Structural Equation Modeling (SEM)

Melinda K. Higgins, Ph.D.

13 April 2009

SEM – descriptions

- SEM is a collection of statistical techniques that allow a set of relationships between one or more IVs (independent variables) (either continuous or discrete) and one or more DVs (dependent variables) (also either continuous or discrete).
- SEM is also called causal modeling; causal analysis; simultaneous equation modeling; analysis of covariance structures; analysis of moments; path analysis or confirmatory factor analysis [last 2 are special types of SEM]
- (example) When you combine EFA (exploratory factor analysis) with multiple regression, you have SEM.
- Some recommend you begin with confirmatory factor analysis (where applicable), then evaluate the various "paths/regression relationships" and, finally, put it all together into one SEM model. [D. Garson]

EMORY UNIVERSITY

SEM – advantages and disadvantages

- ADV – When relationships among factors are examined the relationships are free from measurement error (because error has been estimated and removed leaving only common variance).
- ADV – Thus, reliability of measurement can be accounted for explicitly within the analysis by estimating and removing the measurement error.
- ADV – When phenomena of interest are complex and multidimensional, SEM is the only analysis that allows complete and simultaneous test of all the relationships.
- DISADV – with increased flexibility comes increased complexity (and demands for larger sample sizes).

Structural Equation Modeling

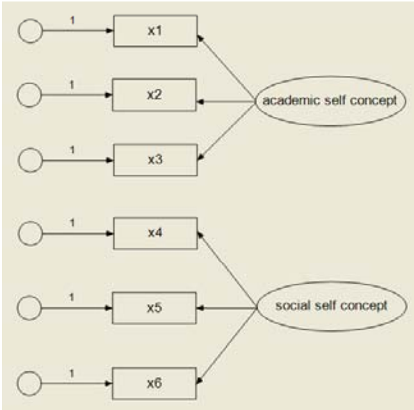
EMORY UNIVERSITY

"Types" of SEM

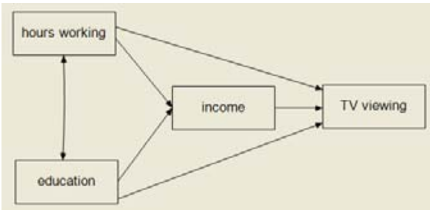
- Measurement Model/Confirmatory Factor Analysis
- Path Analysis
- Structural "Regression" Analysis
- Latent Change (or Growth) Models

Structural Equation Modeling

Confirmatory Factor Analysis

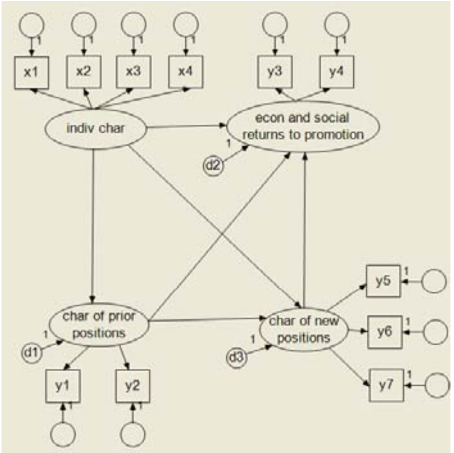


Path Model



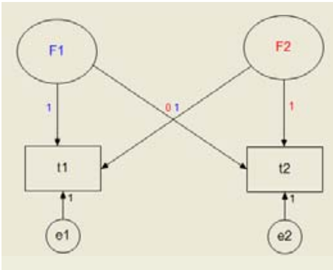
EMORY UNIVERSITY

SEM (CFA and Path Combined)



EMORY UNIVERSITY

Latent Change (Growth) Model



EMORY UNIVERSITY

SEM - goal

- Develop a model the "fits" the data (based on variance/covariance matrix)
  - H0: Model "fits" the data ("fit" function defined that compares sample cov matrix, S, to predicted cov matrix Σ)
  - Ha: Model does not "fit" the data
- Test via a χ^2 - want a non-significant test (p-val >> 0.05)
- χ^2 - is calculated by comparing estimated predicted covariance to sample covariance [ideally = 0, no difference between S and Σ]

Mediators – How to test for …

[Preacher, Hayes] – 3 approaches:

(1) Baron, Kenny: "suffers from low power"

(i) $Y = i_1 + cX$

(ii) $M = i_2 + aX$

(iii) $Y = i_3 + c'X + bM$

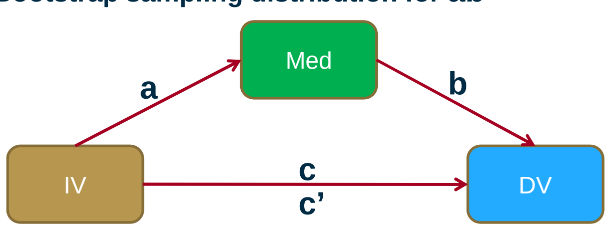
(2) Sobel Test

(i) calculate ab (assumption Normal Distribution)

