Structural Equation Modeling

Dr Wan Nor Arifin

Unit of Biostatistics and Research Methodology,
Universiti Sains Malaysia.
E-mail: wnarifin@usm.my

Contents

1 Objectives

1. Understand and apply the basic knowledge of SEM.
2. Specify SEM models involving correlation, causal, mediation & moderation, and interpret the results.

2 Introduction

- Structural Equation Modeling (SEM) is a mutivariate statistical modeling that aims to explain the structure of relationships among multiple variables (Hair et al., 2010a)
- Needs strong theoretical specification of the model ahead of the analysis → to verify our theory on the relationships.
- SEM basically consists of two components (Bartholomew et al., 2008):
    1. measurement model (CFA): dealing with latent variables (factors) and the relationships between the items and the factors.
    2. structural model (path analysis): dealing with how latent variables are related to each other.
- Similar to CFA, how the variance-covariance matrix produced from the model fits the variance-covariance matrix of the observed data → Goodness of fit of model to the data.

3 Commonly used terms in SEM

- **Exogenous variable** Independent, predictor variable. Could be observed or latent variables.
- **Endogenous variable** Dependent, outcome variable. Also could be observed or latent variables.
- **Path diagram** A visual representation of the SEM model.

4 Path diagram

- **Latent variable** Circle, oval.
- **Observed variable** Square, rectangles.
- **Bidirectional arrow** Correlation, covariance.
- **Unidirectional arrow** Causal relationship. From independent to dependent variables.

5 Relationships in SEM

5.1 Correlation

- Bidirectional correlation/covariance between variables (observed/latent).

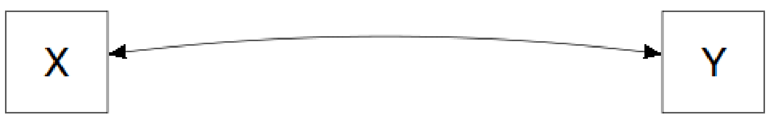

Figure 1: Between observed variables.

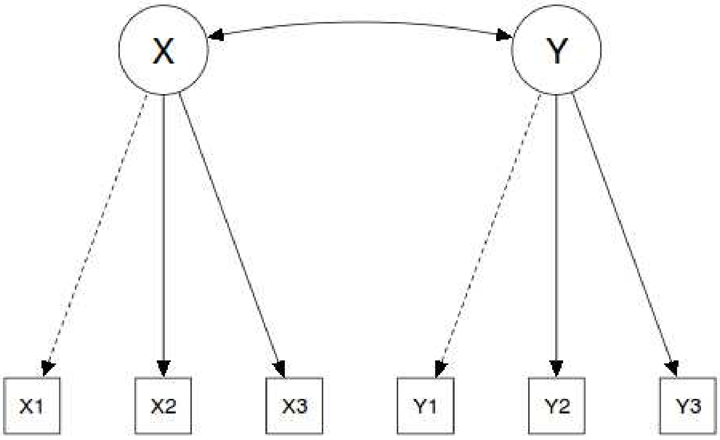

Figure 2: Between latent variables.

5.2 Causal

- Causal relationship between variables (observed/latent). Assign dependent and independent variables.

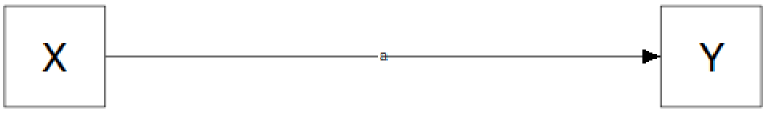

Figure 3: Endogenous/dependent Y caused by exogenous/independent X (observed).

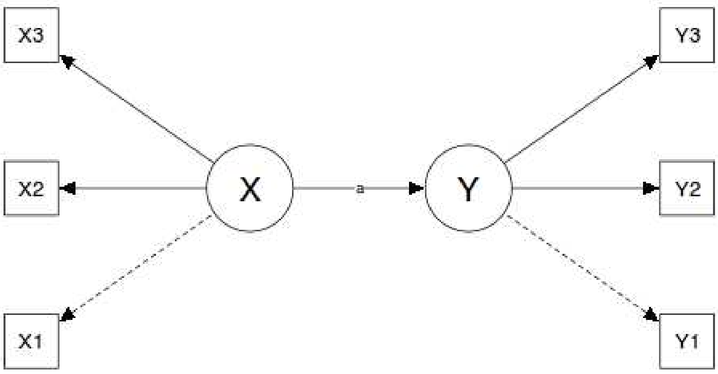

Figure 4: Endogenous/dependent Y caused by exogenous/independent X (latent).

5.3 Mediation

- Figure 3 shows direct causal effect from independent X to dependent Y. Let say the regression coefficient is significant.

