Path Analysis Rudiments

Super rudiments...

Today

- Causality is back
- From multiple regression to path analysis
- Simple example
- Exercise will be online...
- Next week: advanced stats topics...

Causation / Correlation

- A common critique of observational studies is that they do not reveal causal forces.
- "Correlation does not imply causation."
- Keith challenges using this blanket admonition.

"Empirically observed covariation is a necessary but not sufficient condition for causality.

...

Correlation is not causation but it sure is a hint."

E. Tufte

Causality

- Keith: "The term cause is thus a probabilistic statement." (p. 220)
- *By increasing X, you will increase the probability of Y.*

Observational Study

- Imagine conducting a large longitudinal study of children. You hope to uncover the underlying influences of variable Y.
- You measure a variety of other variables during this longitudinal study: X_1, X_2, ...

Four Considerations

- 1) Theory
  - Your work is guided by an explicit theory of the interrelationships among these variables. Theory is carefully reasoned and justified by a variety of means (e.g., previous research, logic, etc.).

Four Considerations

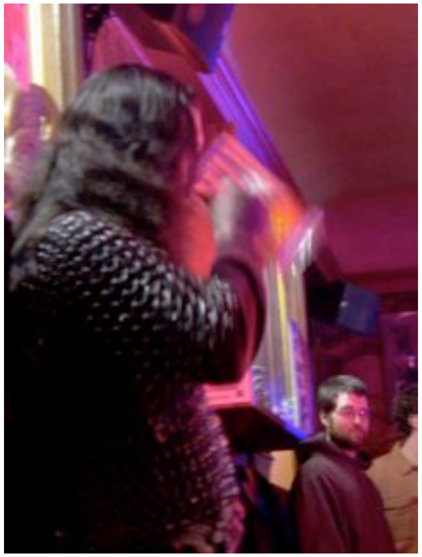
- 2) Temporal precedence
  - Often, your variables have a certain temporal logic to them: It is natural to assume, e.g., that X_1 precedes your outcome variable, and that X_2 precedes both (e.g., SES > homework > test score).

Four Considerations

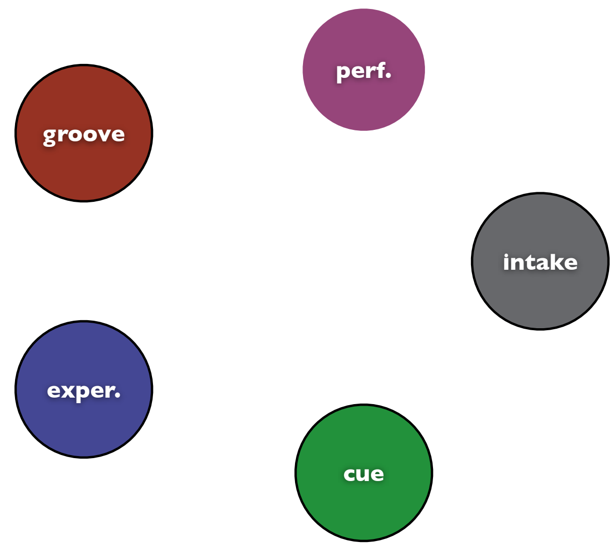
- 3) Previous research
  - Almost always, you have available to you a large set of previous studies to guide you. This previous work informs you how these variables may relate causally.