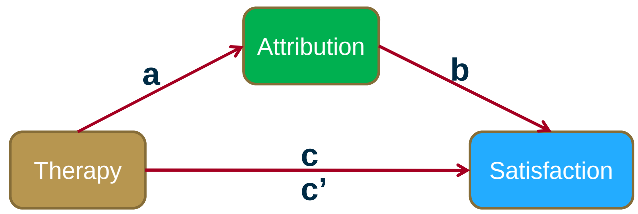
(ii) calculate $s_{ab} = \mathrm{sqrt}\,(b^2 s_a^2 + a^2 s_b^2 + s_a^2 s_b^2)$

(iii) divide ab/s_{ab} → compare to N(0,1) critical values

(3) Bootstrap sampling distribution for ab

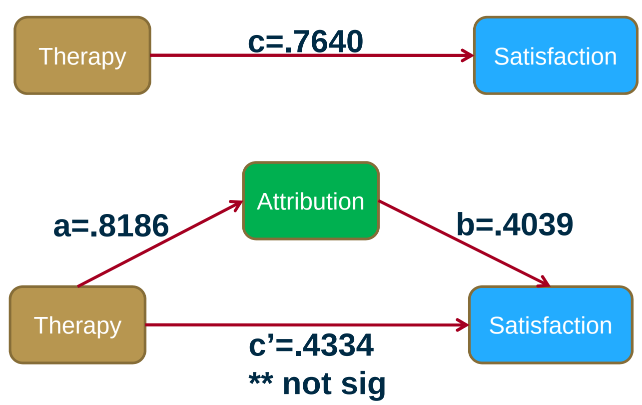


Example [Preacher,Hayes]



“An investigator is interested in the effects of a new cognitive therapy on life satisfaction after retirement. Residents of a retirement home diagnosed as clinically depressed are randomly assigned to receive 10 sessions of a new cognitive therapy or an alternative method. After session 8, the “positivity of attributions” the residents make for a recent failure experience is assessed. After session 10, the residents are given a measure of life satisfaction (questionnaire).”

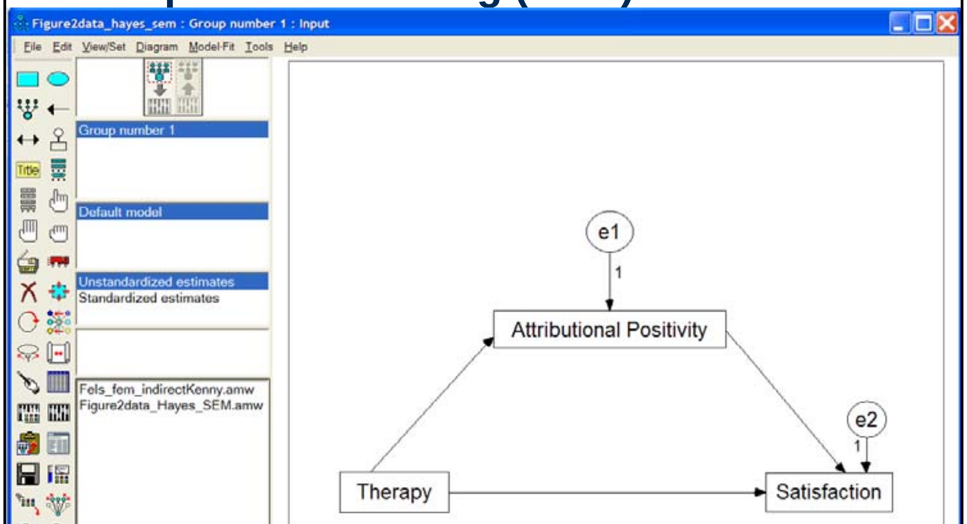
Results – SPSS Macro



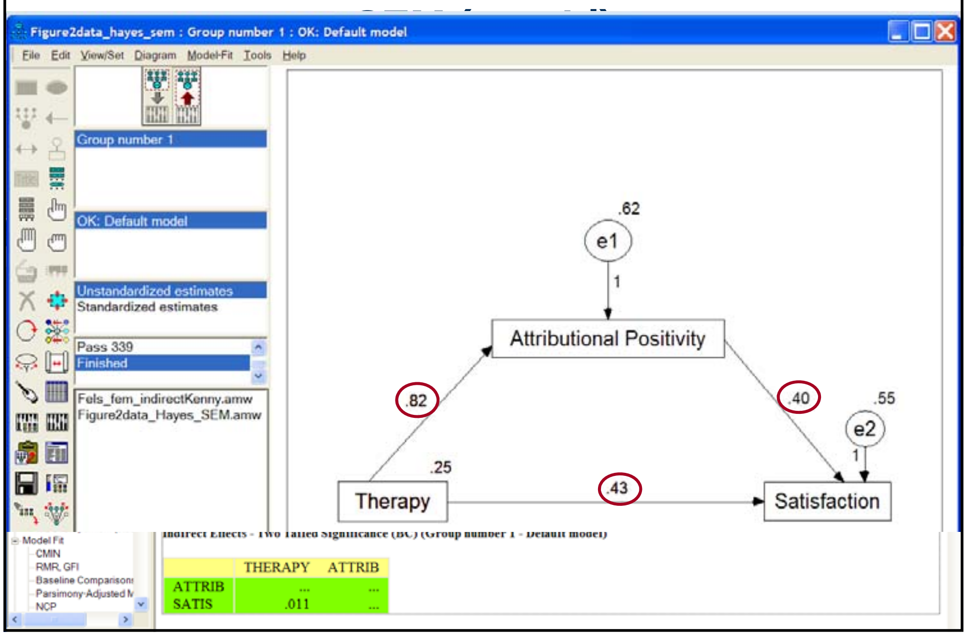
(Baron/Kenny) C’ is not significantly different from 0 (pval=0.1897)

