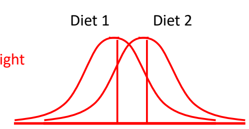
LATENT MEANS MODELS

1

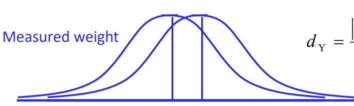
Introduction: Measurement error and Effect Size

True weight



$$d_T = \frac{|\mu_{1T} - \mu_{2T}|}{\sigma_T}$$

Measured weight

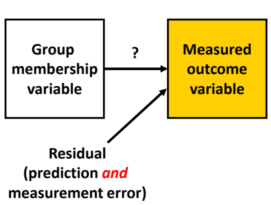


$$d_Y = \frac{|\mu_{1Y} - \mu_{2Y}|}{\sigma_Y}$$

- Measurement error leads to:
  - underestimation of the standardized effect size;
  - decreased power to detect treatment/group differences.

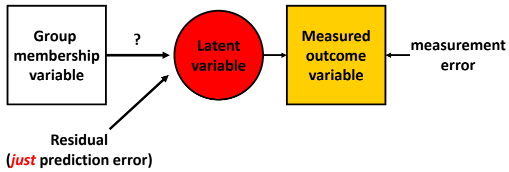
2

Introduction: GLM/ANOVA Outcome



3

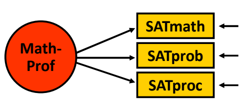
Introduction:
Desired GLM/ANOVA Outcome



4

Introduction:
Example Research Question

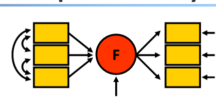
- To diminish the unreliability problem, the ANOVA paradigm can be circumvented: Population differences in the means of a **latent** variable can be modeled. For example …
- … Do 9th grade boys and girls differ in their Mathematics Proficiency, the latent variable believed to underlie scores on the Mathematics, Problem Solving, and Procedure subtests of the Stanford Achievement Test 9?



5

Group Code Analysis

- Special case of MIMIC (Multiple-Indicator, MultIple-Cause) modeling

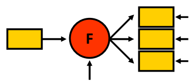


- Similar to conducting a regression analysis with a dummy-coded predictor to address an ANOVA-type question:
  - the outcome variable now is a latent factor, not a measured variable;
  - the analysis is conducted on a combined data set; homogeneity of dispersion across groups must be assumed; mean values on measured variables are not explicitly required.

6

Group Code Analysis vs. Structured Means

- The group code approach uses combined group data, inferring a latent mean difference by the degree of relation of the group code (dummy) variable to the measured outcomes.



- This method assumes invariance of ***all*** covariance structure parameters, down to the last error variance.
- Structured means analysis, although more complex, is also more flexible.
- The structured means analysis approach to assessing latent group differences uses separate group data, inferring a latent mean difference directly from differences in the observed means.

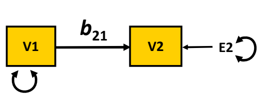
7

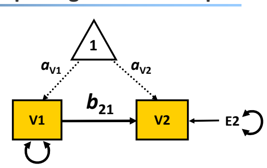
Structured Means Analysis

- Recall from linear regression: $Y = a + bX + e$.
  The unstandardized slope parameter b can be expressed as $b = \sigma_{XY}/\sigma_X^2$, thus requiring no direct mean information.
- In fact, in all of covariance structure modeling (of which linear regression is a special case), means of observed variables are irrelevant; thus, often data are viewed as though they have zero means (i.e., have been mean-centered).
- The intercept parameter a can be expressed as $a = \mu_Y - b\mu_X$, thereby involving information about the variable means.
- Thus, in order to integrate information about variable means into a model, intercept terms must be introduced (back) into our structural equations and modeled accordingly. In turn, this will allow factor means to be modeled.

8

Measured Mean Structure: A simple regression example





$V2 = b_{21}V1 + E2$

"structural equation"

$V2 = a_{V2} + b_{21}V1 + E2$

$V2 = a_{V2}(1) + b_{21}V1 + E2$

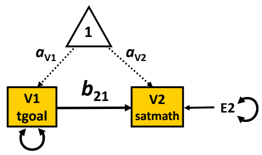
a_{V2} is an intercept term

a_{V1} is a mean

9

Measured Mean Structure: Mplus Syntax

```
DATA:
  FILE IS simple_data.csv;
VARIABLE:
  NAMES ARE tgoal satmath;
MODEL:
  satmath ON tgoal;
  tgoal; satmath;        } default
  [tgoal]; [satmath];    }
OUTPUT:
  SAMPSTAT STDYX;
```



Data from n=1000 9th grade girls on task goal orientation (tgoal) and Stanford Achievement Test math score (satmath).

10

Measured Mean Structure: Mplus Output

```
MODEL FIT INFORMATION

Chi-Square Test of Model Fit

          Value                                0.000
          Degrees of Freedom                       0
          P-Value                             0.0000

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                             0.000
          90 Percent C.I.                      0.000   0.000
          Probability RMSEA <= .05             0.000

CFI/TLI

          CFI                                  1.000
          TLI                                  1.000

SRMR (Standardized Root Mean Square Residual)

          Value                                0.000
```

With mean structures, incremental fit indices are generally ill-advised unless a null model is computed separately and comparative indices are hand-derived.

11

Measured Mean Structure: Mplus Output

```
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
 SATMATH  ON
    TGOAL             -1.428      0.920     -1.553      0.120

 Means
    TGOAL              3.308      0.040     82.160      0.000

 Intercepts
    SATMATH          690.063      3.259    211.711      0.000

 Variances
    TGOAL              1.621      0.072     22.361      0.000

 Residual Variances
    SATMATH         1370.822     61.307     22.360      0.000


STDYX Standardization
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
 SATMATH  ON
    TGOAL             -0.049      0.032     -1.555      0.120
```

12

Measured Mean Structure:
Mplus Output / SPSS Output

```
                                                   Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
 SATMATH  ON
    TGOAL             -1.428      0.920     -1.553      0.120

 Intercepts
    SATMATH          690.063      3.259    211.711      0.000

STDYX Standardization
                                                   Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
 SATMATH  ON
    TGOAL             -0.049      0.032     -1.555      0.120
```