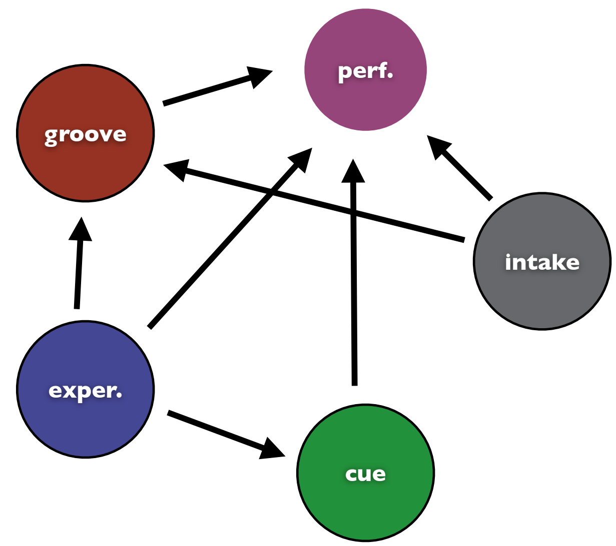
Four Considerations

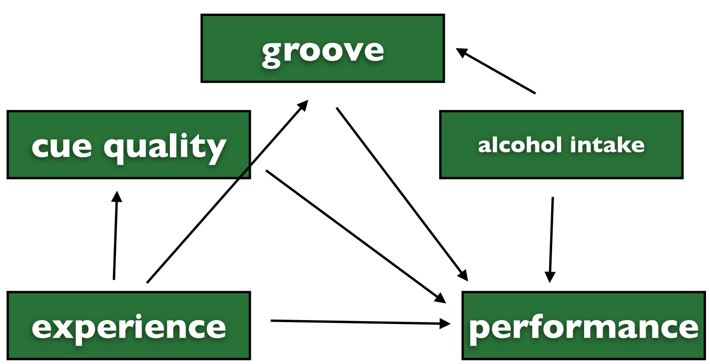
- 4) Logic
  - The variables of interest can be subject to a certain sort of logic or reasoning: It would be unnatural, e.g., to consider X_1 to be caused by the outcome variable, for example (child test scores *causing* family SES?).



- Experience
- Beer
- Cue quality
- Groove



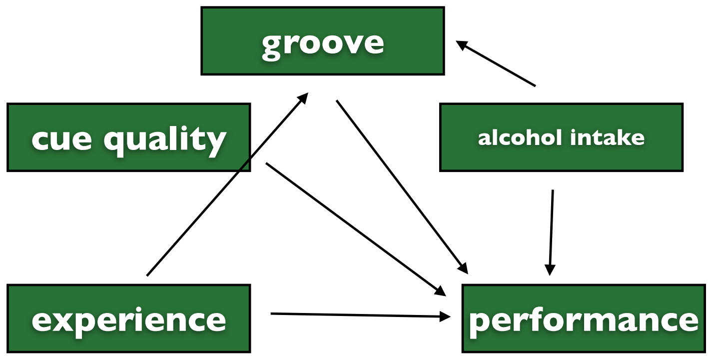




"Solving" the model

- We "solve" the model by attaching specific effect values to our identified paths
- These values reflect the strength of the relationship, using the standardized regression coefficient
- Step 1: Let's get the direct effects

Direct effects

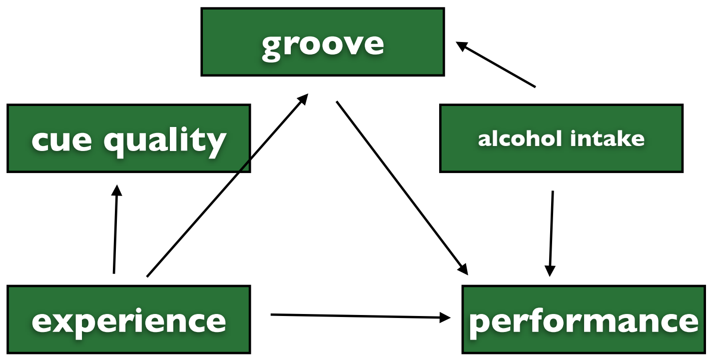


Run a simultaneous multiple regression analysis

"Solving" the model

- Step 1: Let's get the direct effects...
  - ...to the outcome variable.
- But there are direct paths between other variables in the model.
- Step 2: Let's get the other path effects...

