Structural Equation Modeling Using the CALIS Procedure in SAS/STAT® Software: Basic and Advanced Topics

Yiu-Fai Yung
SAS Institute Inc.
Cary, NC 27513 USA

Statistical tutorial presented at the Michigan SAS Users Group Meeting, May 20, 2014, Livonia, Michigan. Email: Yiu-Fai.Yung@sas.com,

SAS/STAT 9.22 or later

is assumed

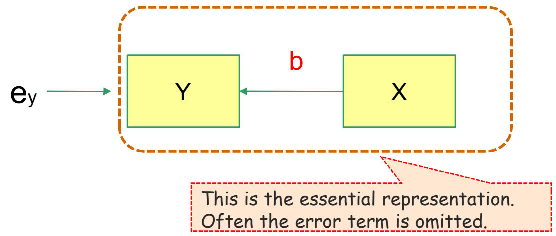
for this workshop

2

Causal Model, Prediction, and Path Diagram

- X causes Y
- X predicts Y
- Linear regression equation

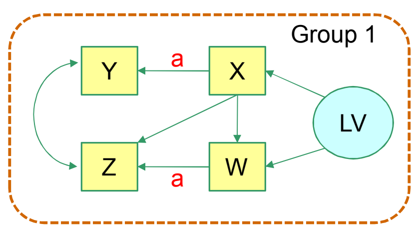
  $Y = b X + e_y$
- Path diagram

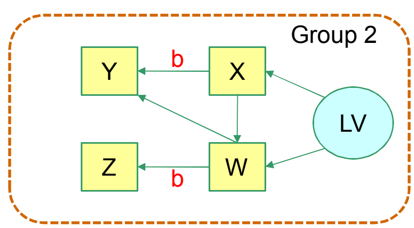


3

Structural Equation Modeling versus Regression Analysis

- More variables
- More equations
- Correlated errors
- Direct and indirect effects
- Latent variables
- Parametric constraints
- Multiple-group analysis





4

Other Names or Closely-Related Analyses for Structural Equation Modeling (SEM)

- Path analysis (usually for observed variables only)
- LISREL model (Jöreskog 1973, Keesling 1972, Wiley 1973)
- Covariance structures analysis
- Analysis of moment structures
- Confirmatory factor analysis
- Causal modeling
- CALIS: Covariance Analysis of Linear Structural Equations

5

A Very Brief History of PROC CALIS

- Older versions: before SAS 9.2
- TCALIS (SAS 9.2, 2008): experimental version
- “New” CALIS (SAS 9.22, 2010): PATH modeling language, multiple-group analysis, mean structures, name-free approach to parameter specifications, and much more
- Current version (SAS/STAT 13.1, 2013): Full information maximum likelihood, robust estimation, case-level residual diagnostics, and path diagram

6

Structure of the Workshop

- First Part: Basic Modeling
  1. A brief description of the process of SEM
  2. The PATH modeling language in PROC CALIS
  3. Specifying models and interpreting results
  4. Extended PATH modeling language
  5. LISMOD – a language tailored to LISREL users
- Second Part: “Advanced” Modeling
  1. Multiple-group analysis
  2. Analyzing direct and indirect effects
  3. Creating Path Diagrams
  4. Testing specific hypotheses
  5. Model modifications
  6. Full information maximum likelihood estimation
  7. Case-level (Observation-level) residual diagnostics

7

Emphases of the Workshop

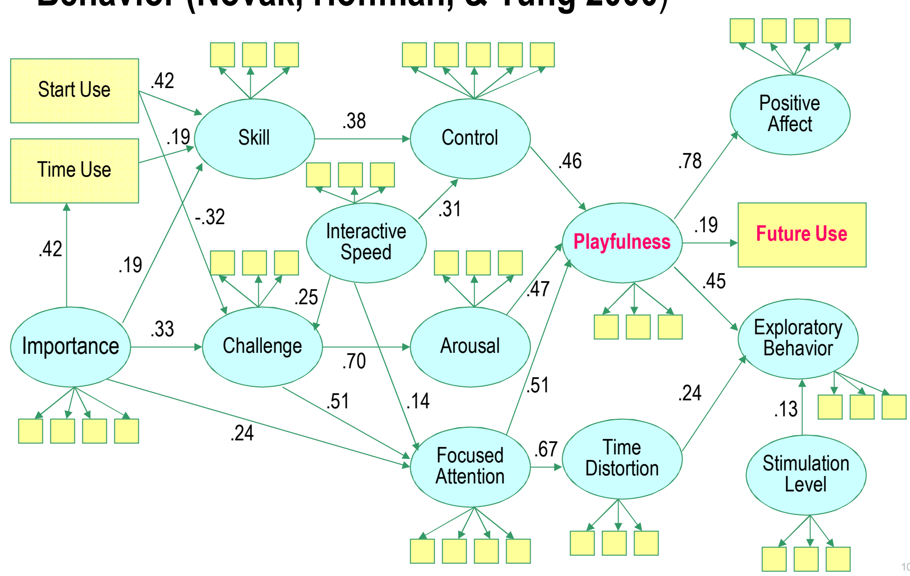
- Introducing the structural equation methodology and applications through examples – What is SEM?
- Analyzing structural equation models with PROC CALIS – How to do SEM?