(Sobel)	**ab = 0.3306**	**95% CI = [-0.0585, 0.7197] not significant**
(Bootstrap)	**ab = 0.3205**	**95% CI = [0.0334, 0.7008] significant**

Example (cont'd) – Mediation Using Structural Equation Modeling (SEM) via AMOS



See http://davidakenny.net/cm/mediate.htm and tutorial (video and sound) at http://amosdevelopment.com/video/indirect/flash/indirect.html



SEM – graphics and terms

Rectangles/squares = measured/observed/manifest variables – can be treated as DVs or IVs

Ovals/Circles = latent/unobserved variables – can be treated as DVs or IVs

Smaller Circles = unobserved “errors” or “residuals”

One way arrows show the direction of prediction*
- variable receiving path = “dependent variable/endogenous”
- variable originating path = “independent variable/exogenous”
- can be both (e.g. indirect effect/mediator)

Two way arrows indicate covariance

*NOTE: Latent variables ALWAYS “predict” Measured variables – the underlying construct/latent variables “drives or creates” the measured/observed variables.

“Measurement Model” portion – relates measured to latent variables (CFA)
“Structural Model” portion – relates constructs (usually) to each other

Example Data (Tabachnick et.al. Fig 14.4)

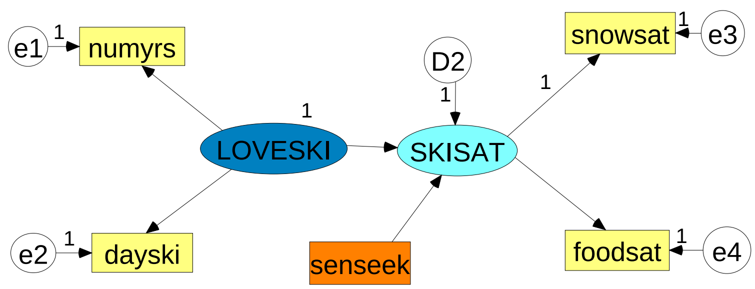
5 continuous measured variables:
- NUMYRS – number of years participant has skied
- DAYSKI – total number days person has skied
- SNOWSAT – Likert scale measure of overall satisfaction with snow conditions
- FOODSAT – Likert scale measure of overall satisfaction with quality of food at resort
- SENSEEK – Likert scale measure of degree of sensation seeking

2 hypothesized latent variables:
- LOVESKI – Love of Skiing
- SKISAT – Ski Trip Satisfaction

It is hypothesized that:
1. Love of Skiing “predicts” number of year skied and number of days skied
2. Ski Trip Satisfaction “predicts” degree of satisfaction with snow conditions and food quality at resort.
3. Love of Skiing and degree of sensation seeking “predict” level of Ski Trip Satisfaction

Figure on next page shows these relationships and the hypothesized model

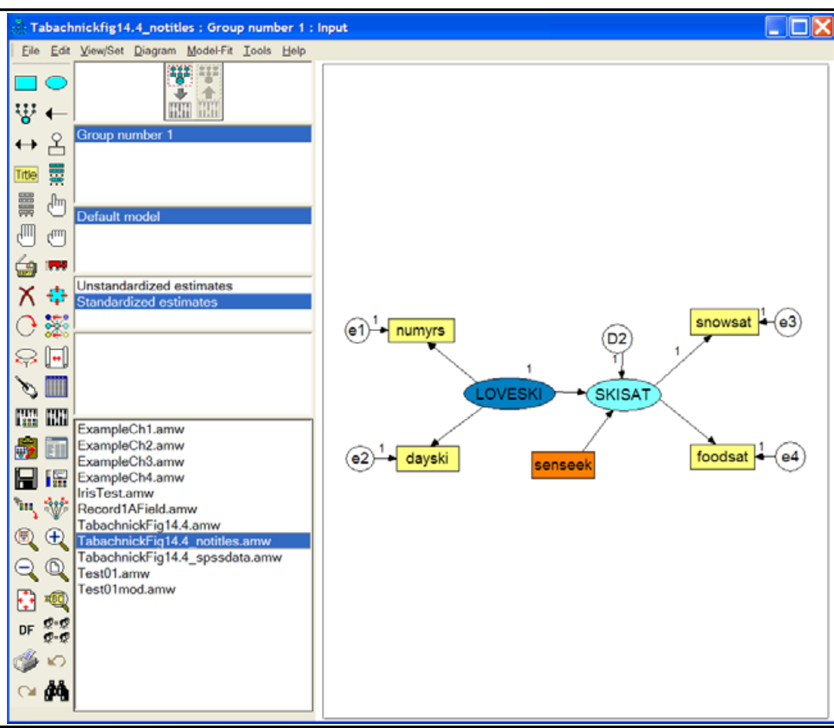
Hypothesized Model for "Small-Sample" Example
Figure 14.4 p.687 (Tabacknick, et.al. 2007)

