Introduction to Path Analysis & Mediation Analyses

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

Why multivariate research designs? ➔ Multicausality

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

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

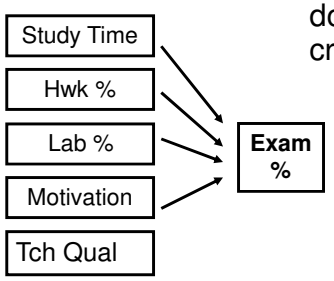
1. Factorial Designs
   - 2 IVs & Interactions
   - does the effect of an IV depend on the value of a 2nd IV?
2. Multiple Regression
   - multiple predictors
   - what does a predictors tell us that other predictors don't?
3. Path Analyses, Structural Models & Mediation Analyses
   - looking at the "temporal causal" relations of behaviors
   - direct and indirect relationships among behaviors

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

- 1 criterion
- 5 predictors

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 tells us the predictor doesn't contribute to understanding the criterion



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.

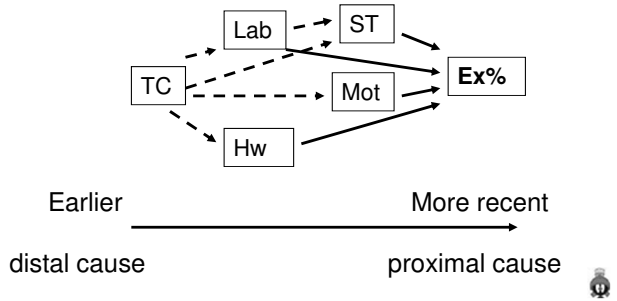
But we don't learn about how the predictors relate to each other!!!

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

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

But it also tells about how the predictors related to each other.

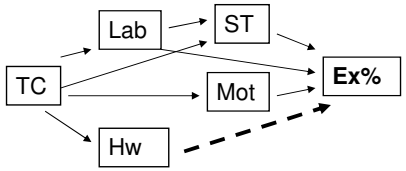
Notice that time is also included in this model – which predicts are causes of which others.



"Direct" and "Indirect" effects …

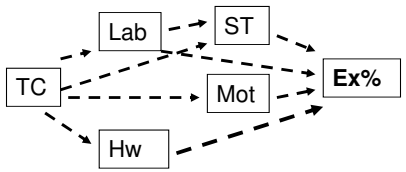
Hw% has a *direct effect* on Crit

• 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" !!!

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

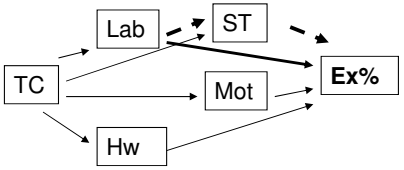


This is a huge advantage of path analysis over multiple regression !!!

Finding that "TQ" doesn't contribute to the regression model could mistakenly lead us to conclude "TQ doesn't matter in understanding Exam Performance"

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!!!

Lab has a *direct* effect on Crit



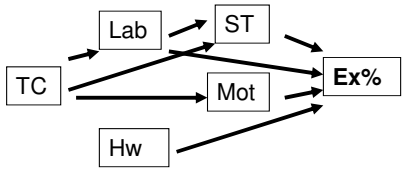
Lab also has an *indirect* effect on Crit

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

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. Pred = TC ➔ Crit = Lab
2. Preds = TC & Lab ➔ Crit = St
3. Pred = TC ➔ Crit = Mot
4. Preds = Lab, Hw, ST & Mot ➔ Crit = Ex%

The path coefficients are just the β weights from the respective multiple regression analyses !!

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

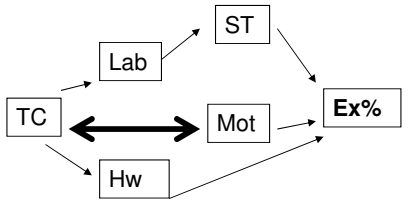
- when it "comes into being for that participant"
- may or may not be before the measure was taken

E.g., State vs. Trait anxiety

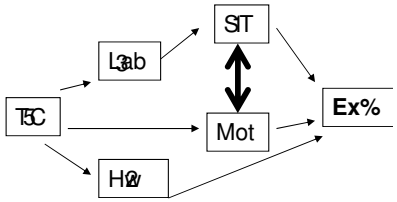
- trait anxiety is intended to be "characterological," "long term" and "context free" → earlier in model
- state anxiety is intended to be "short term" & "contextual" → depends when it was measured

About non-recursive (bi-directional) models

Sometimes we want to consider whether two things that "happen sequentially" might have "iterative causation" – so we want to put in a back-and-forth arrow



Sometimes we want to consider whether two things that "happen at the same time" might have "reciprocal causation" – so we want to put in a sideways arrow



Neither of these can be "handled" by path analysis.

