Structural Equation Modeling

SDS 390 Structural Equation Modeling

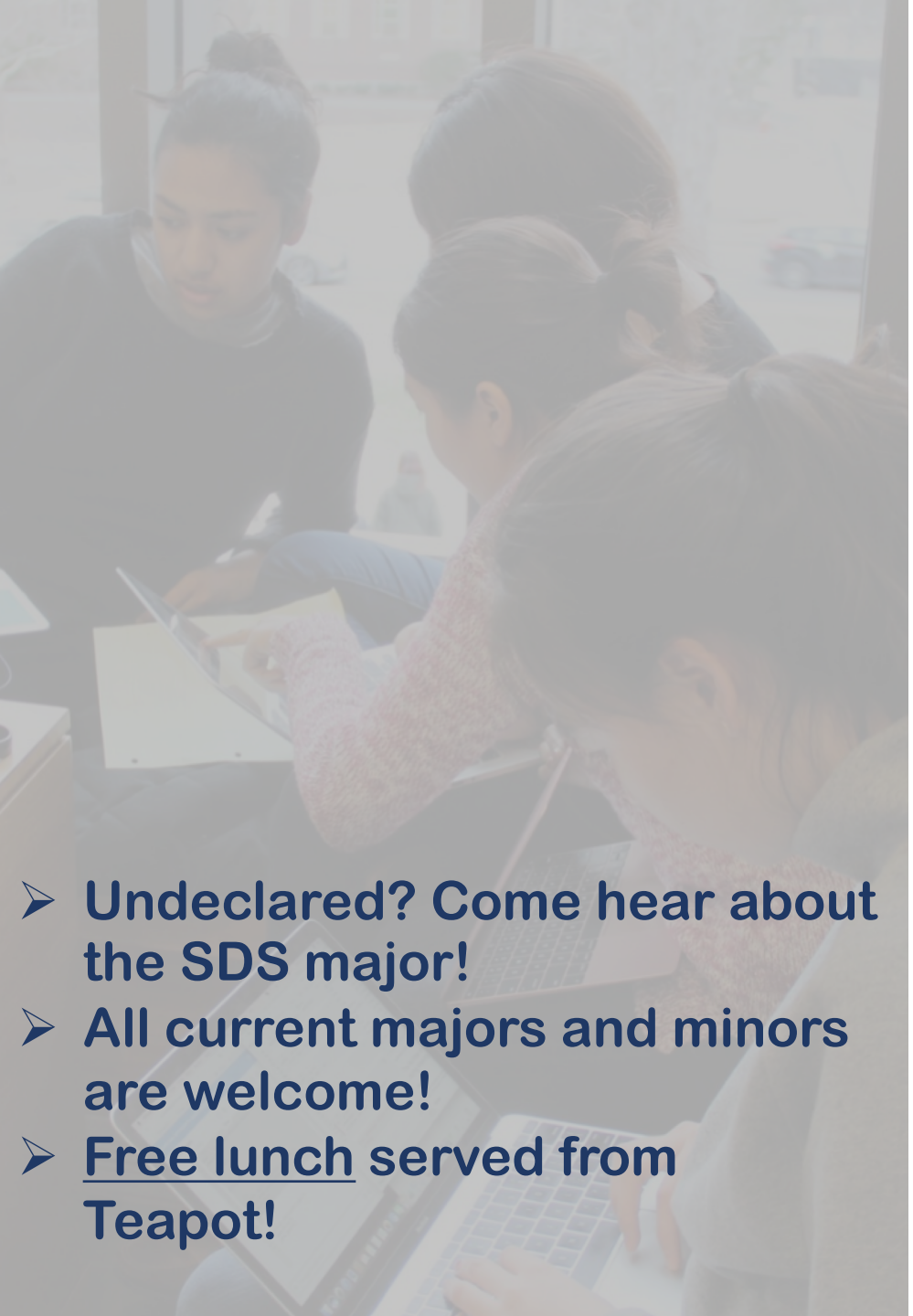
Monday Mar 18, 2019

FROM INTERESTING TO ACTIONABLE: WHY GOOD CONTEXT MATTERS AS MUCH AS GOOD CODE

Friday, Mar 22
McConnell B15
12:15p - 1:00p
Yes, PIZZA!

GINA DELCORAZON '04, DIRECTOR OF DATA & ANALYTICS AT THE NATIONAL MATH AND SCIENCE INITIATIVE AND PHD CANDIDATE AT PRINCETON

Insights from data can transform anyone's work, from school principals to pharmaceutical brand managers, but the framing and focus of their questions vary significantly. Honing a multitude of interesting results down to findings on which a client can take action that moves them closer to meeting their goals is one of the critical learnings of Gina's career. Come hear about this trajectory and get practice developing the kind of questions and analyses that are useful and within the locus of control of organizations who need data scientists.



Statistical and Data Sciences

Presentation of the Major

- Undeclared? Come hear about the SDS major!
- All current majors and minors are welcome!
- Free lunch served from Teapot!