Other effects

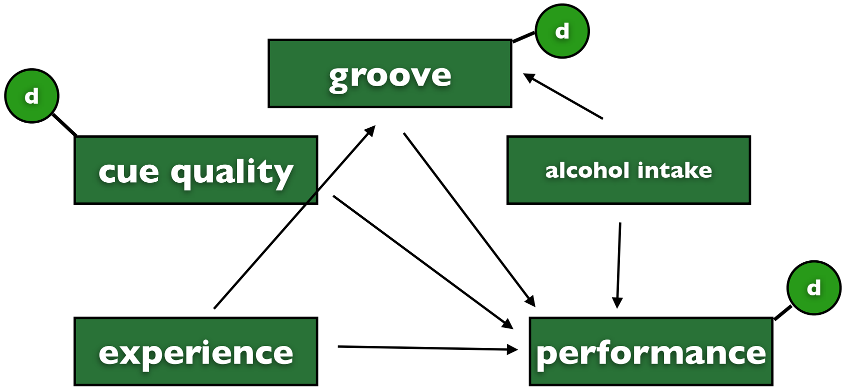


Run a relevant simultaneous multiple regression analysis

"Solving" the model

- Step 1: Let's get the direct effects...
- Step 2: Let's get the other path effects...
- Step 3: Find the "disturbances"
  - Our theory does not account for all the variance of the endogenous variables.
  - Disturbances are the overall influence by "unmeasured variables."

Disturbances



Run a relevant simultaneous multiple regression analysis, and note the % of variance unexplained. Take root.

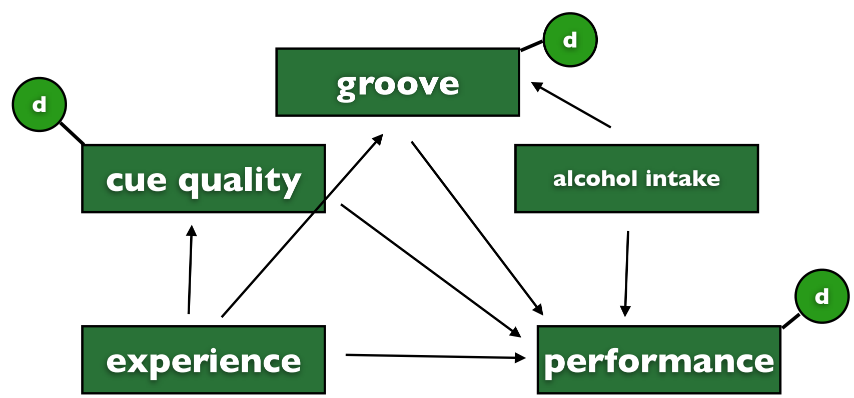
Summary of solution

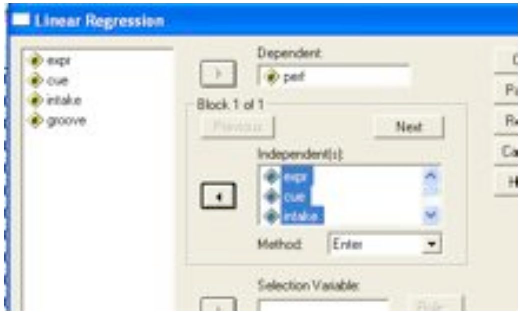
- Step 1: Let's get the direct effects...
  - MR with outcome on all direct predictors
- Step 2: Let's get the other path effects...
  - Relevant MR with endogenous variables
- Step 3: Find the "disturbances"
  - sqrt($1-R^2$) with relevant MR model from above.

Let's do it

	perf	expr	cue	intake	groove
1	2	1.0	1	5	3
2	7	7.0	8	5	6
3	2	1.0	4	3	2
4	0	.0	1	1	0
5	8	9.0	6	6	9
6	6	9.0	7	1	4
7	5	8.0	6	1	3
8	3	3.0	4	2	2
9	3	3.0	3	3	3
10	7	9.0	8	5	6
11	1	.0	4	1	1
12	5	5.0	7	5	5
13	3	3.0	4	1	6
14	3	2.0	3	3	6
15	5	6.0	8	1	5
16	6	7.0	6	4	6