However, this isn't really a problem because both are a misrepresentation of the involved causal paths! The real way to represent both of these is …

The things to remember about "cause"

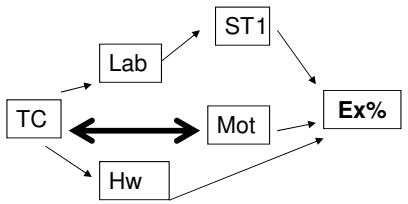
1. "cause takes time" or "cause is not immediate"
   - even the fastest chemical reactions take time
   - behavioral causes take an appreciable amount of time
2. Something must "be" to "cause something else to be"
   - a variable has to be manifested as an effect of some cause before it can itself be the cause of another effect
   - Cause comes before effect → not at the same time

When you put these ideas together, then both "sideways" and "back-and-forth" arrows don't make sense and are not an appropriate portrayal of the causations being represented.

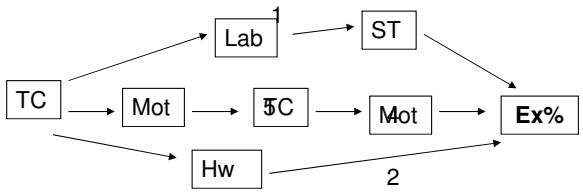
The causal path has to take these two ideas into account…

About non-recursive (bi-directional) models

If "TC" causes "Mot", then "ot" changes "TC", which changes "4" again, all before the criterion is caused, we need to represent that we have 2 "TCs" and 2 "Mots" in a hypothesized sequence.

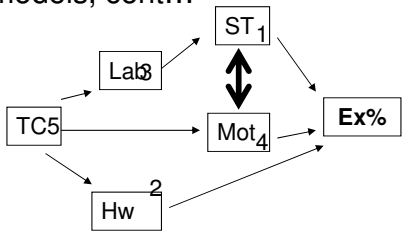


We also have to decide when Lab, ST & Hw enter into the model, temporally &/or causally. Say …

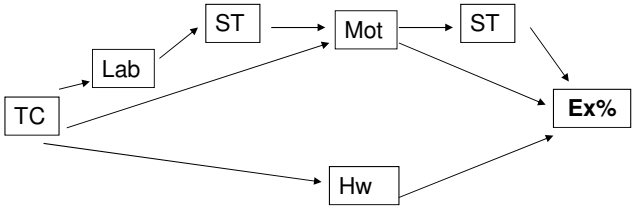


About non-recursive (bi-directional) models, cont…

When applying these ideas to "sideways arrows" we need to remember that the cause comes before the effect.



To do that, we have to decide (& defend) which comes first – often the hardest part) and then add in the second causation, etc.… As well as sort out where the other variables fall temporally &/or causally. Perhaps …



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"

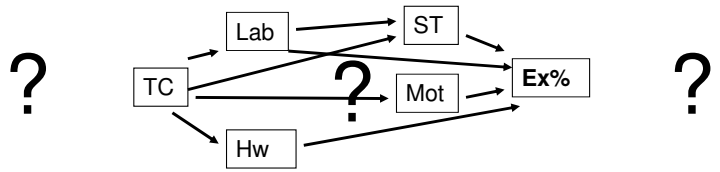
For example, path analysis can not be used to decide which of the following is a better model…

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.

Ways to improve a path analysis (and any program of research) !!



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
   - Additional variables that "come in between" the current causes and effects.

Show path example

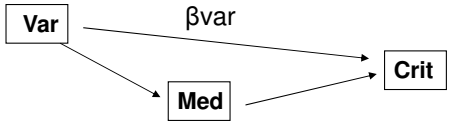
Mediation Analyses

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

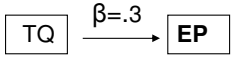
If $\beta Var = r_{Crit,Var}$ → no mediation

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

For example…

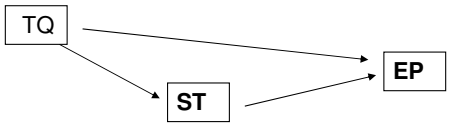
For Teaching Quality & Exam Performance → r = .30, p = .01

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

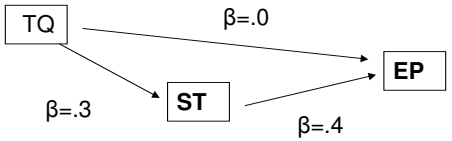


It occurs to one of the researchers that something besides Teaching Quality may influence Exam Performance.

- Study Time (ST) might be such a variable.
- Thinking temporally/causally, 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

An important example of partial mediation!!

Many studies of the relationship between sex (female = 1 male = 2) & salary show a positive correlation, e.g., r = .25, p < .01) → males have significantly higher salaries than females.

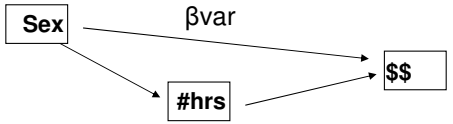
You might suspect that salary has multiple causes, and that this single bivariate correlation might not be telling the whole story.