Tuesday, March 26

12:15 pm – 1:05 pm

Ford Hall Atrium

Agenda

- Mid-semester assessment
- Structural Equation Modeling
  - Estimation and fit
  - Model identification
  - Steps to SEM
  - R example

Mid-Semester Assessment

bit.ly/SDS390msa

Causal Inference with SEM

- Insight regarding causality could be inferred from these analyses, but only after making careful arguments about conditional independence.
- The data can then be evidence of these causal claims, but also if your arguments hold up.
- The scientific community also weighs in on your causal claims through peer review.

Structural Equation Modeling

- SEM considers measurement error by using latent variables in addition to measured variables.
- SEM is a combination of path analysis and CFA.
  - confirmatory factor analysis—measurement model.
  - causal ordering of factors—structural model.

Building & Assessing the Model

- Building the model is a combination of what we did for path analysis and CFA.
  - First we get our measurement model to fit, then we specify our structural model.
- Parameter estimation (next week)
  - Maximum likelihood estimation techniques.
    - Exogenous variables – any type
    - Endogenous variables – continuous, assumed to be normal
- Model fit
  - Popular model fit statistics include the chi-square statistic and test, Comparative Fit Index (CFI), and Root Mean Squared Error of Approximation (RMSEA).

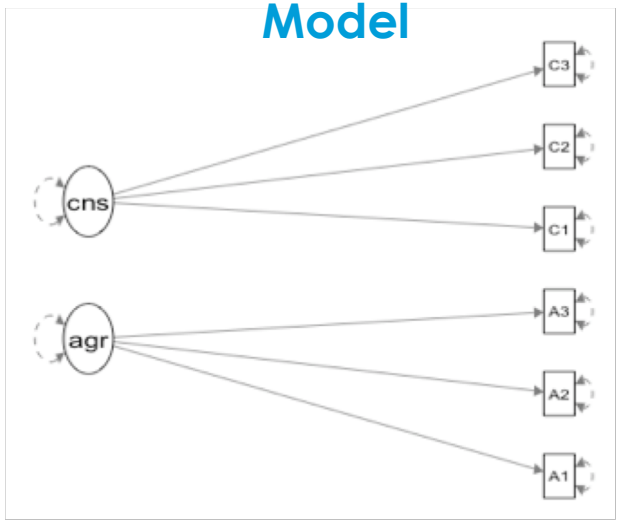
Confirmatory Factor Analysis (CFA)

- Does the model we have in our heads actually fit the data?

Data Cor matrix

	A1	A2	A3	C1	C2	C3
A1	1.000	-0.340	-0.265	0.028	0.016	-0.019
A2	-0.340	1.000	0.485	0.092	0.136	0.192
A3	-0.265	0.485	1.000	0.097	0.141	0.132
C1	0.028	0.092	0.097	1.000	0.428	0.308
C2	0.016	0.136	0.141	0.428	1.000	0.356
C3	-0.019	0.192	0.132	0.308	0.356	1.000

Model



Model implied Cor matrix

	A1	A2	A3	C1	C2	C3
A1	1.000					
A2	-0.337	1.000				
A3	-0.256	0.492	1.000			
C1	-0.063	0.122	0.093	1.000		
C2	-0.074	0.143	0.109	0.418	1.000	
C3	-0.056	0.108	0.082	0.316	0.370	1.000