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	690.063	3.263		211.503	.000
	tgoal	-1.428	.920	-.049	-1.551	.121

a. Dependent Variable: satmath

13

Latent Mean Structure:
A simple one-factor example

Structural equations

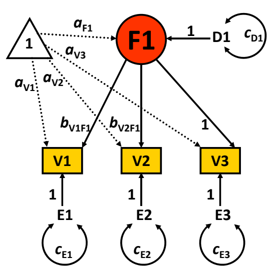
V1 = a_{V1}(1) + b_{V1F1}F1 + E1

V2 = a_{V2}(1) + b_{V2F1}F1 + E2

V3 = a_{V3}(1) + 1 F1 + E3

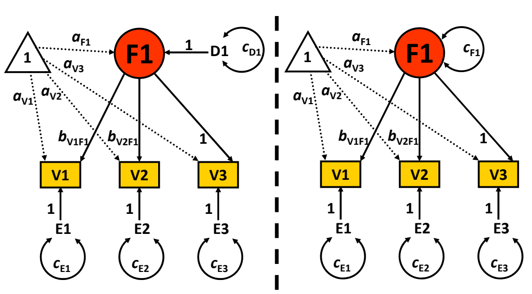
F1 = a_{F1}(1) + D1

represents the latent/factor mean



14

Two Equivalent Representations



15

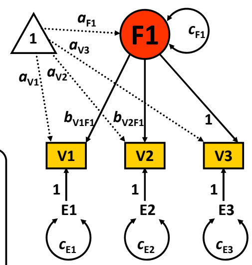
Model-Implied Covariance Matrix

Structural equations

V1 = $a_{V1}(1) + b_{V1F1}$F1 + E1

V2 = $a_{V2}(1) + b_{V2F1}$F1 + E2

V3 = $a_{V3}(1) + 1$ F1 + E3



Model-Implied Covariance Matrix with Six Unknowns

$$\begin{bmatrix} b^2_{V1F1}c_{F1} + c_{E1} & & \\ b_{V1F1}b_{V2F1}c_{F1} & b^2_{V2F1}c_{F1} + c_{E2} & \\ b_{V1F1}c_{F1} & b_{V2F1}c_{F1} & c_{F1} + c_{E3} \end{bmatrix}$$

16

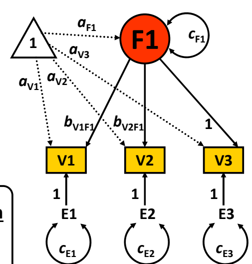
Model-Implied Mean Vector

Structural equations

V1 = $a_{V1}(1) + b_{V1F1}$F1 + E1

V2 = $a_{V2}(1) + b_{V2F1}$F1 + E2

V3 = $a_{V3}(1) + 1$ F1 + E3



Model-Implied Mean Vector with Four Additional Unknowns

$$\begin{bmatrix} a_{V1} + b_{V1F1}a_{F1} & a_{V2} + b_{V2F1}a_{F1} & a_{V3} + a_{F1} \end{bmatrix}$$

17

Solving for the Unknown Parameters: An Identification Problem

Model-implied covariance matrix = Observed covariance matrix

$$\begin{bmatrix} b^2_{V1F1}c_{F1} + c_{E1} & & \\ b_{V1F1}b_{V2F1}c_{F1} & b^2_{V2F1}c_{F1} + c_{E2} & \\ b_{V1F1}c_{F1} & b_{V2F1}c_{F1} & c_{F1} + c_{E3} \end{bmatrix} = \begin{bmatrix} s^2_1 & & \\ s_{21} & s^2_2 & \\ s_{31} & s_{32} & s^2_3 \end{bmatrix}$$

Model-implied mean vector = Observed mean vector

$$\begin{bmatrix} a_{V1} + b_{V1F1}a_{F1} & a_{V2} + b_{V2F1}a_{F1} & a_{V3} + a_{F1} \end{bmatrix} = \begin{bmatrix} \overline{V1} & \overline{V2} & \overline{V3} \end{bmatrix}$$

9 equations: 6 (covariance structure) + 3 (mean structure)
10 unknowns: 6 (covariance structure) + 4 (mean structure)

Thus, the mean structure currently is under-identified.

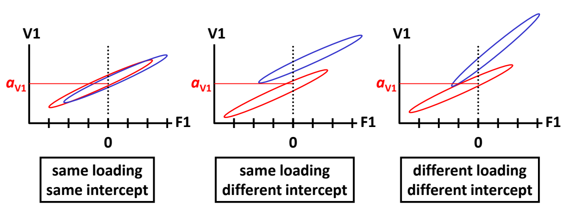
18

Solving the Identification Problem—Partially

- When comparing means across two groups it is assumed, at least initially, that
  - loadings are equal groups/time (i.e., that equal amounts of change in the factor lead to equal amounts of change in the indicators) and
  - intercepts are equal across groups/time (i.e., that equal amounts of the factor lead to equal amounts of the indicators).

19

Solving the Identification Problem—Partially



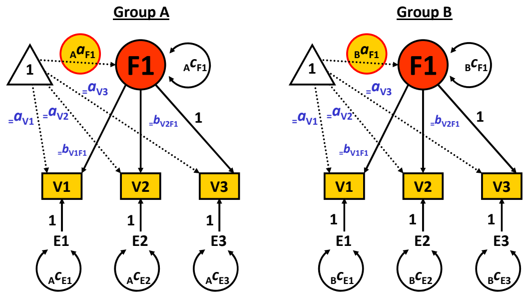
20

Solving the Identification Problem—Partially

- These initial assumptions of invariant loadings and intercepts help, partially, to solve the under-identification problem in the means portion of the model because now there are fewer unconstrained parameters to estimate.
- They will be testable (to a point).

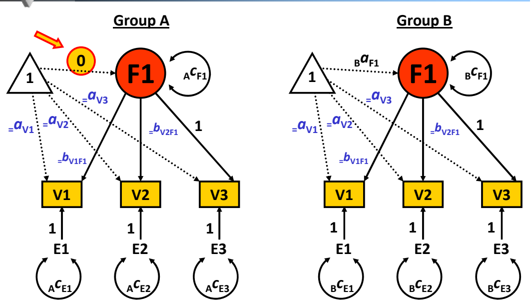
21

The Constrained Means Model Across Two Groups



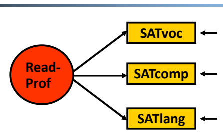
22

Assessing Latent Mean Differences by Using a Reference Group



23

Do Boys and Girls Differ on Mean Latent Reading Proficiency?



