Advanced Experimental Design
Psych 464
Jeffrey D. Leitzel, Ph.D.
Path Analysis

1

Path Analysis as "causal models"

- We talk about path models as "causal models"
- Theory should always be driving model building
- Tells us whether or not the causal model is consistent with the data
- If inconsistent with the data, doubt the theory
  - Requires that the study was validly designed
- Consistency of model with data is NOT proof of a theory
  - may lend support to the theory
  - philosophers of science - our theory simply "survived the test" and was not disconfirmed

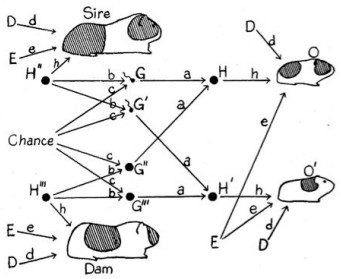
2

The problem of competing causal models

- 1. X->Y->Z
- vs.
- 2. X<-Y->Z
- Both models may be consistent with data
- Nothing in the data that would have us prefer one model over the other

3

- Wright, S. (1920). The Relative Importance of Heredity and Environment in Determining the Piebald Pattern of Guinea-Pigs. *Proceedings of the National Academy of Sciences, 6*(6), 320-332.



Sewell Wright developed path analysis around 1918

4

- A method for studying direct and indirect effects of variables hypothesized as causes of variables treated as effects.
- "The method of path coefficients is not intended to accomplish the impossible task of deducing causal relations from the values of the correlation coefficients. It is intended to combine the quantitative information given by the correlations with such a qualitative information as may be at hand on causal relations to give a quantitative interpretation." Wright 1934, p 193

Wright emphasized theory

Wright, S. (1934). The Method of Path Coefficients. *The Annals of Mathematical Statistics*, 5(3), 161-215.

5

- Path diagram useful for displaying the hypothesized pattern of causal relationships
- Types of variables
  - Exogenous variable is one whose variability is determined by causes outside the model
  - Endogenous variable is one whose variability is explained by the model
- Exogenous variables – independent variables, arrows only exit from, never point to
- Endogenous variables – dependent variables, variables that our path model seeks to explain
- An example similar to the one we worked with in class already

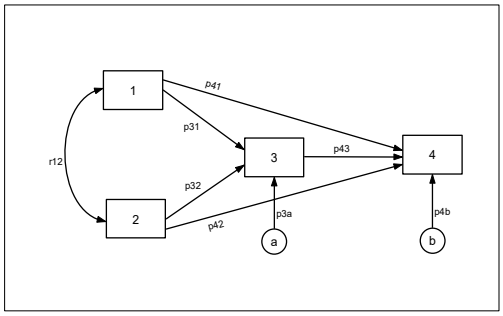
Path models

6

An example path model

- Which variables are exogenous in this diagram?
- Which are endogenous?

Adapted from Pedhazur, E. J. (1997). *Multiple Regression in Behavioral Research: Explanation and Prediction (3rd ed.). New York: Harcourt Brace College Publishers.*



7

Model estimation parameters, observations, and number of cases

- Number of parameters we can estimate is a function of the number of variables we have data for in our model
- n parameters estimated ≤ n observations, which is based on n variables
- $N_{obs} = [k(k+1)]/2$
- How many observations in the dataset we just looked at?
- Parameters we are estimating
  - path coefficients
  - disturbance (error) terms
  - covariances between and variances of exogenous variables
- Path coefficients = β coefficients
- How many parameters are we estimating in model on last slide?
- We will only be dealing with recursive models at this point
- Number of cases recommendations

8

Assumptions of path analysis

1. Relationships are linear, additive, and causal
2. Each residual is uncorrelated with variables that precede it in the model. Endogenous variable is conceptually a linear combination of endogenous & exogenous variables plus a residual. All relevant variables are in the model.
3. One way causal flow – no reciprocal causation
4. Variables are measured on at least an interval scale
5. Variables are measured without error

- Clearly assumptions that all relevant variables are in the model and that measures are error-free are untenable
- Path analysis essentially reduces to solving one or more linear regression analyses.

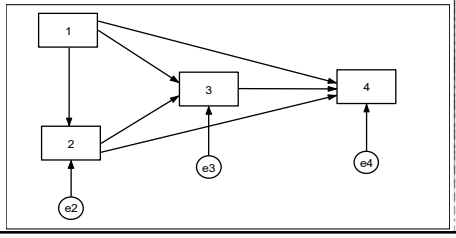
9

Another example path model

- Expressing model equations in z-score form

1. $z_1 = e_1$
2. $z_2 = p_{21}z_1 + e_2$
3. $z_3 = p_{31}z_1 + p_{32}z_2 + e_3$
4. $z_4 = p_{41}z_1 + p_{42}z_2 + p_{43}z_3 + e_4$

Adapted from Pedhazur, E. J. (1997). *Multiple Regression in Behavioral Research: Explanation and Prediction (3rd ed.). New York: Harcourt Brace College Publishers.*



10

Recursive system parameter estimation constraints

- At least half of the path coefficients = 0
- 1. $z_1 = e_1$ implies:
- $z_1 = e_1 + 0_{12}z_2 + 0_{13}z_3 + 0_{14}z_4$
- same would go for the other equations as well
- path coefficients of 0 are implied
- also assumed - none of the residuals are correlated

11

Relationship between b and β coefficients and introducing the covariance

- $\beta = \frac{\sum z_x z_y}{\sum z_x^2}$
- $b = \frac{\sum xy}{\sum x^2}$
- Where: Σxy = $\sum(x-\overline{x})(y-\overline{y})$ or $\sum xy - \frac{\sum x \sum y}{n}$
- $\beta = b\frac{s_x}{s_y}$
- $b = \beta\frac{s_y}{s_x}$
- The sample covariance (unstandardized correlation coefficient) is:
- $S_{xy} = \frac{\sum xy}{n-1}$

12