8

Illustrating the Process of Structural Equation Modeling

9

A Structural Equation Model of Web-Surfing Behavior (Novak, Hoffman, & Yung 2000)



10

Examples of Items

- Playfulness
  - The event was very playful.
  - The event was fun.
- Future Use
  - I would like to engage the same activity in the future.
- Time Distortion Items
  - The experience overwhelmed other senses and thoughts.
  - I forgot about my immediate surroundings when browsing the web-page.
- Control
  - I felt in control.
  - The web-page design allowed me to control the interaction.

11

Key Features of SEM

- Analyzing complicated relationships among variables
- Path diagram representations for models
- Ability to handle latent and observed variables simultaneously
- Testing the model fit and significance of the parameters
- Suggesting ways to improve the model

12

Basics: A Simple Regression Model and the PATH Modeling Language

13

A Simple Linear Regression Model

- $y = b\,x + e_y$
- y: outcome variable
- x: predictor variable
- e_y: error term
- b: effect or regression coefficient

Assumption: Variables are centered.

14

Measures of the Number of Hen Pheasants

- Fuller (1987) p.34
- y : average of the number of birds in August
- x : average of the number of birds in Spring (April/May)
- Averages were based on the number of birds sighted by 15 trained observers
- Goal: How many birds will survive 3 months?

15

Regression Analysis by PROC REG

```
data hens;
   input y x @@;
   datalines;
8     9     6     6.6   9.8   12.3  10.8  11.9  9.7   11.9  9.3   12
9.2   9.6   6.9   7.5   8.1   10.9   8.7  10.4  8.7   10.2  7.4   7.4
10.1  11    10    11.8  7.3    8.2
;
proc reg data=hens;
   model y = x;
run;
```


16

Results Obtained from PROC REG

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value
Intercept	1	2.14227	0.84513	2.53
x	1	0.64941 (b)	0.08275	7.85

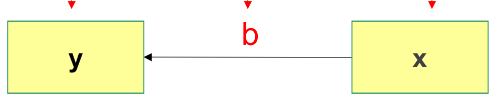
Given a base survival of 2.14 birds, every additional bird in Spring predicts a 0.65 bird surviving in August.

17

Regression Equation, Path Diagram, and the PATH Modeling Language

$$y \quad = \quad b \quad x \quad + \quad e_y$$

Outcome (y) Parameter (b) Predictor (x)




```
PATH    y <=== x = b;
```

Path Statement — Path Relation — Parameter (optional)

18

Regression Model Specified by PROC CALIS

```
proc calis data=hens;
   path
      y <=== x;
run;
```


Results from PROC CALIS for the Pheasant Data

PATH List

--------Path--------			Parameter	Estimate	Standard Error	t Value	Pr > |t|
y	<===	x	_Parm1	0.64941	0.07974	8.1444	<.0001

The same estimate of b by PROC REG

Default variance parameters

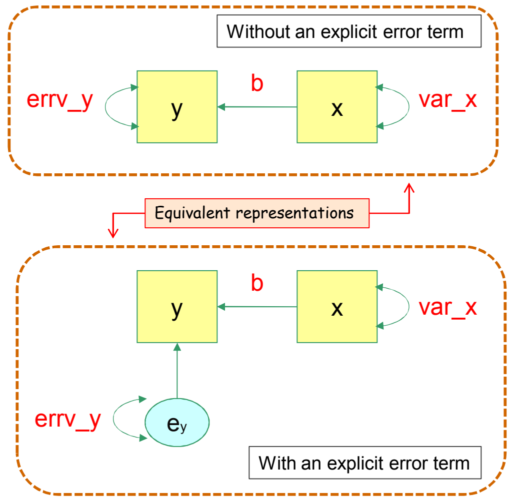
Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Exogenous	x	_Add1	3.62124	1.36870	2.6458	0.0082
Error	y	_Add2	0.32233	0.12183	2.6458	0.0082

Optional Specification with Parameter Names

```
proc calis data=hens;
   path
      y <=== x    = b;
   pvar
      x  = var_x,
      y  = errv_y;
   run;
```

Use the PVAR statement to specify variance or error variance parameters. You can also define parameters explicitly in PROC CALIS.