Girls (n_F = 1,000)

SATvoc SATcomp SATlang

$$\mathbf{S} = \begin{bmatrix} 2105.3 & & \\ 1420.7 & 1575.0 & \\ 1146.7 & 991.5 & 1472.1 \end{bmatrix}$$

means = 709.47 696.67 669.17

Boys (n_M = 1,000)

SATvoc SATcomp SATlang

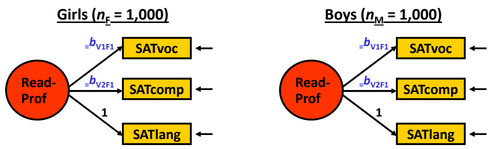
$$\mathbf{S} = \begin{bmatrix} 2323.0 & & \\ 1503.5 & 1625.7 & \\ 1192.5 & 1017.7 & 1530.9 \end{bmatrix}$$

means = 692.36 680.63 654.31

25

I. Are Factor Loadings Invariant?

- First, use a multi-group CFA to assess the measurement model fit across groups. Here, data-model fit is good with the loadings constrained to be equal across groups.

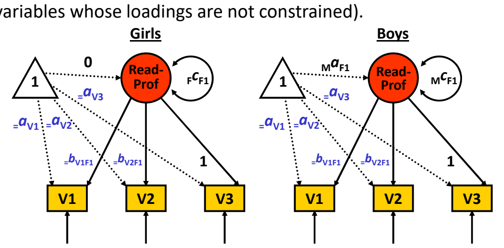


- If a loading were judged to be noninvariant (e.g., through modification indices), its constraint should be released.

26

II. Are Intercepts Invariant?

- Second, impose a model with mean structure and assess the intercept constraints (but do not impose intercept constraints on variables whose loadings are not constrained).



- If an intercept is judged to be noninvariant (e.g., through modification indices), its constraint should be released.

27

The Unknown Parameters

${}_{=}b_{V1F1}$, ${}_{=}b_{V2F1}$ } **loadings**

${}_{F}c_{E1}$, ${}_{M}c_{E1}$, ${}_{F}c_{E2}$, ${}_{M}c_{E2}$, ${}_{F}c_{E3}$, ${}_{M}c_{E3}$, ${}_{F}c_{F1}$, ${}_{M}c_{F1}$ } **variances**

} **From covariance structure: 10 unknown parameters**

${}_{=}a_{V1}$, ${}_{=}a_{V2}$, ${}_{=}a_{V3}$ } **intercepts**

${}_{M}a_{F1}$ } **latent mean**

} **From mean structure: 4 unknown parameters**

14 parameters

28

Estimating the Covariance Structure

Model-implied covariance matrices | **Observed covariance matrices**

Girls

$$\begin{bmatrix} b^2_{\mathrm{V1F1}\,\mathrm{F}}c_{\mathrm{F1}} + {}_{\mathrm{F}}c_{\mathrm{E1}} & & \\ b_{\mathrm{V1F1}}b_{\mathrm{V2F1}\,\mathrm{F}}c_{\mathrm{F1}} & b^2_{\mathrm{V2F1}\,\mathrm{F}}c_{\mathrm{F1}} + {}_{\mathrm{F}}c_{\mathrm{E2}} & \\ b_{\mathrm{V1F1}\,\mathrm{F}}c_{\mathrm{F1}} & b_{\mathrm{V2F1}\,\mathrm{F}}c_{\mathrm{F1}} & {}_{\mathrm{F}}c_{\mathrm{F1}} + {}_{\mathrm{F}}c_{\mathrm{E3}} \end{bmatrix} = \begin{bmatrix} {}_{\mathrm{F}}s_1^2 & & \\ {}_{\mathrm{F}}s_{21} & {}_{\mathrm{F}}s_2^2 & \\ {}_{\mathrm{F}}s_{31} & {}_{\mathrm{F}}s_{32} & {}_{\mathrm{F}}s_3^2 \end{bmatrix}$$

Boys

$$\begin{bmatrix} b^2_{\mathrm{V1F1}\,\mathrm{M}}c_{\mathrm{F1}} + {}_{\mathrm{M}}c_{\mathrm{E1}} & & \\ b_{\mathrm{V1F1}}b_{\mathrm{V2F1}\,\mathrm{M}}c_{\mathrm{F1}} & b^2_{\mathrm{V2F1}\,\mathrm{M}}c_{\mathrm{F1}} + {}_{\mathrm{M}}c_{E2} & \\ b_{\mathrm{V1F1}\,\mathrm{M}}c_{\mathrm{F1}} & b_{\mathrm{V2F1}\,\mathrm{M}}c_{\mathrm{F1}} & {}_{\mathrm{M}}c_{\mathrm{F1}} + {}_{\mathrm{M}}c_{\mathrm{E3}} \end{bmatrix} = \begin{bmatrix} {}_{\mathrm{M}}s_1^2 & & \\ {}_{\mathrm{M}}s_{21} & {}_{\mathrm{M}}s_2^2 & \\ {}_{\mathrm{M}}s_{31} & {}_{\mathrm{M}}s_{32} & {}_{\mathrm{M}}s_3^2 \end{bmatrix}$$

12 equations

29

Estimating the Mean Structure

Model-implied mean vectors | **Observed mean vectors**

Girls $$\left[a_{\mathrm{V1}} + b_{\mathrm{V1F1}}(0) \quad a_{\mathrm{V2}} + b_{\mathrm{V2F1}}(0) \quad a_{\mathrm{V3}} + (0)\right] = \left[{}_{\mathrm{F}}\overline{\mathrm{V1}} \quad {}_{\mathrm{F}}\overline{\mathrm{V2}} \quad {}_{\mathrm{F}}\overline{\mathrm{V3}}\right]$$