Disturbances



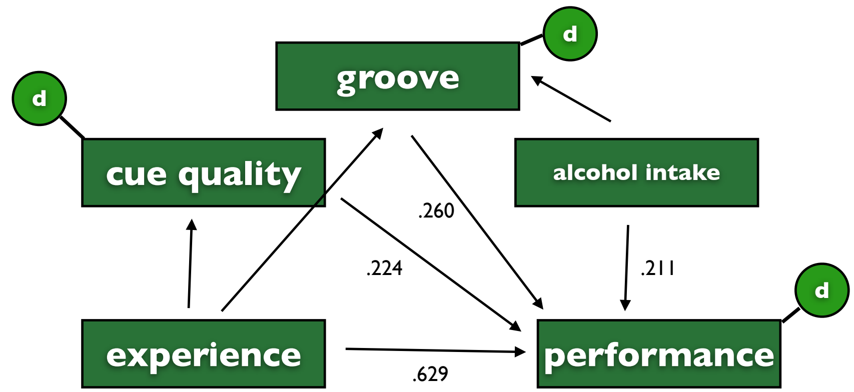


Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-.161	.103		-1.562	.122
	expr	.378	.014	.629	26.702	.000
	cue	.212	.020	.224	10.579	.000
	intake	.200	.016	.211	12.672	.000
	groove	.237	.018	.260	12.826	.000

a. Dependent Variable: perf

Disturbances



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.987[a]	.975	.974	.289

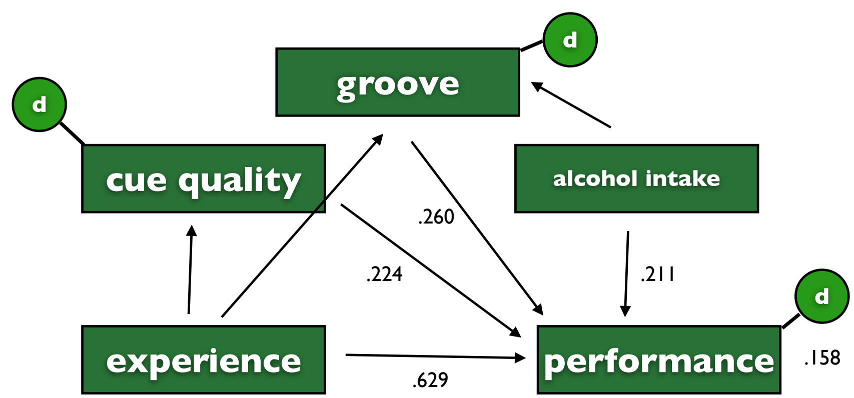
a. Predictors: (Constant), groove, intake, cue, expr

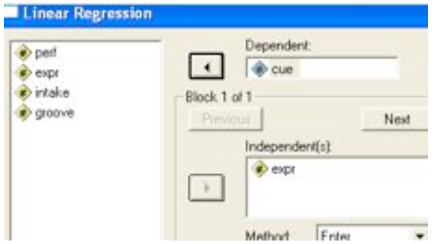
$$d = \text{sqrt}(1 - R^2)$$

$$= \text{sqrt}(1 - .975)$$

$$= .158$$

Disturbances





sqrt(1-.405) = .771

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.636[a]	.405	.399	1.462

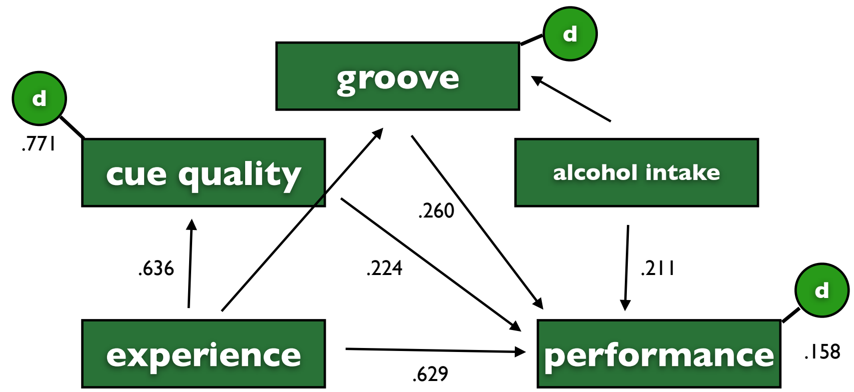
a. Predictors: (Constant), expr

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	3.050	.295		10.346	.000
	expr	.404	.050	.636	8.167	.000