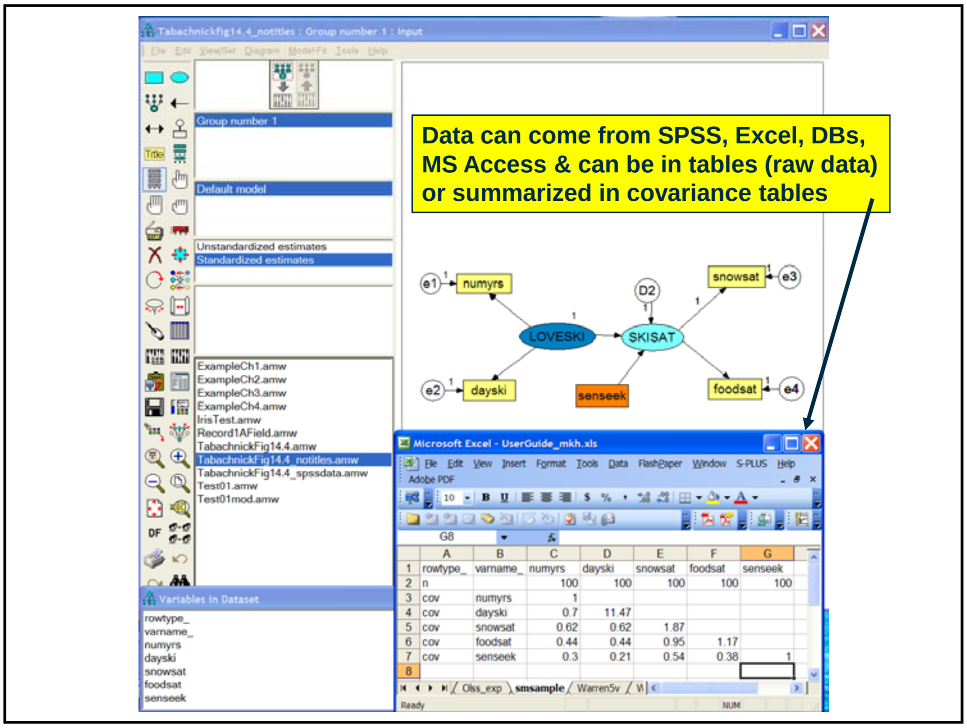


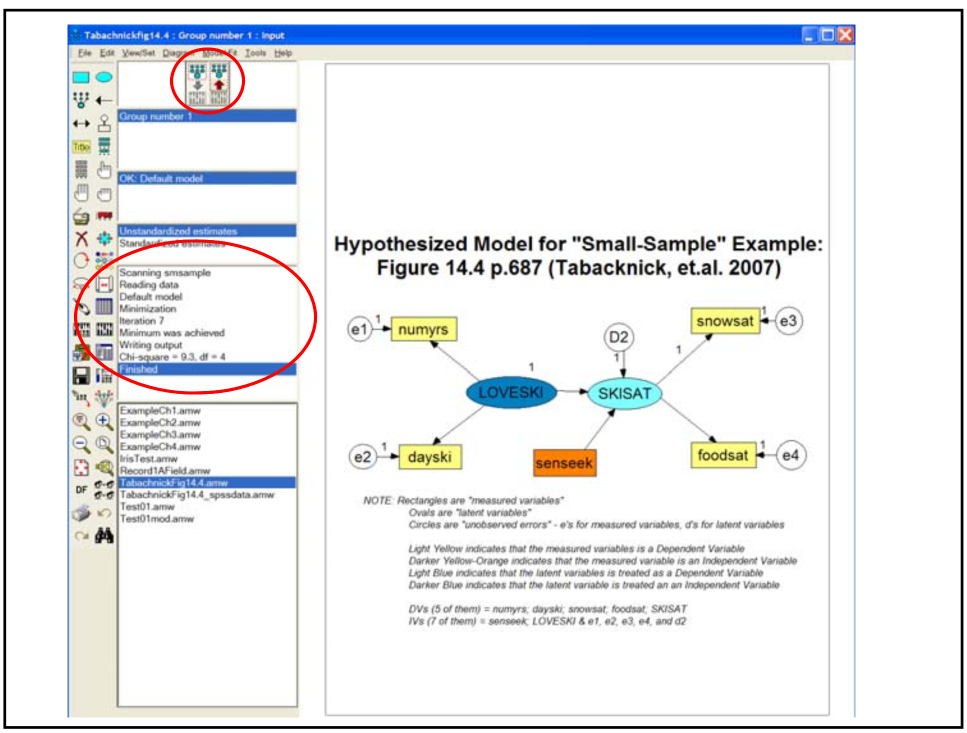
NOTE: Rectangles are "measured variables"
Ovals are "latent variables"
Circles are "unobserved errors" - e's for measured variables, d's for latent variables