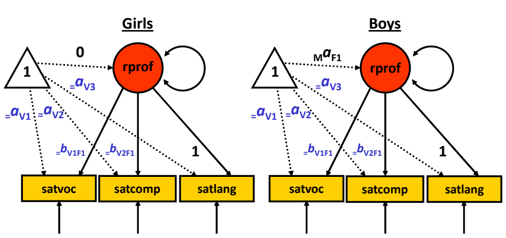
Boys $$\left[a_{\mathrm{V1}} + b_{\mathrm{V1F1}}({}_{\mathrm{M}}a_{\mathrm{F1}}) \quad a_{\mathrm{V2}} + b_{\mathrm{V2F1}}({}_{\mathrm{M}}a_{\mathrm{F1}}) \quad a_{\mathrm{V3}} + ({}_{\mathrm{M}}a_{\mathrm{F1}})\right] = \left[{}_{\mathrm{M}}\overline{\mathrm{V1}} \quad {}_{\mathrm{M}}\overline{\mathrm{V2}} \quad {}_{\mathrm{M}}\overline{\mathrm{V3}}\right]$$

6 equations

Thus, u = **18 = 12 (covariance structure) + 6 (mean structure)**
t = **14 parameters to be estimated**
df = $u - t$ **= 18 – 14 = 4**

30

The Model for the Software



31

Running Mplus: Model Syntax (separate data files)

```
DATA:
  FILE (female) IS means_data_girls.txt;
  FILE (male) IS means_data_boys.txt;
VARIABLE:
  NAMES ARE satvoc satcomp satlang;
MODEL:
  rprof BY satvoc* satcomp* satlang@1;
MODEL female:
  [rprof@0];
MODEL male:
  [rprof];
OUTPUT:
  sampstat modindices(3.841);
```

32

Running Mplus: Model Syntax (combined data file)

```
DATA:
  FILE IS means_data_girls_boys.txt;
VARIABLE:
  NAMES ARE gender satvoc satcomp satlang;
  GROUPING IS gender(1=female 2=male);
MODEL:
  rprof BY satvoc* satcomp* satlang@1;
MODEL female:
  [rprof@0];
MODEL male:
  [rprof];
OUTPUT:
  sampstat modindices(3.841);
```

33

Running Mplus: Data-model fit

```
MODEL FIT INFORMATION

Chi-Square Test of Model Fit
          Value                              5.174
          Degrees of Freedom                     4
          P-Value                           0.2699

Chi-Square Contributions From Each Group
          GIRLS                              2.142
          BOYS                               3.031

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.017
          90 Percent C.I.                    0.000  0.053
          Probability RMSEA <= .05           0.928

CFI/TLI
          CFI                                1.000
          TLI                                0.999
```

Generally ill-advised unless a null model is computed separately and comparative indices are hand-derived.

34

Running Mplus: Parameter estimates (girls)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
RPROF BY				
SATVOC	1.433	0.038	38.079	0.000
SATCOMP	1.240	0.033	38.104	0.000
SATLANG	1.000	0.000	999.000	999.000
Means				
RPROF	0.000	0.000	999.000	999.000
Intercepts				
SATVOC	710.029	1.410	503.654	0.000
SATCOMP	696.598	1.220	570.964	0.000
SATLANG	668.183	1.101	606.882	0.000
Variances				
RPROF	798.589	51.525	15.499	0.000
Residual Variances				
SATVOC	463.593	39.922	11.612	0.000
SATCOMP	345.174	29.923	11.536	0.000
SATLANG	672.466	35.147	19.133	0.000

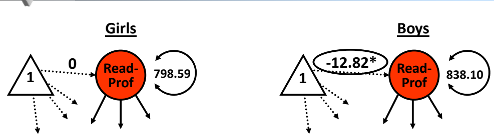
35

Running Mplus: Parameter estimates (boys)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
RPROF BY				
SATVOC	1.433	0.038	38.079	0.000
SATCOMP	1.240	0.033	38.104	0.000
SATLANG	1.000	0.000	999.000	999.000
Means				
RPROF	-12.816	1.383	-9.266	0.000
Intercepts				
SATVOC	710.029	1.410	503.654	0.000
SATCOMP	696.598	1.220	570.964	0.000
SATLANG	668.183	1.101	606.882	0.000
Variances				
RPROF	838.099	54.031	15.512	0.000
Residual Variances				
SATVOC	575.527	44.982	12.795	0.000
SATCOMP	336.133	31.306	10.737	0.000
SATLANG	720.011	37.711	19.093	0.000

36

Do Boys and Girls Differ on Mean Latent Reading Proficiency?



Model χ^2 = 5.17, df = 4, p = 0.27

RMSEA = 0.017, CI: (.00, .053)

*p<.05

- The fit results implicitly support intercepts' invariance across groups
- The mean latent reading proficiency is statistically significantly higher for girls than for boys.

37

Estimated Standardized Effect Size: Hancock (2001)

- **Estimated standardized effect size:**

$$\hat{d} = \frac{\left| {}_{\text{F}}\hat{a}_{\text{F1}} - {}_{\text{M}}\hat{a}_{\text{F1}} \right|}{\sqrt{\hat{c}_{\text{F1}}}} = \frac{\left| {}_{\text{M}}\hat{a}_{\text{F1}} \right|}{\sqrt{\hat{c}_{\text{F1}}}} \Rightarrow \hat{c}_{\text{F1}} = \frac{n_{\text{F}}({}_{\text{F}}\hat{c}_{\text{F1}}) + n_{\text{M}}({}_{\text{M}}\hat{c}_{\text{F1}})}{n_{\text{F}} + n_{\text{M}}}$$

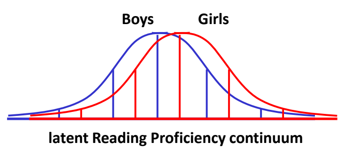
= pooled variance estimate of F1

$$\hat{d} = \frac{\left| -12.82 \right|}{\sqrt{1000\,(798.59) + 1000\,(838.10) \big/ (1000 + 1000)}} = .45$$

38

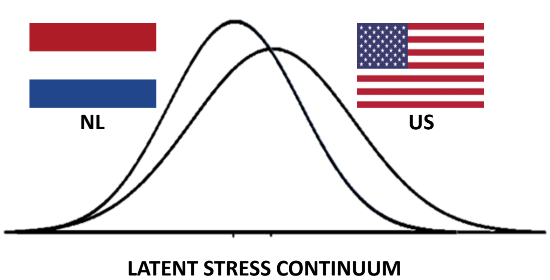
How Much do Boys and Girls Differ on Mean Latent Reading Proficiency?

- Girls are, on average, almost half a standard deviation higher than boys on the latent Reading Proficiency continuum.