a. Dependent Variable: cue

Disturbances



sqrt(1-.352) = .804

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.593[a]	.352	.339	1.593

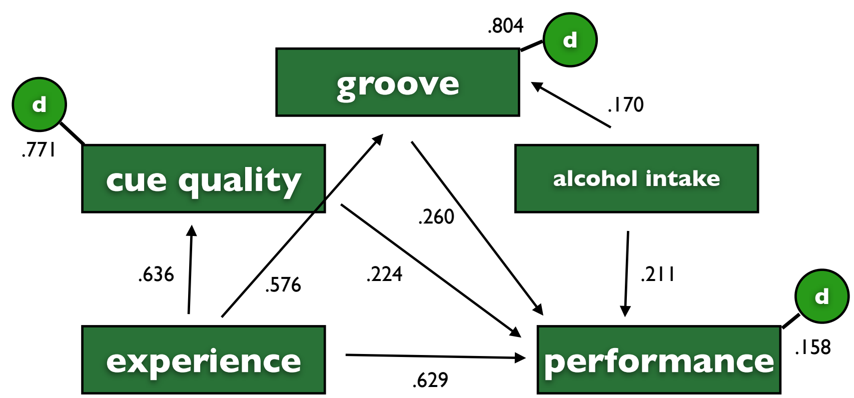
a. Predictors: (Constant), intake, expr

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	2.317	.414		5.593	.000
	expr	.380	.054	.576	7.045	.000
	intake	.177	.085	.170	2.072	.041

a. Dependent Variable: groove

Disturbances



Interpretation

- Our hypothetical notion of "groove" seems to have a strong influence on performance (.260).
- There is very strong mediation with experience (.576). This means: More experienced players are more likely to exhibit groove, thus more likely to exhibit high performance.

Interpretation

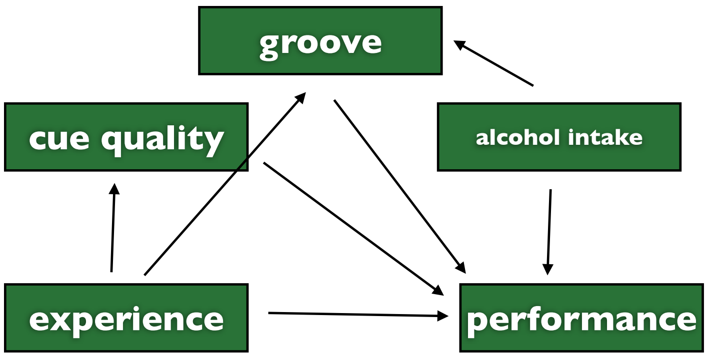
- Moderate mediation with intake (.170). This seems to indicate that alcohol intake during pool player indirectly influences performance through groove -- the more beer, the more groove, and this seems to have an influence on performance.

Paths + thinking

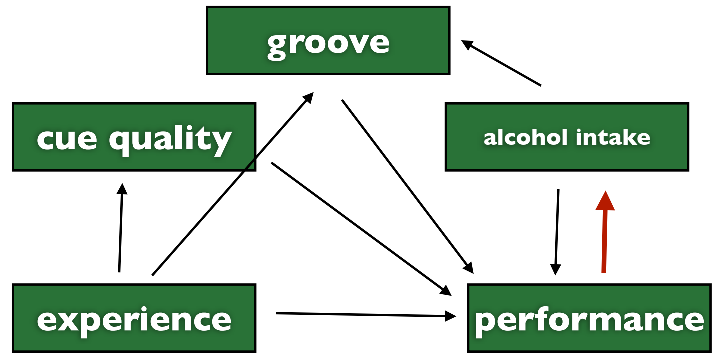
- Theory, temporal precedence, logic, previous research
- You can use MR to estimate the various paths in your model based on careful reasoning about those directional influences
- Those *causes*

Other issues

- Recursive vs. nonrecursive models
- Identification
- Indirect vs. direct effects, total effects
- SPSS tricks...