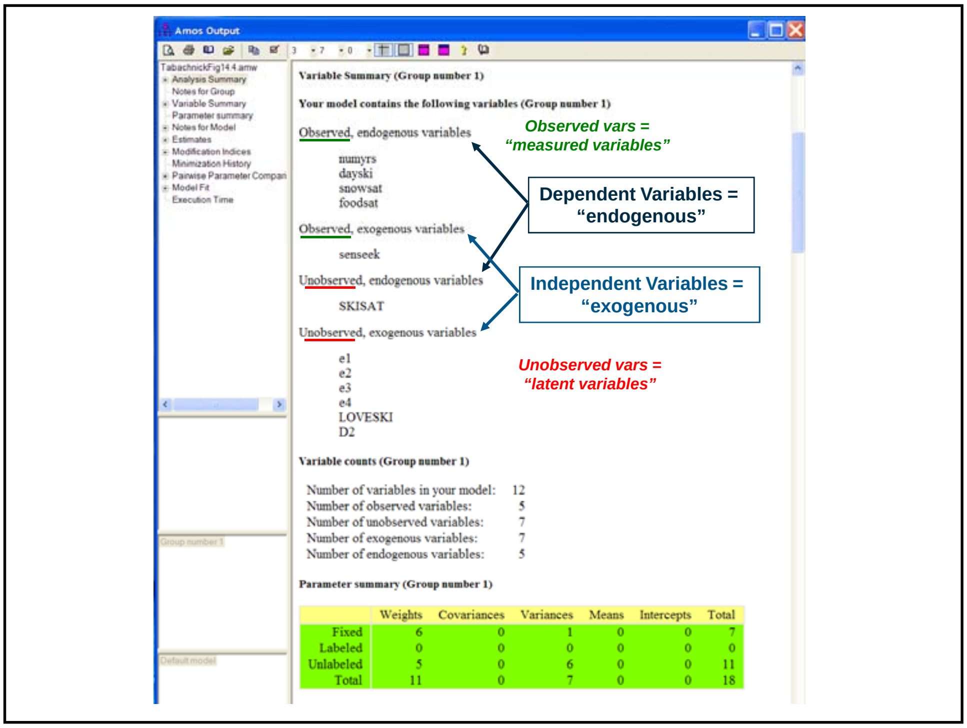
Light Yellow indicates that the measured variables is a Dependent Variable
Darker Yellow-Orange indicates that the measured variable is an Independent Variable
Light Blue indicates that the latent variables is treated as a Dependent Variable
Darker Blue indicates that the latent variable is treated an an Independent Variable

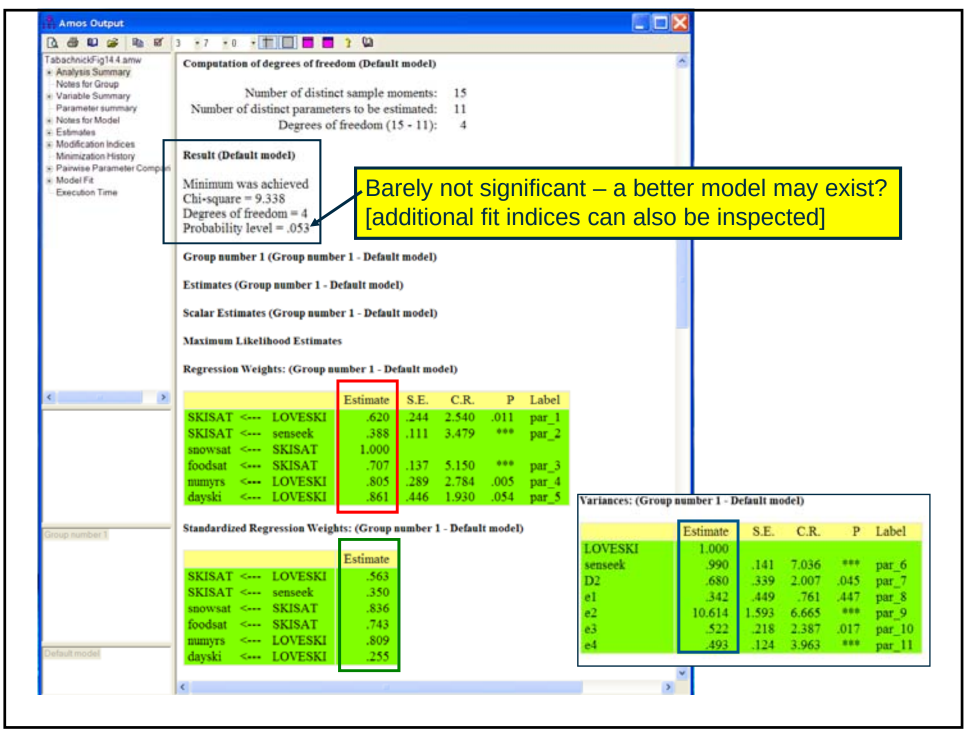
DVs (5 of them) = numyrs; dayski; snowsat; foodsat; SKISAT
IVs (7 of them) = senseek; LOVESKI & e1, e2, e3, e4, and d2