39

Side Note: Graphic Representation of Latent Effect Size



40

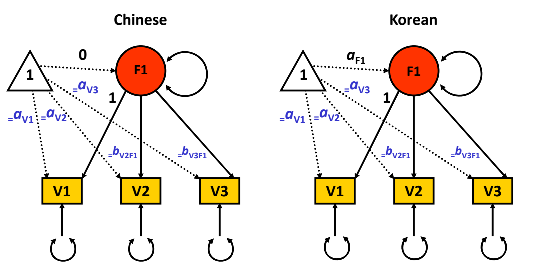
Structured Means Analysis: Hands-on Exercise

- Conduct structured means model to assess the latent mean difference between Chinese and Korean adults on English Reading.
- Data were collected on three measured indicators per factor from n = 86 Chinese adults and n = 96 Korean adults.
- Does there appear to be acceptable data-model fit across both groups?
- Which group appears to be higher on average on the latent construct? Is the group difference statistically significant?
- How many standard deviations would you estimate separate the population means along the latent continuum? That is, what is the estimated standardized effect size? [*This is a hand calculation*.]
- Start with the *partial* syntax file

  < Mplus English Latent Means Exercise >

41

The Model



STOP

42

Running Mplus: Data-model fit

```
MODEL FIT INFORMATION

Chi-Square Test of Model Fit

          Value                              2.675
          Degrees of Freedom                     4
          P-Value                           0.6137

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.000
          90 Percent C.I.                    0.000  0.132
          Probability RMSEA <= .05           0.708

CFI/TLI

          CFI                                1.000
          TLI                                1.005

SRMR (Standardized Root Mean Square Residual)

          Value                              0.036
```

43

Running Mplus: Unstandardized parameter estimates

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Group CHINESE				
READ BY				
READ1	1.000	0.000	999.000	999.000
READ2	1.001	0.061	16.516	0.000
READ3	0.900	0.059	15.252	0.000
Means				
READ	0.000	0.000	999.000	999.000
Intercepts				
READ1	16.208	0.720	22.521	0.000
READ2	17.335	0.724	23.932	0.000
READ3	18.180	0.676	26.898	0.000
Variances				
READ	37.287	6.810	5.475	0.000

44

Running Mplus: Unstandardized parameter estimates

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Group KOREAN				
READ BY				
READ1	1.000	0.000	999.000	999.000
READ2	1.001	0.061	16.516	0.000
READ3	0.900	0.059	15.252	0.000
Means				
READ	2.948	0.927	3.178	0.001
Intercepts				
READ1	16.208	0.720	22.521	0.000
READ2	17.335	0.724	23.932	0.000
READ3	18.180	0.676	26.898	0.000
Variances				
READ	32.034	5.605	5.716	0.000

45

Computing the Estimated Standardized Effect Size

- **Estimated standardized effect size:**

$$\hat{d} = |\hat{a}_{F1}| \Big/ \sqrt{\hat{c}_{F1}} \Longrightarrow \hat{c}_{F1} = \frac{n_C({}_C\hat{c}_{F1}) + n_K({}_K\hat{c}_{F1})}{n_C + n_K}$$

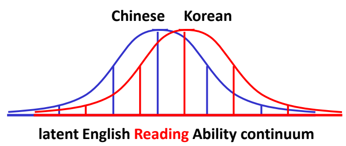
= pooled variance estimate of F1

$$\hat{d} = \frac{|2.95|}{\sqrt{86(37.29)+96(32.03) \Big/ (86+96)}} = .499$$

46

How Much do Chinese & Korean Adults Differ on Latent English Reading Ability?

- Korean native speakers are, on average, about .50 of a standard deviation higher than Chinese native speakers on the latent English Reading Ability continuum.



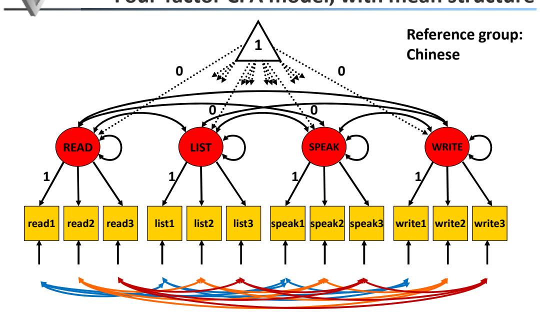
47

How Much do Chinese & Korean Adults Differ on Other Factors?

- Similar analyses with other latent outcomes yielded the following results:
  - *d*≈.63 on latent English Listening Ability.
  - *d*≈.26 on latent English Speaking Ability (NS).
  - *d*≈.26 on latent English Writing Ability (NS).
- Would it be possible to do all four outcomes (reading, listening, speaking, writing) simultaneously?

48

Can We Do Multiple Factors at Once?
Four-factor CFA model, with mean structure

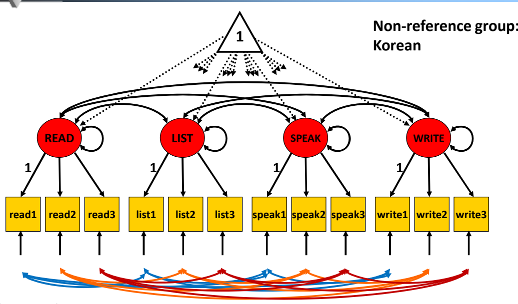


49

Can We Do Multiple Factors at Once?

Four-factor CFA model, with mean structure

Non-reference group: Korean



loadings and intercepts constrained across groups

50

Running Mplus: Model Syntax

```
DATA:
  FILE (chinese) IS chinese_means.txt;
  FILE (korean) IS korean_means.txt;

VARIABLE:
  NAMES ARE read1-read3 list1-list3
            speak1-speak3 write1-write3;

MODEL:
  read BY read1-read3; list BY list1-list3;
  speak BY speak1-speak3; write BY write1-write3;
  read1-read3 PWITH list1-list3;
  read1-read3 PWITH speak1-speak3;
  read1-read3 PWITH write1-write3;
  list1-list3 PWITH speak1-speak3;
  list1-list3 PWITH write1-write3;
  speak1-speak3 PWITH write1-write3;
```

