Introduction to Path Analysis

- Review of Multivariate research & an Additional model
- Structure of Regression Models & Path Models
- Direct & Indirect effects
- Mediation analyses
- "When" & some words of caution
- Where path coefficients come from
- Some ways to improve a path analysis model

Why multivariate research designs? ➔ Multicausality

Multicausality is the idea that behavior has multiple causes, and so, can be better studied using multivariate research designs !!!

The fundamental questions about multicausality that are asked in multivariate research…

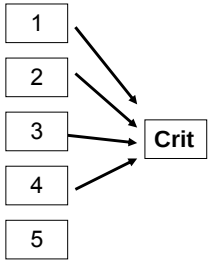
1. Interactions
   - does the effect of an IV upon the DV depend upon the value of a 2nd IV?
   - Studied using Factorial Designs
2. Unique contributions
   - Does an IV tell us something about a DV that other IVs don't?
   - Studied using Multiple Regression
3. Causal Structures
   - Is a DV an IV for another DV?
   - Behaviors are effects of some things and causes of others
   - Structural Modeling & Path Analysis

Here is the "structure" of a multiple regression model…

- 5 predictors
- 1 criterion

This structure shows the RH:

- of these 5 predictors, only 4 of them are hypothesized to make a unique contribution to understanding the criterion
- leaving a "path" out hypothesizes the predictor doesn't contribute to the model



In a multiple regression model, the collinearity (correlation) among the predictors it taken into account, to help us identify which variables have a unique contribution to understanding the criterion.

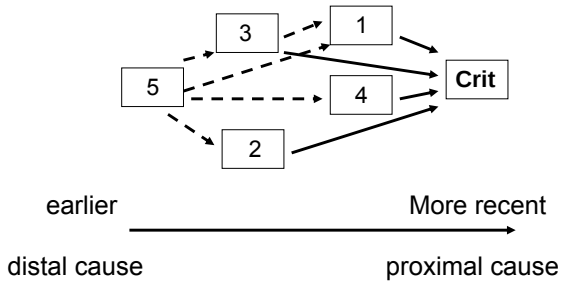
But we don't learn about how the predictors relate to each other!!!

Here is the "structure" of the path model of the same set of variables…

It includes the RH: from the multiple regression model → that only 4 of the 5 predictors have a unique contribution to understanding the criterion.

But it also has RH: about how the predictors related to each other. Notice that not all the possible paths are included.

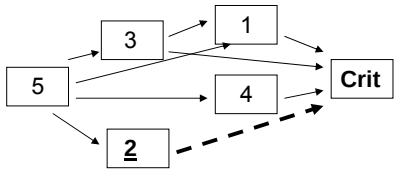
Notice that time is also included in this model – which predicts are causes of which others.



"Direct" and "Indirect" effects …

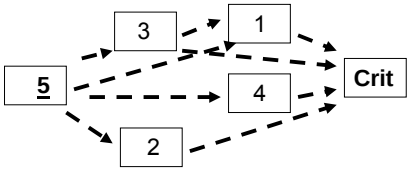
2 has a *direct effect* on Crit

- a "contributor" in both the regression & the path models



Please note: The term "effect" is commonly used in path analyses. It means "statistical effect" not "causal effect" !!!

5 does not have a direct effect on Crit – but does have multiple *indirect effects*

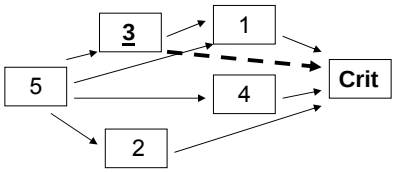


This is a huge advantage of path analysis over multiple regression !!!

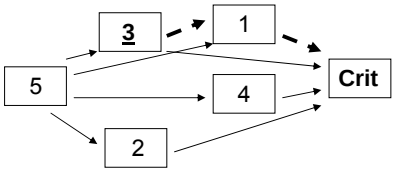
Finding that "5" doesn't contribute to the regression model could mistakenly lead us to conclude "5 doesn't matter in understanding Crit"

Finding that a predictor has a ***only*** an indirect effect in a path model is like finding that an IV has no main effect but is ***only*** involved in an interaction → more complicated analyses show us things that simpler analyses don't!!!

3 has a *direct* effect on Crit



3 also has an *indirect* effect on Crit



There's more to the 3 → Crit relationship than was captured in the regression model

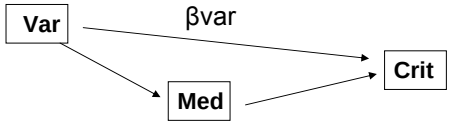
Mediation Analyses

The basic mediation analysis is a 3-variable path analysis.
A correlation shows that "var" is related to the "crit" .

But we wonder if we have the "whole story" – is it really that variable that causes Crit ???