So, we might look for other variables that "come in between" and "mediate" the relationship between sex and salary → many folks have

One variable that has been examined is "number of hours worked / week"

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

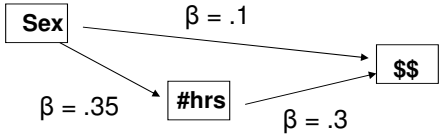


An important example of partial mediation!!

Remember the bivariate results ➔ r = .25, p < .01
→ males have significantly higher salaries than females.

Here are the mediation analysis results.

There is much to notice!



Sex is related to #hrs → men average more hours than women

#hrs is related to $$ → folks who work more hours earn more $$

The sex → $$ relationship is mediated by #hrs…

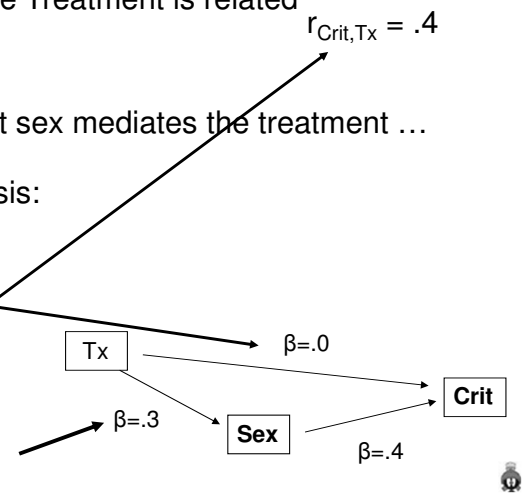
β = .1 is less than r = .25

But the mediation is ***ONLY PARTIAL MEDIATION*** !!!
Even after controlling for #hr there is still a relationship between Sex & $$ → there is more to this complex story!!

Be careful when selecting "mediating variables" …

Mediating variables must occur after what they are mediating

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



Example of Path Analysis in Programmatic Research ….

Nearly everybody who looks for it finds a relationship between "practice" and "performance".

For example, in a recent semester the correlation between % Pink Things completed and Exam 1% grade was r(157) = .33, p<.01. This would be interpreted as, "Those who completed more of the Pink Things tended to have higher grades on Exam 1.

However, this is not an experimental study, so the "Pink Thing effect" may be confounded by lots of other variables.

While we can't measure or even think all of the possible confounding variables, we can consider what are things that might be related to both % EDUs completed and Exam scores ????

We chose 3 for study: motivation, exam study time, GPA

Bivariate & Multivariate contributions – DV = Exam 1% grade

predictor➔	Motiv	St. Time	GPA	% Pink
r(p)	.28(<.01)	.45 (<.01)	.46 (<.01)	.33(<.01)

All of these predictors have substantial correlations with Exam grades!!

β(p)	.32(.02)	-.25(.04)	.09(.51)	.58 (.01)

GPA does not have a significant regression weights – after taking the other variables into account, it has no unique contribution!

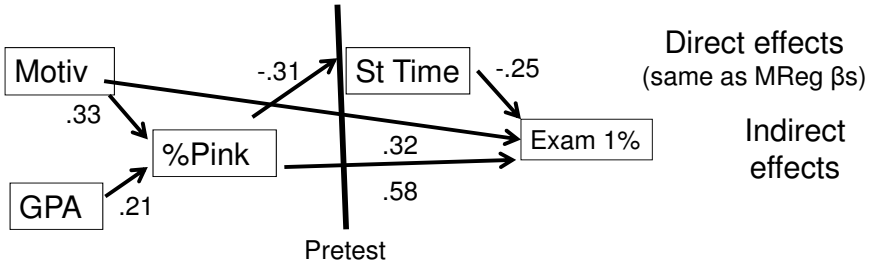
%Pink does have a significant regression weight. Even after taking the other variables into account, those who do more EDU exercises tend to do better on the exam.

Motivation does have a significant regression weight. Even after taking the other variables into account, those who are more motivated tend to do better on the exam.

Exam study time has a significant regression weight, however, notice that it is part of a suppressor effect! After taking the other variables into account, those who study more for the test actually tend to do poorer on the exam.

Notice that only two of the 4 predictors had the same "story" from the bivariate and multivariate analysis!!!!

Path Analysis – allows us to look at how multiple predictors relate to the criterion – considering both "direct" and "indirect" relationships!!



Direct effects
(same as MReg βs)

Indirect effects

GPA → no direct effect – but indirect effects thru %pink & St Time

Motiv → direct effect – also indirect effects thru %pink & St Time

%Pink → direct effect – also indirect effect thru St Time

-β for St Time? Less %Pink predicts more St Time, suggesting that those who study more were those who did less work before they started to study for the exam, and they also did poorer on the exam!