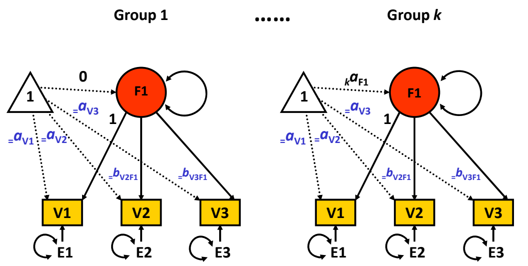
51

Running Mplus: Model Syntax (cont.)

```
MODEL chinese:
   [read@0]; [list@0]; [speak@0]; [write@0];

MODEL korean:
   [read]; [list]; [speak]; [write];

OUTPUT:
  sampstat modindices(3.841);
```

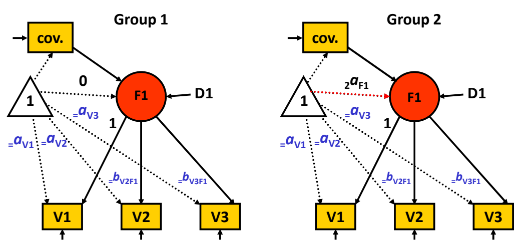
52

Mean Structure Modeling Extensions: More Than Two Groups



53

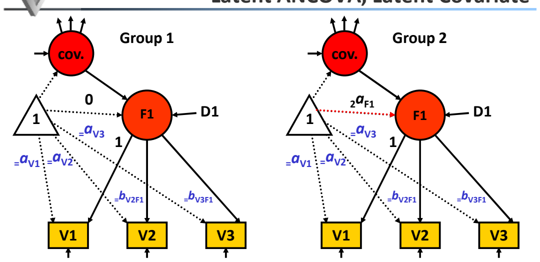
Mean Structure Modeling Extensions: Latent ANCOVA, Measured Covariate



The path from the pseudo-variable (1) to the factor is the test of latent mean differences above and beyond the measured covariate; the assumption of homogeneity of slope is testable in this model.

54

Mean Structure Modeling Extensions: Latent ANCOVA, Latent Covariate



The path from the pseudo-variable (1) to the outcome factor is the test of latent mean differences above and beyond the latent covariate; the assumption of homogeneity of slope is testable in this model.

55

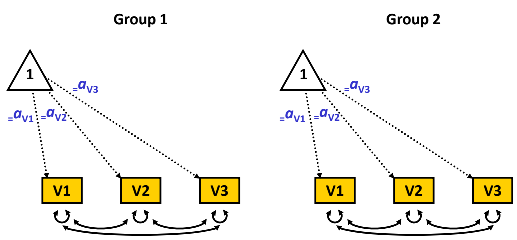
Mean Structure Modeling Extensions: What About MANOVA?

MANOVA:

- is often mistakenly used when researchers really should be using a latent means model;
- is a measured variable model, not a latent variable model, seeking population differences on any measured variable or linear combination of measured variables;
- makes strong assumptions about homogeneity of dispersion across populations;
- can be conducted better within the SEM framework (if you still insist on conducting something like a MANOVA).

56

Mean Structure Modeling Extensions: MANOVA Using Structured Means Analysis

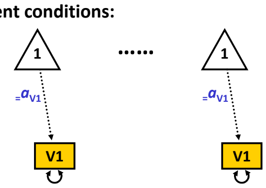


The χ^2 for the above model, with intercept constraints, corresponds to the omnibus MANOVA test, but without requiring the assumption of homogeneity of dispersion.

57

Mean Structure Modeling Extensions: ANOVA Using Structured Means Analysis

Same measured variable at different points in time, or under different conditions:

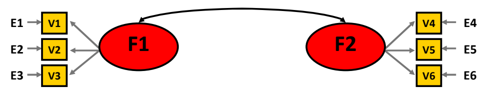


The χ^2 for this model, with intercept constraints, corresponds to the between-subjects omnibus test of means, but without requiring homogeneity of variance. It has also been shown to outperform ANOVA and adjusted forms of ANOVA (Fan & Hancock, 2012).

58

Repeated Measure Means Models:
Latent Variable Research Questions

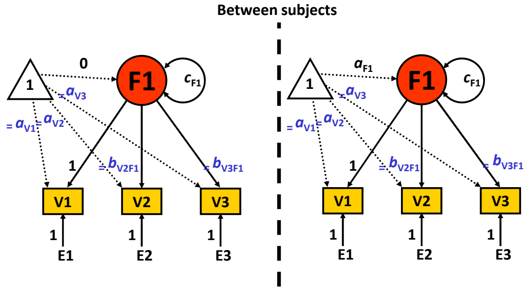
- Does a population differ with respect to the average amount of a particular latent construct across time (or, more generally, across conditions)?
- Do populations of matched individuals (e.g., mothers/daughters) differ with respect to the average amount of a particular latent construct?