21

Hen Pheasants Results with Parameter Names Specified

PATH List

--------Path--------			Parameter	Estimate	Standard Error	t Value	Pr > |t|
y	<===	x	b	0.64941	0.07974	8.1444	<.0001

Parameter name specified

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Exogenous	x	var_x	3.62124	1.36870	2.6458	0.0082
Error	y	errv_y	0.32233	0.12183	2.6458	0.0082

Parameter names specified

22

Keys to the PATH Modeling Language

- As easy as drawing a path diagram
- PATH statement specifies the functional relationships – required specification
- PROC CALIS sets variances and error variances by default – optional specification
- Naming free parameters is optional

> Most of the time, you only need to specify the functional relationships by using the PATH statement.

23

Measurement Errors in Predictors

- Bird counting might involve measurement errors in x
- $x = f_x + e_x$
- f_x : true score, but not observed
- x : observed, but with measurement error e_x

24

A Measurement Error Model for the Pheasant Data

- Structural Equation

  $y = b\ f_x + e_y$

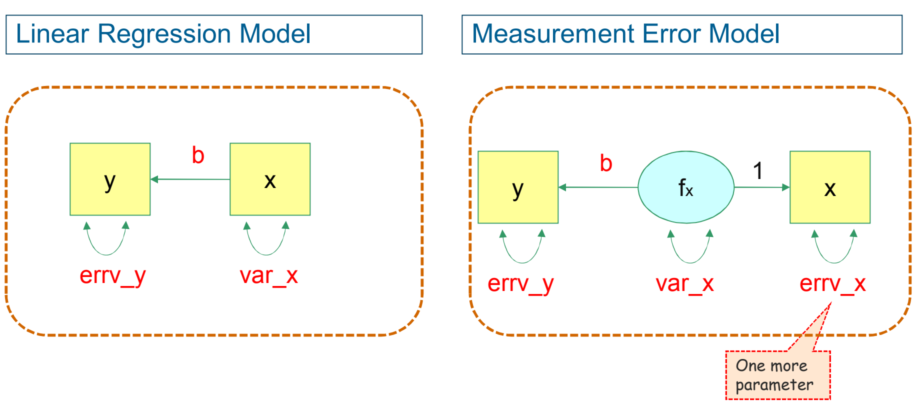
- Measurement equation

  $x = f_x + e_x$

- Can you estimate b?
- Problem: The measurement equation introduces an additional parameter: Var(e_x) (variance of e_x or error variance of x)

25

Path Diagram Representations



26

Constraining the Error Variances

- Bird counting is more accurate in fall (y) than in spring (x)
- In an independent study, error variance (for x) in spring is six times as much as that (for y) in fall
- Fuller's suggestion: Var(ex) = 6 Var(ey)

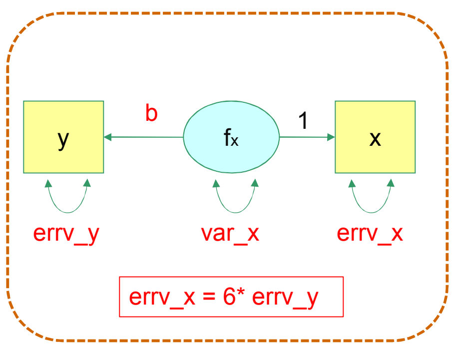
errv_x = 6* errv_y

27

A Measurement Error Model with a Constraint for the Pheasant Data

```
proc calis data=hens;
   path
      y  <=== fx    = b,
      fx ===> x     = 1;
   pvar
      y   = errv_y,
      fx  = var_x,
      x   = errv_x;
   errv_x = 6 * errv_y;
run;
```

The required constraint is specified as a SAS programming statement.