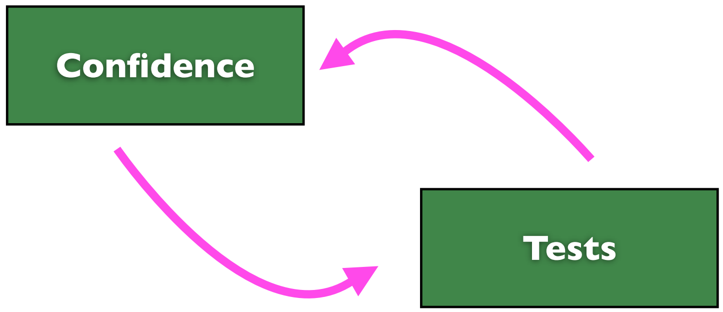
Known as a “recursive” model



Known as a “non-recursive” model

Keith Text

- *“It is tempting to...solve difficult questions of presumed cause and effect by deciding that such effects are reciprocal. Can’t decide...draw paths for both directions! Generally, however, this is equivocation rather than decision.”*



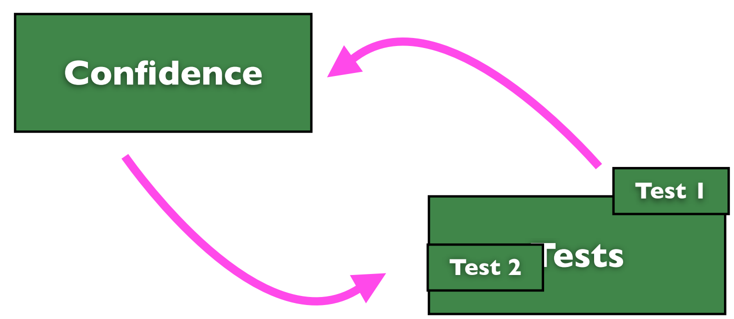
1 Impredicativity, Dynamics, and the Perception-Action Divide

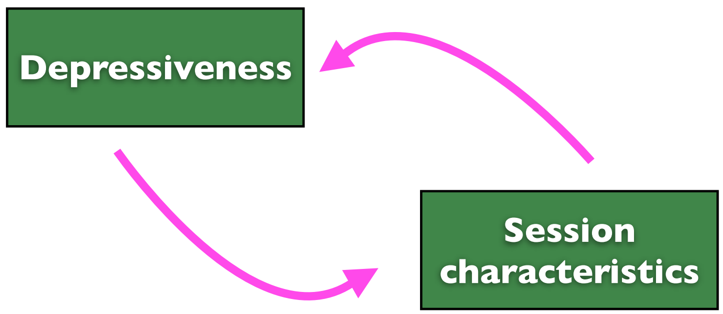
M. T. Turvey

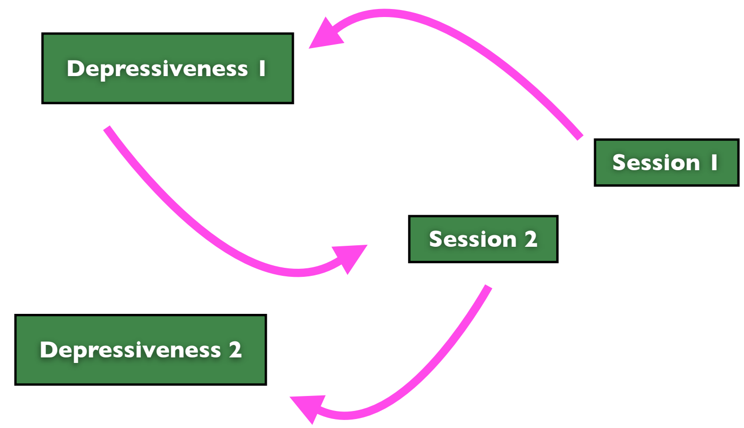
Center for the Ecological Study of Perception and Action, University of Connecticut, Storrs CT, USA and Haskins Laboratories, New Haven CT, USA