So, we run a path analysis including all 3 variables and compare

- $r_{Crit,Var}$ from the bivarate model &
- β_{Var} from the multivariate model



If $\beta_{Var} = .00$ → complete mediation

If $.00 < \beta Var < r_{Crit,Var}$ → partial mediation

If $\beta Var = r_{Crit,Var}$ → no mediation

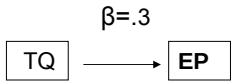
… to investigate "mediation effects"…

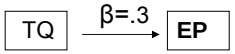
Mediation effects and analyses highlight the difference between bivariate and multivariate relationships between a variable and a criterion (collinearity & suppressor effects).

For example…

For Teaching Quality & Exam Performance → r = .30, p = .01

- for binary regression β = r
- so we have the path model…

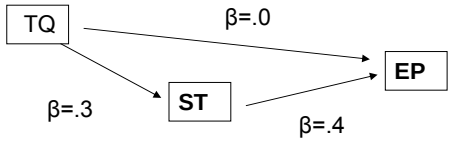




After thinking about the findings for a while, it occurs to one of the researchers that there just might be something else besides Teaching Quality that influences Exam Performance.

- The researcher decides that Study Time (ST) might be such a variable.
- Thinking temporally/causally, the researcher considers that Study Time "comes in between" Teaching and Testing.
- So the researcher builds a mediation model

The resulting model …



Notice that TQ does not have direct effect upon EP !

- Study time completely mediates the TQ effect !

However: Notice that TQ is "still very important" because it is part of understanding Exam Performance …

- it has an indirect effect upon Exam Performance
- TQ is related to ST, which in turn, is related to EP

The "when" of variables and their place in the model …

When a variable is "measured" → when we collect the data:

- usually we collect all the variables at one time

When a variable is "manifested" → when the value of the variable came into being

- when it "comes into being for that participant"
- may or may not be before the measure was taken

E.g., State vs. Trait anxiety

- trait anxiety is intended to be "characterological," "long term" and "context free" → earlier in model
- state anxiety is intended to be "short term" & "contextual" → depends when it was measured

A word of caution …

Structural Models & Path Models are also sometimes called "**Causal Models**" ?!?!?!?

As has always been the case, statistical relationships between variables can only be causally interpreted if …

- an experimental research design (RA & IV manip) is used
- there are no confounds

Data from path models are rarely from experimental designs

- the data are almost always from non-experimental designs
- usually most, if not all, the variables are subject variables

So, "causal models" still only show associations among a set of variables – not their causal relationships !!!

Another word of caution …

Structural Models & Path Models can not be used to test hypotheses about different "structural paths"

For example, path analysis can not be used to decide which of the following is a better model…

Motivation → Study Time → Test Score

Study Time → Motivation → Test Score

You have to convince your audience that the causal/temporal ordering of the variables makes sense – then path analysis can be used to decide which paths do and do not contribute to the model.

Another word of caution …

Mediating variables must occur after what they are mediating

E.g. A correlation shows the Treatment is related to the criterion.

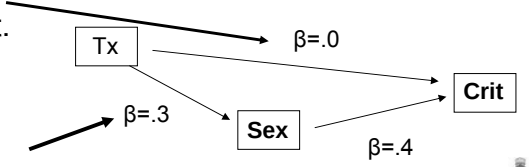
$r_{Crit,Tx} = .4$

But the researcher thinks that sex mediates the treatment …

So we run a mediation analysis:

Looks like a participant's sex mediates the treatment.

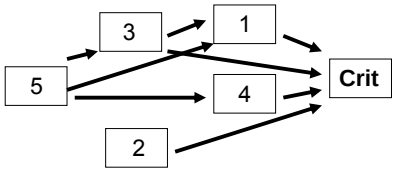
But it also looks like treatment causes a participant's sex ???



Where do the path coefficients come from?

One way is to run a series of multiple regressions…

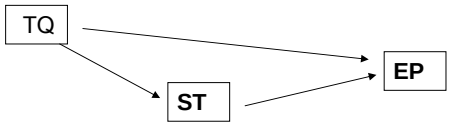
for each analysis: a variable with arrows pointing at it will be the criterion variable and each of the variables having arrows pointing to it will be the predictors



1. Crit = 3 Pred = 5
2. Crit = 1 Preds = 3 & 5
3. Crit = 4 Pred = 5
4. Crit = Crit Preds = 1, 2, 3 & 4

The path coefficients are the β weights from the respective regression analyses (remember that β = r for bivariate models)

Ways to improve a path analysis



1. Antecedents to the current model
   - Variables that “come before” or “cause” the variables in the model
2. Effects of the current model
   - Variables that “come after” or “are caused by” the variables in the model
3. Intermediate causes
   - Variables that “come in between” the current causes and effects.