28

Measurement Error Model Results for the Pheasant Data

PATH List

--------Path--------			Parameter	Estimate	Standard Error	t Value	Pr > |t|
y	<===	fx	b	0.75158	0.09228	8.1443	<.0001
fx	===>	x		1.00000			

A larger estimated effect than the one estimated without taking the measurement error into account (0.649)

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Error	y	errv_y	0.08205	0.03101	2.6458	0.0082
Exogenous	fx	var_x	3.12893	1.36180	2.2976	0.0216
Error	x	errv_x	0.49231	0.18608	2.6458	0.0082

Six times as much as the estimate for errv_y

29

Some Features of the PATH Modeling Language

- PATH statement for specifying paths or relationships
- PVAR statement for specifying variances and error variances
- PCOV statement for specifying covariances and error covariances (to be shown)
- Parameter dependency can be specified by SAS programming statements. For example,

```
parm1 = 4 * parm2 + exp(parm4) ** parm6;
```

30

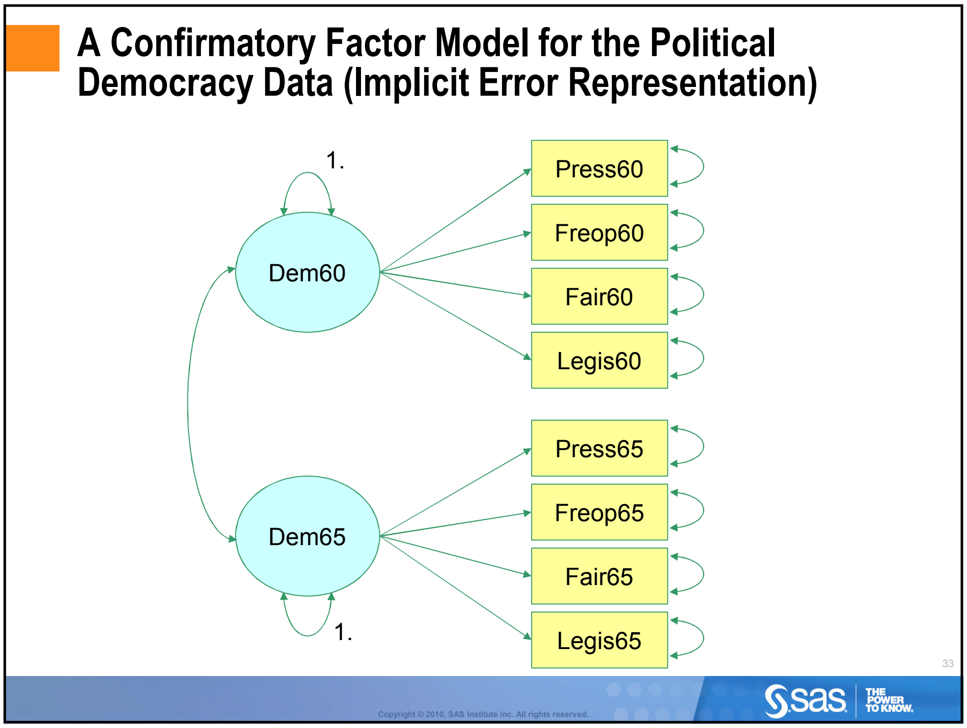
A Confirmatory Factor Model

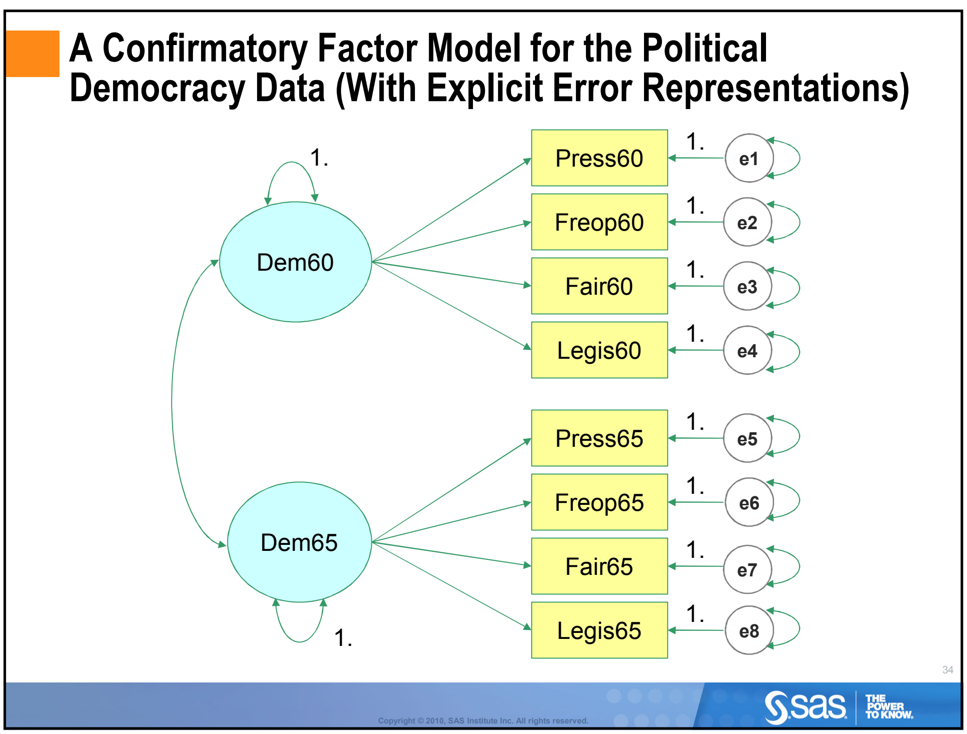
31

Political Democracy Data

- Bollen (1989) Chapter 7
- Two latent factors: political democracy in 75 developing countries in 1960 and 1965
- Four indicator measures for the latent factors in each year:
  - Freedom of press (Press60, Press65)
  - Freedom of group oppositions (Freop60, Freop65)
  - Fairness of elections (Fair60, Fair65)
  - Elective nature of the legislative body (Legis60, Legis65)
- Purpose of the confirmatory factor analysis: Validate the measurement indicators

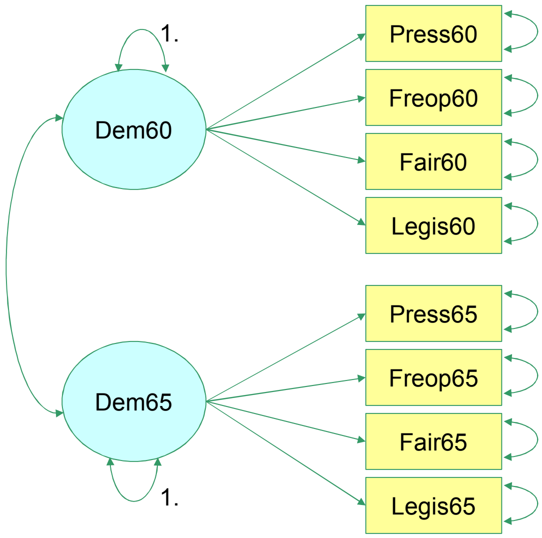
32


A Confirmatory Factor Model for the Political Democracy Data (Implicit Error Representation)