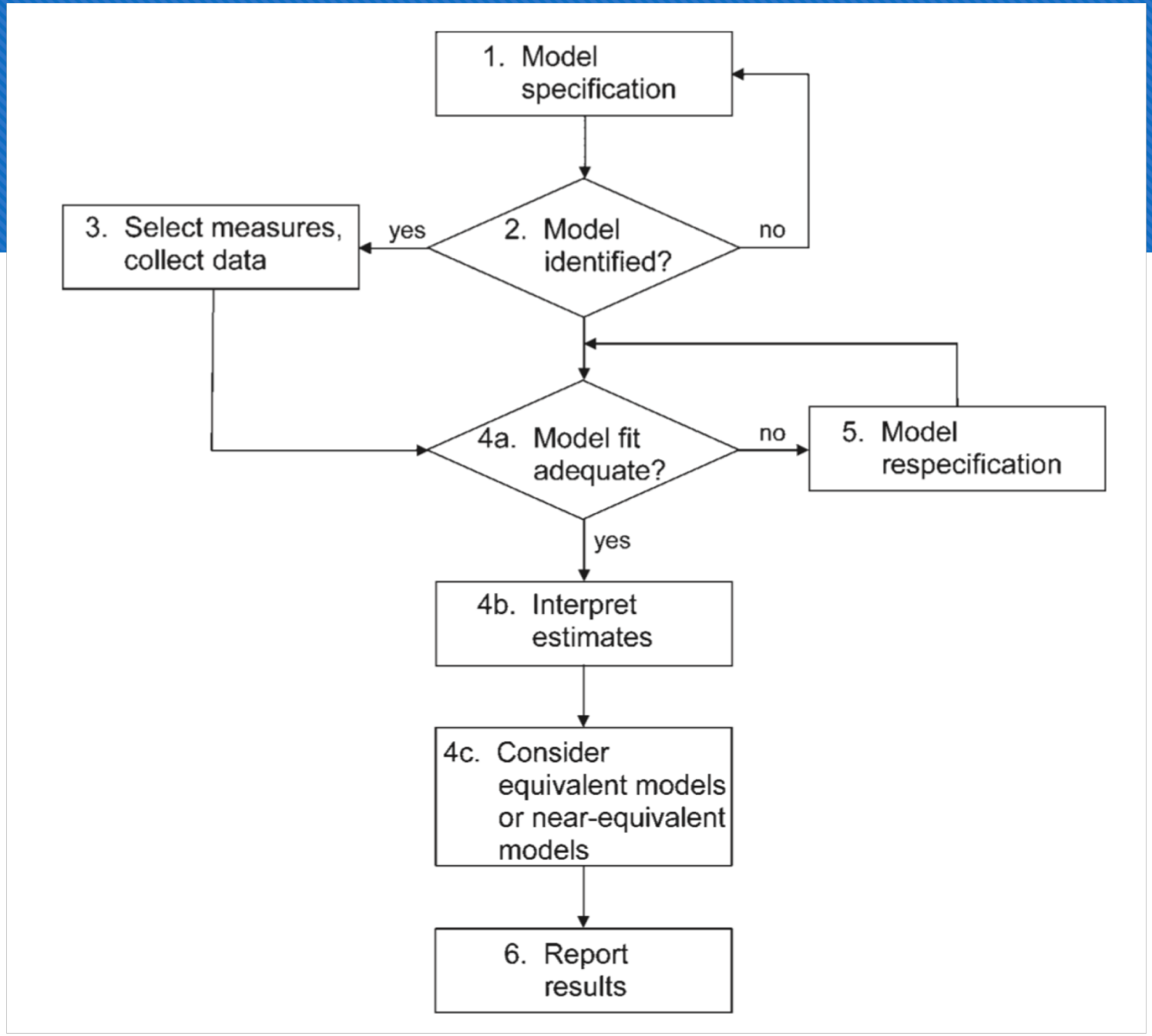
Fit?

Fit of the Measurement Model

- The measurement model is the more complex of the two models.
- If the measurement model does not fit the data, then knowing the relationships among the latent variables is useless.
  - If the measurement model does not fit, then your whole SEM will not fit.

Six Steps in SEM

1. Specification
2. Model identification
3. Estimation and fit
4. Re-specification



Model Identification

- Number of observations for SEM is the lower diagonal elements in the covariance matrix plus the variances
- Model degrees of freedom = knowns - unknowns
  - Knowns = $\frac{v(v+1)}{2}$
  - Unknowns =
    - **Paths**
    - **Covariances between the exogenous variables, between the disturbances, and between exogenous variables and disturbances, and**
    - **Variances of the exogenous variables and disturbances of endogenous variables**
    - **Minus the number of linear constraints**

Identification Worksheet

Model Identification

- Assigning factors scales (marker variables)
  - Underidentification – knowns less than unknowns
  - Just-identified or saturated – knowns equal unknowns
  - Over-identified – knowns greater than unknowns

Model Identification

- Assigning factors scales (marker variables)
  - Underidentification
  - Just-identified or saturated
  - **Over-identified:**

$$a + b = 6$$
$$2a + b = 10$$
$$3a + b = 12$$

Find values of *a* and *b* that are positive and yield total scores such that the sum of the squared differences between the observations (6, 10, 12) and these totals is as small as possible.

No single solution:

- (a = 4, b = 2)
- (a = 2, b = 6)

Scaling Latent Variables

- Three methods
  1. Marker variable
  2. Latent variable variance set to 1
  3. Little, Slegers, and Card (2006), or LSC method, constrain the average (mean) loading of a set of indicators on their common factor to equal 1.0 in the unstandardized solution

Identification in Measurement Models

- A standard CFA model is identified if any of the following are true
  - It has at least 3 indicators per factor,
  - It has >= 2 factors and there are >= indicators per factor,

SEM Steps

1. Specify and re-Specify the Measurement Model
   - Remove ANY items with standardized loadings greater than 1
   - Remove items with weak standardized loadings
   - Make other modifications
2. Specify the Structural part of the Model
3. Re-Specify the Structural part of the Model
   - Trim any non-significant path estimates
   - Add any paths that are statistically significant and/or are needed for adequate model fit

Advice for Re-Specifying Models

- Lavaan provides output that suggests additions that could be made to a model called **modification indices**
- By evaluating the parameter estimates, identify linkages that could be removed.
- The order that parameters are added or removed can affect the significance of the remaining parameters.
  - removal and addition should be done one parameter at a time.

Note of Caution

- A modification of a model solely on the basis of the results of a previous model:
  - possibility of capitalizing on chance
- Must articulate a persuasive rationale as to why the modification is theoretically and practically defensible.
- Any respecified model should be tested on an independent sample.

To R!

Copy the Structural Equation Modeling code into an .Rmd file