Data can come from SPSS, Excel, DBs, MS Access & can be in tables (raw data) or summarized in covariance tables
Group number 1
Default model
Unstandardized estimates
Standardized estimates
ExampleCh1.amw
ExampleCh2.amw
ExampleCh3.amw
ExampleCh4.amw
IrisTest.amw
Record1AField.amw
TabachnickFig14.4.amw
TabachnickFig14.4_notitles.amw
TabachnickFig14.4_spssdata.amw
Test01.amw
Test01mod.amw
Variables in Dataset
rowtype_
varname_
numyrs
dayski
snowsat
foodsat
senseek
e1
numyrs
e2
dayski
LOVESKI
D2
SKISAT
senseek
snowsat
e3
foodsat
e4
Microsoft Excel - UserGuide_mkh.xls
rowtype_ varname_ numyrs dayski snowsat foodsat senseek
n 100 100 100 100 100
cov numyrs 1
cov dayski 0.7 11.47
cov snowsat 0.62 0.62 1.87
cov foodsat 0.44 0.44 0.95 1.17
cov senseek 0.3 0.21 0.54 0.38 1
Olss_exp
smsample
Warren5v
Ready
NUM


Tabachnickfig14.4 : Group number 1 : Input
Group number 1
OK: Default model
Unstandardized estimates
Scanning smsample
Reading data
Default model
Minimization
Iteration 7
Minimum was achieved
Writing output
Chi-square = 9.3, df = 4
Finished
ExampleCh1.amw
ExampleCh2.amw
ExampleCh3.amw
ExampleCh4.amw
IrisTest.amw
Record1AField.amw
TabachnickFig14.4.amw
TabachnickFig14.4_spssdata.amw
Test01.amw
Test01mod.amw
Hypothesized Model for "Small-Sample" Example:
Figure 14.4 p.687 (Tabacknick, et.al. 2007)
e1
numyrs
D2
snowsat
e3
LOVESKI
SKISAT
e2
dayski
senseek
foodsat
e4
NOTE: Rectangles are "measured variables"
Ovals are "latent variables"
Circles are "unobserved errors" - e's for measured variables, d's for latent variables
Light Yellow indicates that the measured variables is a Dependent Variable
Darker Yellow-Orange indicates that the measured variable is an Independent Variable
Light Blue indicates that the latent variables is treated as a Dependent Variable
Darker Blue indicates that the latent variable is treated an an Independent Variable
DVs (5 of them) = numyrs; dayski; snowsat; foodsat; SKISAT
IVs (7 of them) = senseek; LOVESKI & e1, e2, e3, e4, and d2

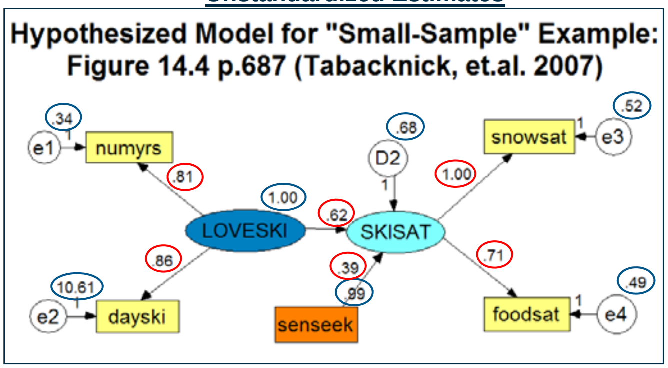

Amos Output
TabachnickFig14.4.amw
Analysis Summary
Notes for Group
Variable Summary
Parameter summary
Notes for Model
Estimates
Modification Indices
Minimization History
Pairwise Parameter Compari
Model Fit
Execution Time
Group number 1
Default model
Variable Summary (Group number 1)
Your model contains the following variables (Group number 1)
Observed, endogenous variables
numyrs
dayski
snowsat
foodsat
Observed, exogenous variables
senseek
Unobserved, endogenous variables
SKISAT
Unobserved, exogenous variables
e1
e2
e3
e4
LOVESKI
D2
Observed vars = "measured variables"
Dependent Variables = "endogenous"
Independent Variables = "exogenous"
Unobserved vars = "latent variables"
Variable counts (Group number 1)
Number of variables in your model: 12
Number of observed variables: 5
Number of unobserved variables: 7
Number of exogenous variables: 7
Number of endogenous variables: 5
Parameter summary (Group number 1)
	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	6	0	1	0	0	7
Labeled	0	0	0	0	0	0
Unlabeled	5	0	6	0	0	11
Total	11	0	7	0	0	18


Amos Output
Computation of degrees of freedom (Default model)
Number of distinct sample moments: 15
Number of distinct parameters to be estimated: 11
Degrees of freedom (15 - 11): 4
Result (Default model)
Minimum was achieved
Chi-square = 9.338
Degrees of freedom = 4
Probability level = .053
Barely not significant – a better model may exist? [additional fit indices can also be inspected]
Group number 1 (Group number 1 - Default model)
Estimates (Group number 1 - Default model)
Scalar Estimates (Group number 1 - Default model)
Maximum Likelihood Estimates
Regression Weights: (Group number 1 - Default model)
			Estimate	S.E.	C.R.	P	Label
SKISAT	<---	LOVESKI	.620	.244	2.540	.011	par_1
SKISAT	<---	senseek	.388	.111	3.479	***	par_2
snowsat	<---	SKISAT	1.000				
foodsat	<---	SKISAT	.707	.137	5.150	***	par_3
numyrs	<---	LOVESKI	.805	.289	2.784	.005	par_4
dayski	<---	LOVESKI	.861	.446	1.930	.054	par_5