- In SEM it is also possible to examine the effect of a mediating variable Z on this established relationship.
- A mediating variable Z is a third variable that intervenes between two related variables (Hair et al., 2010a). It plays two roles, as an DV in one equation and IV in another as it mediates between X and Y (Awang, 2012).
- If relationship between X and Y is mediated by Z, it indicates <u>indirect</u> causal effect of X to Y (X $\rightarrow$ Z $\rightarrow$ Y).
- The role of Z is clearly shown in Figure 5 below:

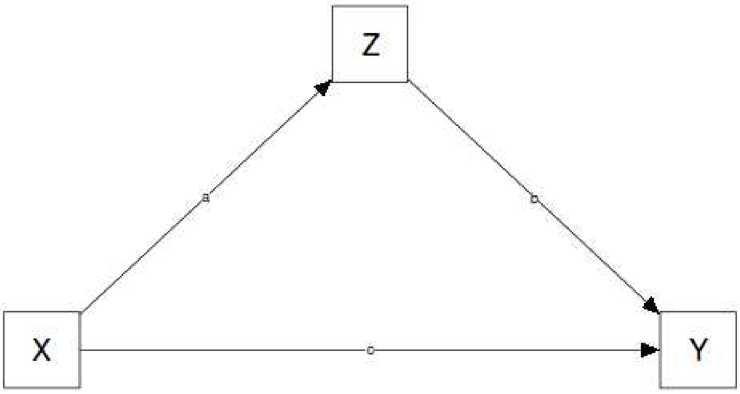

Figure 5: Z mediates relationship of X to Y.

- Mediation effect can be classified into three outcomes (Awang, 2012):

  1. Complete mediation.
     - Z completely mediates the effect of X on Y; X indirectly causes Y.
     - Conditions:
       (a) a & b regression coefficients are significant.
       (b) c is not significant, i.e. c becomes insignificant once Z is included in the model.
  2. Partial mediation.
     - Z partially mediates the effect of X on Y; X directly and indirectly causes Y.
     - Conditions:
       (a) All a, b and c regression coefficients are significant.
       (b) c (with Z) $<$ c (without Z) (Hair et al., 2010a).
  3. No mediation.
     - Z does not mediate the effect of X on Y; X directly causes Y.
     - Conditions:
       (a) c regression coefficient is significant.
       (b) a or b, or both a and b are not significant. This indicates the mediation path through a and b is interrupted.

5.4 Moderation

- A moderating variable M is a third variable that changes the relationship between two related variables X and Y (Hair et al., 2010a). It moderates the causal effect that an IV has on a DV (Awang, 2012).
- In medical statistics, this is commonly known as an interaction or effect modifier.

- In SEM, moderation/interraction is specified as shown in Figure 6:

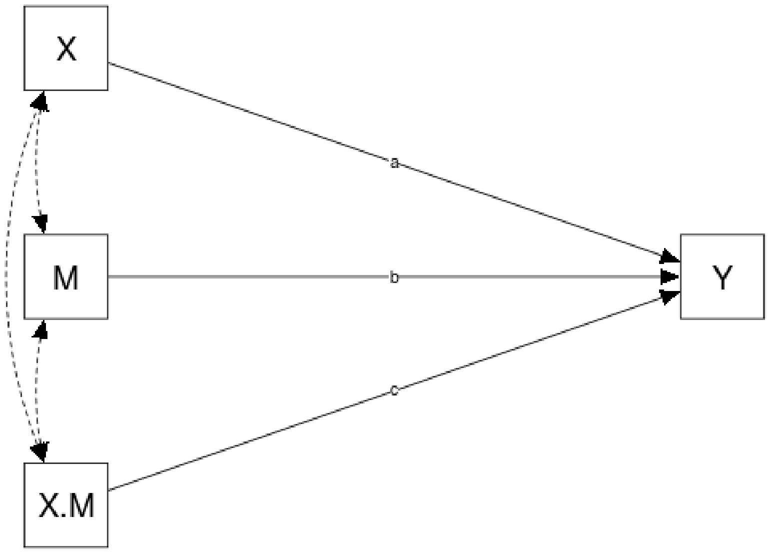

Figure 6: Specifying moderation in SEM.

- Moderation effect can be classified into three outcomes (Awang, 2012):
  1. Complete moderation.
     - M completely moderates the effect of X on Y.
     - Conditions:
       (a) c (regression coefficient of the interaction term X*M) is significant.
       (b) a is not significant, i.e. a becomes insignificant once M is included in the model.
  2. Partial moderation.
     - M partially moderates the effect of X on Y.
     - Conditions:
       (a) a and c regression coefficients are significant, i.e. both the main effect and interaction terms.
       (b) a (with X*M) $<$ a (without X*M).
  3. No moderation.
     - M does not moderate the effect of X on Y.
     - Conditions:
       (a) c is not significant.
       (b) a (with X*M) $=$ a (without X*M).
- However, the model specification in Figure 6 can be applied to observed variables only.
- For latent variables, it is complicated and requires multi-group CFA for categorical moderator (Hair et al. 2010b; Awang, 2012) or adding a latent variable consisting of interaction items (e.g. items in latent X×M, items in latent X×items in latent M) in the model for numerical moderator (Hair et al., 2010b).

References

Awang, Z. (2012). *Structural equation modeling using Amos Graphic.* Selangor, Malaysia: UiTM Press.

Bartholomew, D. J., Steele, F., Moustaki, I., and Galbraith, J. I. (2008). *Analysis of multivariate social science data.* USA: CRC Press.

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2010a). *Multivariate data analysis.* New Jersey: Prentice Hall.

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2010b). Sem basics: A supplement to multivariate data analysis.