## Center for Causal Discovery:

# Summer Workshop - 2015



#### June 8-11, 2015

Carnegie Mellon University

# Outline

- 1) Motivation
- 2) Representing/Modeling Causal Systems
- 3) Estimation and Model fit
- 4) Hands on with Real Data

# Estimation

	PUF			(m)	1 10 -	a 16 1			
▲	untitled	L.tet - Tetra	d 5.0.0-1						
Fi	le Edit	Logging	Template	Window	Help				
	\lambda untitle	d1.tet 🔛							
	+	*@)							
	Gr	aph		X → Grap	·Y h1 ∋				
	Gr Manip	aph pulation							
	Com	parison		X	Y				
	Para M	metric odel		PM SEM	<b>1</b> PM				
	Insta M	ntiated odel							
	D	ata			• Ƴ 1 ™			× Y ····································	
	Da Manip	ata pulation					L		
	Esti	mator		Data	2 3 1				
	Up	dater		SEM (	Data				

# Estimation



# **Tetrad Demo and Hands-on**

- 1) Select Template: "Estimate from Simulated Data"
- Build the SEM shown below all error standard deviations = 1.0 (go into the Tabular Editor)
- 3) Generate simulated data N=1000
- 4) Estimate model.
- 5) Save session
  - as "Estimate1"



# Estimation



# Coefficient inference vs. Model Fit

Coefficient Inference: Null: coefficient = 0, e.g.,  $\beta_{X1 \rightarrow X3} = 0$ p-value = p(Estimated value  $\widehat{\beta}_{X1 \rightarrow X3} \ge .4788 | \beta_{X1 \rightarrow X3} = 0 \& \text{rest of model correct}$ ) Reject null (coefficient is "significant") when p-value <  $\alpha$ ,  $\alpha$  usually = .05



## Coefficient inference vs. Model Fit

Coefficient Inference: Null: coefficient = 0, e.g.,  $\beta_{X1 \rightarrow X3} = 0$ p-value = p(Estimated value  $\widehat{\beta}_{X1 \rightarrow X3} \ge .4788 | \beta_{X1 \rightarrow X3} = 0 \& \text{rest of model correct})$ Reject null (coefficient is "significant") when p-value < <  $\alpha$ ,  $\alpha$  usually = .05,

Model fit: Null: Model is *correctly specified* (constraints true in population) p-value =  $p(f(Deviation(\Sigma_{ml}, S)) \ge 5.7137 | Model correctly specified)$ 



## Coefficient inference vs. Model Fit

	coefficient $\widehat{\beta}_{X1 \rightarrow X3}$	Model fit $\chi^2_{df}$
	Null: $\beta_{X1 \rightarrow X3} = 0$	Null: Model is correctly specified
p-value < .05	Can reject 0 Significant edge	Can reject correct specification, Model not correctly specified
p-value >.05	Can't reject 0, insignificant edge	Can't reject correct specification, model <i>may be</i> correctly specified

# Model Fit

Specified Model



True Model



Implied Covariance Matrix						
	X1	X2	X3			
X1	1					
X2	β1	1				
X3	β1*β2	β2	1			

#### **Population Covariance Matrix**

	X1	X2	X3
X1	1		
X2	.6	1	
X3	.3	.5	1

$$\widehat{\beta 1} = r_{X1,X2} = \sim .6$$
$$\widehat{\beta 2} = r_{X2,X3} = \sim .5$$
$$\widehat{\rho}_{X1,X3} = \widehat{\beta 1} \ \widehat{\beta 2} = \sim .3$$

10

# Model Fit

X1 M1 B1 X2 M2 X3 M3 M3

**Specified Model** 



**True Model** 

Implied Covariance Matrix					Population Covariance Mat			ix	
	X1	X2	X3			X1	X2	X3	
X1	1				X1	1			
X2	β1	1			X2	.6	1		
X3	β1*β2	β <b>2</b>	1		X3	.32	.5	1	

#### Unless *r<sub>X1,X3* **=** *r<sub><b>X1,X2</sub> <i>r<sub>X2,X3</sub>*</sub>

Estimated Covariance Matrix *≠* Sample Covariance Matrix

# Model Fit

#### **Specified Model**

True Model





Implied Covariance Matrix					Population Covariance Mat			ix	
	X1	X2	X3			X1	X2	X3	
X1	1				X1	1			
X2	β1	1			X2	.6	1		
X3	β1*β2	β <b>2</b>	1		X3	.32	.5	1	

Model fit: Null: Model is *correctly specified* (constraints true in population)  $\rho_{X1,X3} = \rho_{X1,X2} \rho_{X2,X3}$ 

p-value =  $p(f(\text{Deviation}(\Sigma_{ml}, S)) \ge \chi 2 \mid \text{Model correctly specified})$ 

# Tetrad Demo and Hands-on

- Create two DAGs with the same variables each with one edge flipped, and attach a SEM PM to each new graph (copy and paste by selecting nodes, Ctl-C to copy, and then Ctl-V to paste)
- 2) Estimate each new model on the data produced by original graph
- 3) Check p-values of:
  - a) Edge coefficients
  - b) Model fit
- 4) Save session as:"estimation2"



# Break

### Charitable Giving

#### What influences giving? Sympathy? Impact?

"The Donor is in the Details", Organizational Behavior and Human Decision Processes, Issue 1, 15-23, C. Cryder, with G. Loewenstein, R. Scheines.

	N = 94
[1,0]	Randomly assigned experimental condition
[17]	How concrete scenario I
[17]	How much sympathy for target
[17]	How much impact will my donation have
[05]	How much actually donated
	[1,0] [17] [17] [17] [05]

N = 0.4

### **Theoretical Hypothesis**



# **Tetrad Demo and Hands-on**

- 1) Load charity.txt (tabular not covariance data)
- 2) Build graph of theoretical hypothesis
- 3) Build SEM PM from graph
- 4) Estimate PM, check results

![](_page_17_Figure_5.jpeg)

### Foreign Investment

Does Foreign Investment in 3<sup>rd</sup> World Countries inhibit Democracy?

Timberlake, M. and Williams, K. (1984). Dependence, political exclusion, and government repression: Some cross-national evidence. American Sociological Review 49, 141-146.

#### N = 72

- PO degree of political exclusivity
- CV lack of civil liberties
- EN energy consumption per capita (economic development)
- FI level of foreign investment

#### Case Study: Foreign Investment Alternative Models

![](_page_19_Figure_1.jpeg)

There is no model with testable constraints (df > 0) that is not rejected by the data, in which FI has a positive effect on PO.

![](_page_19_Figure_3.jpeg)

# Tetrad Demo and Hands-on

- 1) Load tw.txt (this IS covariance data)
- 2) Do a regression
- 3) Build an alternative hypothesis, Graph SEM PM, SEM IM
- 4) Estimate PM, check results

# Hands On Lead and IQ

- Lead: Lead concentration in baby teeth
- CIQ: child's IQ score at 7
- PIQ: Parent's average IQ
- MED: mother's education (years)
- NLB: number of live births prior to child
- MAB: mother's age at birth of child
- FAB: father's age at birth of child

# Hands On Lead and IQ

- 1) Load leadiq1.tet
- 2) Specify different hypotheses, test the model fit on each
- See if you can find a model (without using search), that is not rejected by the data