Standardized Regression Weights: (Group number 1 - Default model)
			Estimate
SKISAT	<---	LOVESKI	.563
SKISAT	<---	senseek	.350
snowsat	<---	SKISAT	.836
foodsat	<---	SKISAT	.743
numyrs	<---	LOVESKI	.809
dayski	<---	LOVESKI	.255

Variances: (Group number 1 - Default model)
	Estimate	S.E.	C.R.	P	Label
LOVESKI	1.000				
senseek	.990	.141	7.036	***	par_6
D2	.680	.339	2.007	.045	par_7
e1	.342	.449	.761	.447	par_8
e2	10.614	1.593	6.665	***	par_9
e3	.522	.218	2.387	.017	par_10
e4	.493	.124	3.963	***	par_11

Unstandardized Estimates

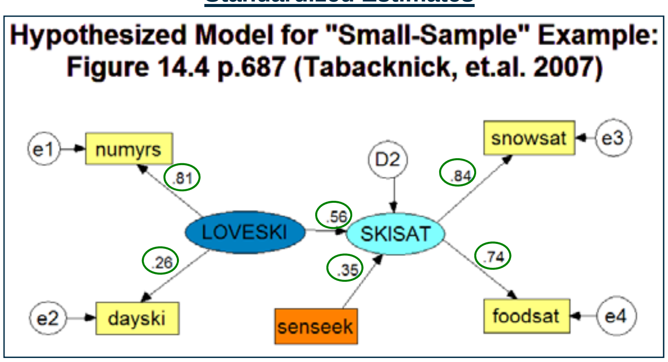


Numbers in circled in blue are variances

Numbers in circled in red are (unstandardized) regression coefficients

1. NUMYRS and DAYSKI are both significant indictors of LOVESKI [see Tabachnick, et.al. p.695 for z-score tests on regression parameters – all have pval < or = to 0.05]
2. FOODSAT is a significant indicator of SKISAT
3. [Since SKISAT → SNOWSAT was set=1, this significance cannot be determined – rerun with SKISAT → FOODSAT set=1]
4. SENSEEK is a significant indicator of SKISAT
5. LOVESKI is a significant indicator of SKISAT

Standardized Estimates



Now all variances = 1 [not shown]

Numbers in circled in green are (standardized) regression coefficients

1. Standardized regression coefficients can be compared to each other since the “scales” of the variables have been removed.
2. E.g. can compare that NUMYRS is a stronger indicator (.81) of LOVESKI than DAYSKI (.26) although both were tested to be significant indicators.
3. Likewise both SNOWSAT and FOODSAT are almost equivalently strong indicators of SKISAT
4. LOVESKI while significant is a weaker indicator (.56) of SKISAT followed by SENSEEK with an even lower standardized regression weight of 0.35.

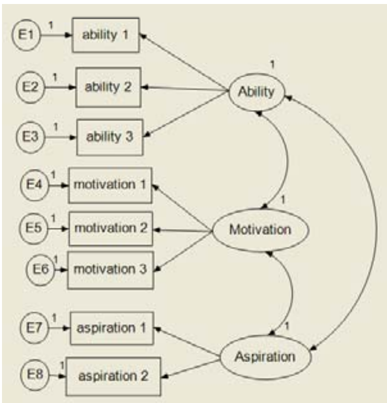
EMORY UNIVERSITY

6 rules

1. All variances of independent variables are model parameters
2. All covariances between independent variables are model parameters
3. All factor loadings are model parameters
4. All regression coefficients between observed or latent variables are model parameters
5. Variances and covariances between dependent variables and/or covariances between dependent variables and independent variables are NEVER model parameters (because these variances and covariances are themselves explained in terms of other model parameters)
6. For each latent variable, the "metric" of its latent "scale" needs to be set. [usually either the variance = constant (1) or the path leaving the latent variable = constant (1).]

EMORY UNIVERSITY

Measurement Model Confirmatory Factor Analysis



CHECK 6 RULES

(1) 8 error variances (e1 → e8) and 3 factor variances (1)
(2) 3 factor covariances
(3) 8 factor loadings
(4) n/a – no regression relationships
(5) No 2-way arrows between DVs nor between IV-DV.
(6) Metric of 3 factors is "fixed" by setting their variances = 1 (see (1) above).

So, there are 8 + 3 + 8 = 19 "free" parameters to estimate

EMORY UNIVERSITY

Degrees of Freedom (“Identification”)

For a model to be “identified,” there must be the same or fewer number of parameters than non-redundant elements in the covariance matrix (i.e. degrees of freedom >= 0).

DF = [P(P+1)/2] – number of parameters to be estimated

DF (example) = [8(9)/2] – 19 = 36 – 19 = 17 degrees of freedom

NOTE: If DF=0, the model is said to be saturated.