The real world

- Over time, the rule is rich integration and feedback loops in worldly contexts
- Your goal in designing studies is to measure variables that permit you to invoke unidirectionality
- E.g., setup experimental conditions
- E.g., measure an instance of test performance



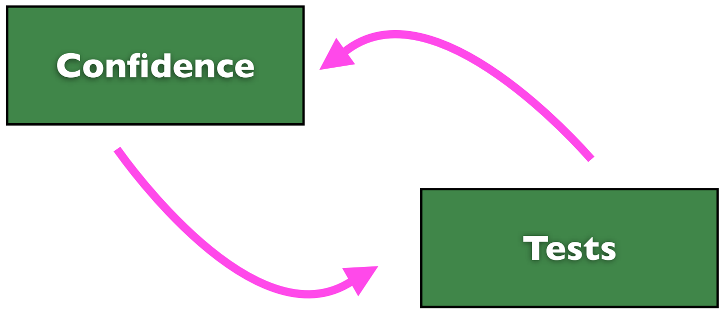




Issue: Time

Identification

- One reason for unidirectionality (for "recursive" models) is that path analysis is equivalent to a path-estimation process using variable correlations.
- When you draw a model we can therefore describe its identification: Is it overidentification or underidentified?

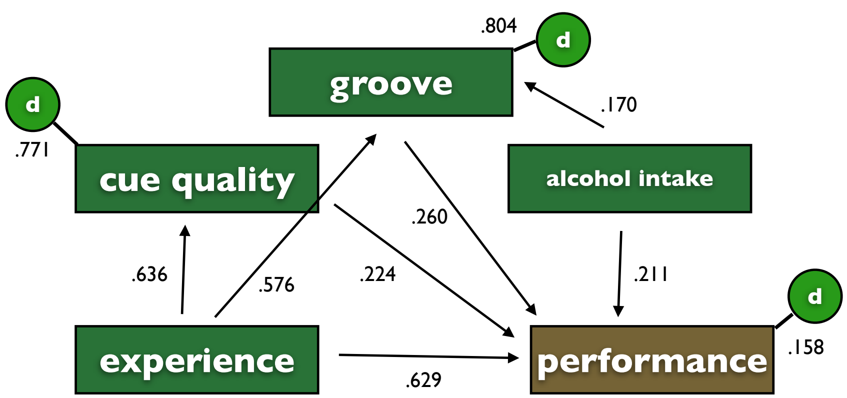


1 correlation, 2 paths: underidentified model

Indirect effects

- Indirect effect *on the outcome variable*.
- The *indirect effect* of a variable is estimated by multiplying the relevant component paths of that indirect effect.
- The total effect of a variable is the sum of indirect effects and the direct effect (on the outcome variable).

Our pool model



Experience on performance

- Direct effect: .629
- Indirect effects:
  - Through cue: .636 * .224 = .142
  - Through groove: .576 * .260 = .150
- Total effect of experience:
  - .629 + .142 + .150 = .921

What about sequential?

- Remember sequential MR?
- Its offerings give us the *total effects* of the entered variable onto the outcome variable.
- We would enter experience first and just the standardized coefficient. This is the total effect of experience.
- Within errors of rounding, it should be the same value.

Significance

- Goodness-of-fit statistics to assess an overall model.
- For now, you can assess the statistical significance of the paths using the standard procedure.
- Some suggest "model trimming" by removing non-significant coefficients.