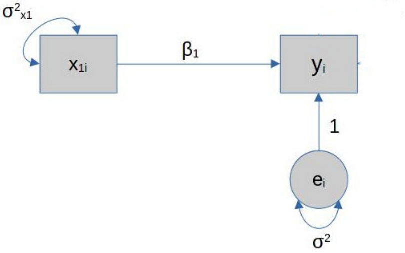
No Intercept Model



This model adds a regression path between x_1 and y with parameter β_1.

- The added path creates the regression equation
  $y_i = \beta_1 x_{1i} + 1 * e_i$,
  $e_i \sim N(0, \sigma^2)$.
- e_i is the error term of the regression.
- Circle notation: e_i is unobserved.
- Equation derived by looking at all paths into y_i.

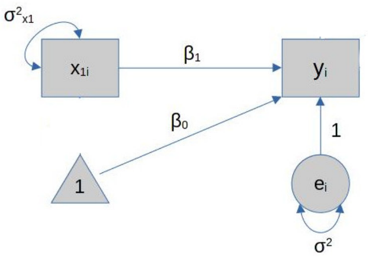
No Intercept Model



- x_{1i} and e_i assumed independent because no path connects them.
- Unobserved variables like e_i are not included in Σ.
- The covariance matrix is only for observed variables.
- $\text{Cov}(y_i, x_{1i}) = \text{Cov}(\beta_1 x_{1i} + e_i, x_{1i}) = \beta_1 \sigma^2_{x1}$

$$\Sigma = \text{Var}\begin{pmatrix} x_{1i} \\ y_i \end{pmatrix} = \begin{pmatrix} \sigma^2_{x1} & \beta_1 \sigma^2_{x1} \\ \beta_1 \sigma^2_{x1} & \sigma^2 \end{pmatrix}$$

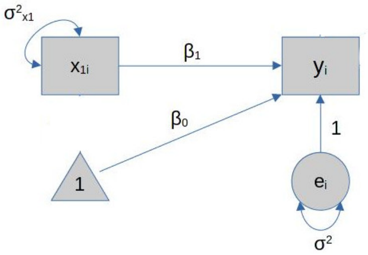
Simple Linear Regression



We add the intercept β_0 to the previous model.

- Triangle: a constant term for all subjects.
- Constant is usually set to 1.
- A path from a triangle will add an intercept.
- $y_i = \beta_0 + \beta_1 x_{1i} + e_i$,
  $e_i \sim N(0, \sigma^2)$

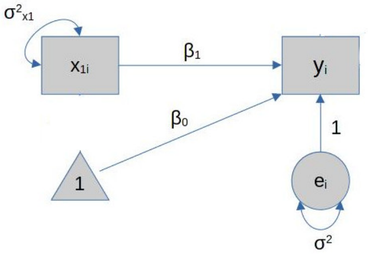
Simple Linear Regression



What is the covariance of x_{1i} and y_i?

- $\text{Cov}(x_{1_i}, y_i) = \beta_1 \sigma^2_{x1}$
- Is this the same as $\beta_1 \sigma^2_{x1}$ for the no-intercept model?
- No: β_1 is different between the two models.

Simple Linear Regression



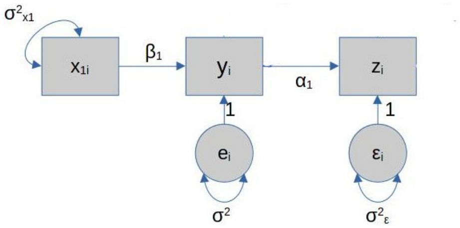
Let $r_{x1,y}$ be the population correlation between x_{1i} and y_i.

$$r_{x1,y} = \frac{\mathrm{Cov}(x_{1i}, y_i)}{\sqrt{\sigma^2_{x1}}\sqrt{\sigma^2_y}}$$

β_1 for this model is

$$\beta_1 = r_{x1,y}\frac{\sigma_y}{\sigma_{x1}}$$

Mediation



Path analysis can accommodate more than 1 regression equation.

- This mediation model specifies 2 regression equations.
- $y_i = \beta_1 x_{1i} + e_i$
  $z_i = \alpha_1 y_i + \epsilon_i,$
  $e_i \sim N(0, \sigma^2),$
  $\epsilon_i \sim N(0, \sigma_\epsilon^2)$
- We can show $\text{Cov}(x_{1i}, z_i) = \alpha_1 \beta_1 \sigma_{x1}^2$

Estimation and Inference

To fit covariance models, we estimate Σ.

- One popular way to find the estimate $\hat{\Sigma}$ is to minimize

$$F_{ML} = \log|\hat{\Sigma}| + \mathrm{tr}(S\hat{\Sigma}^{-1}) - \log|S| - k$$

- S is the observed covariance matrix, k is the number of observed variables, log is the natural logarithm, $\mathrm{tr}(\cdot)$, $|\cdot|$ are the trace and determinant.
- Assuming multivariate normality, $(n-1)F_{ML}$ has a chi-square distribution with degrees of freedom equal to the number of parameters estimated.
- Other estimators exist for non-normal data, missing data, etc.

Conclusion

- We drew path diagrams for simple 2 variable models.
- It is easy to draw more complicated diagrams for bigger models.
- F_{ML} provides an easy way to fit many of these models.

References

- Wang, Jichuan, and Xiaoqian Wang. 2012. Structural Equation Modeling with Mplus : Methods and Applications. Wiley Series in Probability and Statistics. Chichester, West Sussex: Wiley/Higher Education Press.