Structural Equation Modeling

Your model contains the following variables (Group number 1)

Observed, endogenous variables

ab1 (ability 1)
ab2 (ability 2)
ab3 (ability 3)
mot1 (motivation 1)
mot2 (motivation 2)
mot3 (motivation 3)
asp1 (aspiration 1)
asp2 (aspiration 2)

Unobserved, exogenous variables

E1
E2
E3
E4
E5
E6
E7
E8
F2 (Motivation)
F3 (Aspiration)
F1 (Ability)

Variable counts (Group number 1)

Number of variables in your model:	19
Number of observed variables:	8
Number of unobserved variables:	11
Number of exogenous variables:	11
Number of endogenous variables:	8

Parameter summary (Group number 1)

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	8	0	3	0	0	11
Labeled	0	0	0	0	0	0
Unlabeled	8	3	8	0	0	19
Total	16	3	11	0	0	30

Computation of degrees of freedom (Default model)

Number of distinct sample moments:	36
Number of distinct parameters to be estimated:	19
Degrees of freedom (36 - 19):	17

Result (Default model)

Minimum was achieved
Chi-square = 20.581
Degrees of freedom = 17
Probability level = .246

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
mot1 (motivation 1)	<---	F2 (Motivation)	.510	.045	11.323	***	par_1
mot2 (motivation 2)	<---	F2 (Motivation)	.592	.049	12.189	***	par_2
mot3 (motivation 3)	<---	F2 (Motivation)	.594	.046	12.835	***	par_3
asp1 (aspiration 1)	<---	F3 (Aspiration)	.613	.049	12.492	***	par_4
asp2 (aspiration 2)	<---	F3 (Aspiration)	.634	.051	12.486	***	par_5
ab3 (ability 3)	<---	F1 (Ability)	.520	.039	13.452	***	par_9
ab2 (ability 2)	<---	F1 (Ability)	.614	.042	14.464	***	par_10
ab1 (ability 1)	<---	F1 (Ability)	.518	.039	13.376	***	par_11

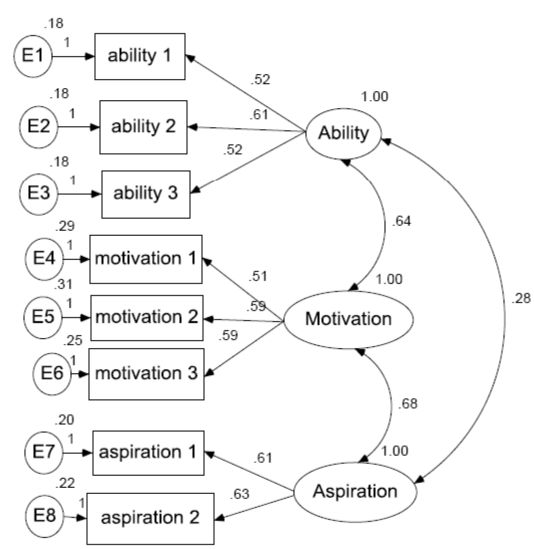
Standardized Regression Weights: (Group number 1 - Default model)

			Estimate
mot1 (motivation 1)	<---	F2 (Motivation)	.690
mot2 (motivation 2)	<---	F2 (Motivation)	.731
mot3 (motivation 3)	<---	F2 (Motivation)	.762
asp1 (aspiration 1)	<---	F3 (Aspiration)	.807
asp2 (aspiration 2)	<---	F3 (Aspiration)	.806
ab3 (ability 3)	<---	F1 (Ability)	.777
ab2 (ability 2)	<---	F1 (Ability)	.822
ab1 (ability 1)	<---	F1 (Ability)	.773

Covariances: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
F2 (Motivation)	<-->	F1 (Ability)	.636	.055	11.659	***	par_6
F2 (Motivation)	<-->	F3 (Aspiration)	.681	.054	12.554	***	par_7
F3 (Aspiration)	<-->	F1 (Ability)	.276	.074	3.760	***	par_8

EMORY UNIVERSITY



Structural Equation Modeling

EMORY UNIVERSITY

References

- Tabachnick, Barbara G.; Fidell, Linda S. "Using Multivariate Statistics," 5th edition, Pearson Education Inc., 2007. {"Chapter 14: SEM" good worked examples – with discussion of covariance math}
- Raykov, Tenko; Marcoulides, George. "A First Course in Structural Equation Modeling." Lawrence Erlbaum Associates Publishers, New Jersey, 2000. {good examples with LISREL and EQS codes and some info on AMOS}
- Kline, Rex. "Principles and Practice of Structural Equation Modeling." The Guilford Press, New York, 1998. {some info on AMOS, EQS and LISREL}
- Kaplan, David. "Structural Equation Modeling – Foundations and Extensions." 2nd edition, SAGE Publications, Los Angeles, 2009. {more advanced text, but good indepth discussions}
- Duncan, T; Duncan, S; Strycker, L; Li, F; Alpert, A. "An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues and Applications." Lawrence Erlbaum Associates Publishers, New Jersey, 1999. {also good worked examples with EQS and LISREL codes}

EMORY UNIVERSITY

VIII. Statistical Resources and Contact Info

SON S:\Shared\Statistics_MKHiggins\website2\index.htm

[updates in process]

Working to include tip sheets (for SPSS, SAS, and other software), lectures (PPTs and handouts), datasets, other resources and references

Statistics At Nursing Website: [website being updated] http://www.nursing.emory.edu/pulse/statistics/

And Blackboard Site (in development) for "Organization: Statistics at School of Nursing"

Contact

Dr. Melinda Higgins

Melinda.higgins@emory.edu

Office: 404-727-5180 / Mobile: 404-434-1785