59

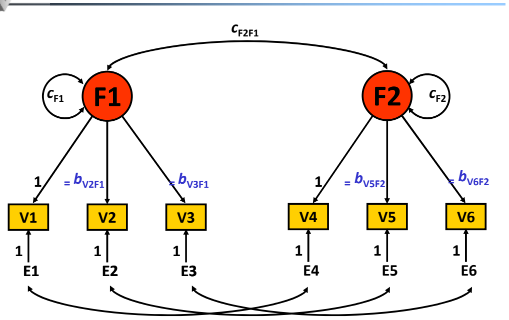
Between vs. Within Subjects Designs

Between subjects



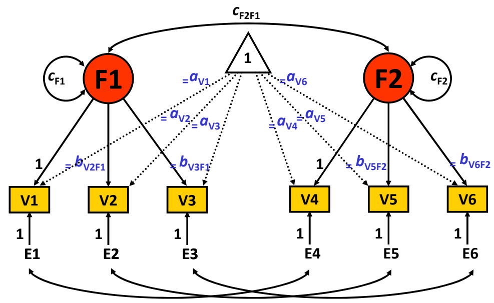
60

Repeated Measure Designs:
Covariance Structure



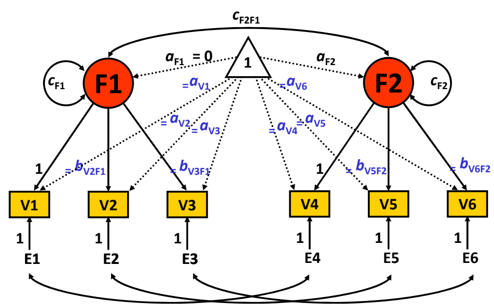
61

Repeated Measure Designs:
Mean Structure



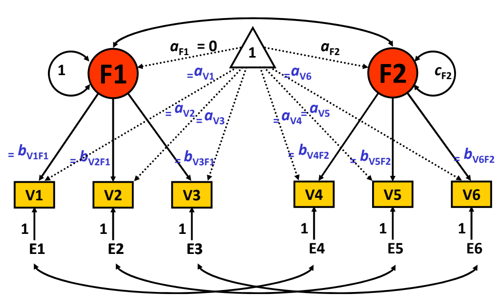
62

Repeated Measure Designs:
Mean Structure



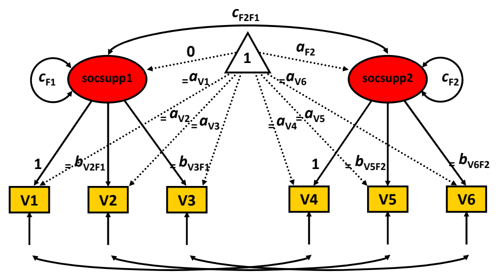
63

Repeated Measure Designs:
Mean Structure (alternative)



64

Duncan & Stoolmiller (1993) Example:
Model Set-up



65

Duncan & Stoolmiller (1993) Example:
Summary Statistics

n = 84

$$\mathbf{R} = \begin{bmatrix} 1 & & & & & \\ .812 & 1 & & & & \\ .819 & .752 & 1 & & & \\ .672 & .616 & .621 & 1 & & \\ .464 & .620 & .514 & .680 & 1 & \\ .612 & .640 & .719 & .819 & .676 & 1 \end{bmatrix}$$

$$\mathbf{s} = [2.46 \quad 1.76 \quad 2.74 \quad 2.63 \quad 1.89 \quad 2.84]$$

$$\mathbf{m} = [10.96 \quad 11.83 \quad 9.90 \quad 11.03 \quad 12.14 \quad 10.12]$$

66

Running Mplus:
Model Syntax

```
DATA:
  FILE IS repeated_means.txt;
  TYPE IS MEANS STD CORR;
  NOBS IS 84;

VARIABLE:
  NAMES ARE v1-v6;

MODEL:
  socsupp1 BY v1
              v2 (a)
              v3 (b);
  socsupp2 BY v4
              v5 (a)
              v6 (b);

  v1-v3 PWITH v4-v6;
```

$$\mathbf{m} = [10.96 \quad 11.83 \quad 9.90 \quad 11.03 \quad 12.14 \quad 10.12]$$

$$\mathbf{s} = [2.46 \quad 1.76 \quad 2.74 \quad 2.63 \quad 1.89 \quad 2.84]$$

$$\mathbf{R} = \begin{bmatrix} 1 & & & & & \\ .812 & 1 & & & & \\ .819 & .752 & 1 & & & \\ .672 & .616 & .621 & 1 & & \\ .464 & .620 & .514 & .680 & 1 & \\ .612 & .640 & .719 & .819 & .676 & 1 \end{bmatrix}$$

67

Running Mplus:
Model Syntax (cont.)

```
  [v1] (c); [v4] (c);
  [v2] (d); [v5] (d);
  [v3] (e); [v6] (e);

  [socsupp1@0];
  [socsupp2];

OUTPUT:
  sampstat modindices(3.841);
```

68

Running Mplus:
Data-model fit

```
MODEL FIT INFORMATION

Chi-Square Test of Model Fit

          Value                            11.967
          Degrees of Freedom                    9
          P-Value                          0.2152

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                          0.063
          90 Percent C.I.                   0.000  0.146
          Probability RMSEA <= .05          0.360

CFI/TLI

          CFI                               0.993
          TLI                               0.988

SRMR (Standardized Root Mean Square Residual)

          Value                             0.044
```

69

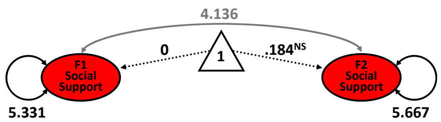
Running Mplus:
Unstandardized parameter estimates

```
MODEL RESULTS
                                                  Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
 SOCSUPP1 BY
    V1                 1.000      0.000    999.000    999.000
    V2                 0.638      0.049     13.116      0.000
    V3                 1.039      0.073     14.235      0.000

 SOCSUPP2 BY
    V4                 1.000      0.000    999.000    999.000
    V5                 0.638      0.049     13.116      0.000
    V6                 1.039      0.073     14.235      0.000

 SOCSUPP2 WITH
    SOCSUPP1           4.136      0.849      4.874      0.000

 Means
    SOCSUPP1           0.000      0.000    999.000    999.000
    SOCSUPP2           0.184      0.199      0.925      0.355

Variances
    SOCSUPP1           5.331      0.947      5.631      0.000
    SOCSUPP2           5.667      1.036      5.469      0.000
```

70

Exercise Behavior Means Example:
Key Results

Model χ^2=11.97, *df*=9, *p*=.22
RMSEA=.063; SRMR=0.044



$$\hat{d} = \frac{|.184|}{\sqrt{84(5.331) + 84(5.667) / (84+84)}} = .078$$

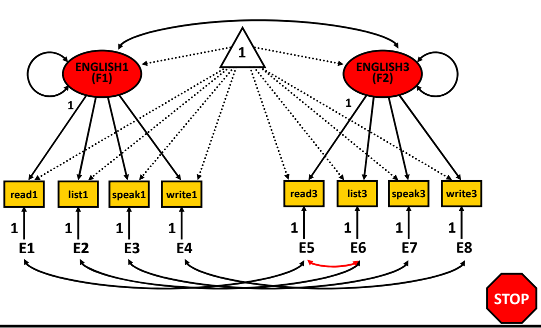
71

Repeated Measure Latent Means:
Hands-on Exercise

- Conduct a repeated latent means model to assess the latent mean difference from time 1 to time 3 for a sample of Chinese adults on English Ability.
- Data were collected on four measured indicators per factor (reading, listening, speaking, writing) from n = 86 Chinese adults.
- Researchers have a theoretical reason (due to test administration) that the reading and listening measures' errors should covary at time 3.
- Does there appear to be acceptable data-model fit?
- Is the latent mean difference from time 1 to time 3 statistically significant?
- How many standard deviations would you estimate separate time 1 and time 3 along the latent continuum? That is, what is the estimated standardized effect size? [*This is a hand calculation.*]
- Start with the *partial* syntax file
  < *Mplus English Repeated Latent Means Exercise* >

72

The Model:
Constraints not shown



73

Running Mplus:
Data-model fit

```
MODEL FIT INFORMATION

Chi-Square Test of Model Fit

          Value                              32.814
          Degrees of Freedom                     20
          P-Value                            0.0354

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.086
          90 Percent C.I.                     0.023  0.138
          Probability RMSEA <= .05            0.132

CFI/TLI

          CFI                                 0.975
          TLI                                 0.965

SRMR (Standardized Root Mean Square Residual)

          Value                               0.061
```

74

Running Mplus:
Unstandardized parameter estimates

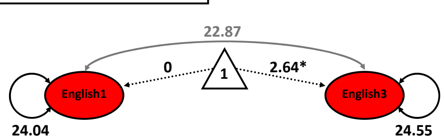
MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ENGLISH1 BY				
READ1	1.000	0.000	999.000	999.000
LIST1	1.028	0.102	10.087	0.000
SPEAK1	0.430	0.052	8.225	0.000
WRITE1	0.774	0.074	10.518	0.000
ENGLISH3 BY				
READ3	1.000	0.000	999.000	999.000
LIST3	1.028	0.102	10.087	0.000
SPEAK3	0.430	0.052	8.225	0.000
WRITE3	0.774	0.074	10.518	0.000
Means				
ENGLISH1	0.000	0.000	999.000	999.000
ENGLISH3	2.639	0.370	7.124	0.000
Variances				
ENGLISH1	24.042	5.491	4.378	0.000
ENGLISH3	24.550	5.783	4.245	0.000

75

Repeated Measure Exercise:
Key Results

Model χ^2=32.80, *df*=20, *p*=.036
RMSEA=.086; SRMR=0.063



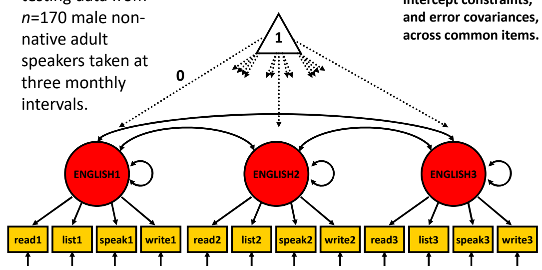
$$\hat{d} = \frac{|2.64|}{\sqrt{\left[86(24.04) + 86(24.55)\right] / (86+86)}} = .532$$

76

Mean Structure Modeling Extensions: More than two time points

- English language testing data from *n*=170 male non-native adult speakers taken at three monthly intervals.

Plus loading and intercept constraints, and error covariances, across common items.



77

Mean Structure Modeling Extensions: More than two time points

```
DATA:
  FILE IS latent_means2_data.txt;
VARIABLE:
  NAMES ARE read1-read3 list1-list3
            speak1-speak3 write1-write3;
MODEL:
  ENGLISH1 BY read1 list1 (a)
              speak1 (b)
              write1 (c);
  ENGLISH2 BY read2 list2 (a)
              speak2 (b)
              write2 (c);
  ENGLISH3 BY read3 list3 (a)
              speak3 (b)
              write3 (c);
```

78

Mean Structure Modeling Extensions: More than two time points

```
  read1 WITH read2-read3;
  read2 WITH read3;
  list1 WITH list2-list3;
  list2 WITH list3;
  speak1 WITH speak2-speak3;
  speak2 WITH speak3;
  write1 WITH write2-write3;
  write2 WITH write3;
  [read1-read3] (d);
  [list1-list3] (e);
  [speak1-speak3] (f);
  [write1-write3] (g);
```

79

Mean Structure Modeling Extensions: More than two time points

```
  [ENGLISH1@0];
  [ENGLISH2] (eng2);
  [ENGLISH3] (eng3);

MODEL CONSTRAINT:
  NEW(engdiff);
  engdiff = eng2 - eng3;

OUTPUT:
  SAMPSTAT MODINDICES(3.841);
```

80

Mean Structure Modeling Extensions: More than two time points

```
Chi-Square Test of Model Fit

        Value                          108.271
        Degrees of Freedom                  51
        P-Value                         0.0000

RMSEA (Root Mean Square Error Of Approximation)

        Estimate                         0.081
        90 Percent C.I.                  0.060  0.103
        Probability RMSEA <= .05         0.010

SRMR (Standardized Root Mean Square Residual)

        Value                            0.071
```

81

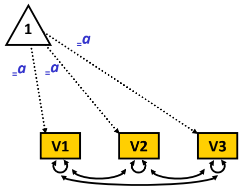
Mean Structure Modeling Extensions: More than two time points

```
                                                  Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value

Means
   ENGLISH1           0.000      0.000    999.000    999.000
   ENGLISH2           1.122      0.195      5.766      0.000
   ENGLISH3           1.906      0.232      8.222      0.000

Variances
   ENGLISH1          19.046      3.461      5.503      0.000
   ENGLISH2          16.175      3.002      5.389      0.000
   ENGLISH3          15.964      2.879      5.546      0.000



New/Additional Parameters
 ENGDIFF             -0.784      0.158     -4.959      0.000
```

82

Mean Structure Modeling Extensions:
Repeated ANOVA Using Means Analysis

Same measured variables at different points in time, or under different conditions:



The χ^2 for this model, with intercept constraints, corresponds to the repeated measure omnibus test, but without requiring the assumption of sphericity.

83

Summary

- Latent means models allow for inferences about population mean differences at the latent level.
- By parsing out measurement error, latent means models have more power than the measured variable techniques that they subsume (e.g., MANOVA, ANOVA).
- Multiple groups, multiple time points, as well as measured and latent covariates, may be accommodated within this analytic framework.
- Methods of modeling, effect size calculations, power and sample size estimations – exist for simple as well as more complex models.

84

Supplemental Readings

- Hancock, G. R. (2004). Experimental, quasi-experimental, and nonexperimental design and analysis with latent variables. In D. Kaplan (Ed.), *SAGE handbook of quantitative methodology for the social sciences* (pp. 317-334). Thousand Oaks, CA: Sage.
- Thompson, M. S., & Green, S. B. (2013). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed.) (pp. 163-218). Charlotte, NC: Information Age Publishing.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology*, *27*, 229-239.

85