1.
Dem60
Press60
Freop60
Fair60
Legis60
Dem65
Press65
Freop65
Fair65
Legis65
1.
33
Copyright © 2010, SAS Institute Inc. All rights reserved.
sas
THE POWER TO KNOW.


A Confirmatory Factor Model for the Political Democracy Data (With Explicit Error Representations)
1.
Dem60
Press60
Freop60
Fair60
Legis60
Dem65
Press65
Freop65
Fair65
Legis65
1.
e1
e2
e3
e4
e5
e6
e7
e8
34
Copyright © 2010, SAS Institute Inc. All rights reserved.
sas
THE POWER TO KNOW.

Basic Confirmatory Factor Model for the Political Democracy Data

```
proc calis data=polidem;
path
  Dem60 ===> Press60,
  Dem60 ===> Freop60,
  Dem60 ===> Fair60,
  Dem60 ===> Legis60,
  Dem65 ===> Press65,
  Dem65 ===> Freop65,
  Dem65 ===> Fair65,
  Dem65 ===> Legis65;
pvar
  Dem60 = 1, Dem65 = 1,
  Press60 Freop60 Fair60
  Legis60 Press65 Freop60
  Fair65 Legis65;
pcov
  Dem60 Dem65;
run;
```



35

Error Variances, Variances, and Exogenous Covariances Are Free Parameters by Default

Specifying All Variance and Covariance Parameters

```
proc calis data=polidem;
   path
      Dem60 ===> Press60,
      Dem60 ===> Freop60,
      Dem60 ===> Fair60,
      Dem60 ===> Legis60,
      Dem65 ===> Press65,
      Dem65 ===> Freop65,
      Dem65 ===> Fair65,
      Dem65 ===> Legis65;
   pvar
      Dem60 = 1, Dem65 = 1,
      Press60 Freop60 Fair60
      Legis60 Press65 Freop65
      Fair65 Legis65;
   pcov
      Dem60 Dem65;
run;
```

These could have been set automatically by default.

Default Variance and Covariance Parameters in Effect

```
proc calis data=polidem;
   path
      Dem60 ===> Press60,
      Dem60 ===> Freop60,
      Dem60 ===> Fair60,
      Dem60 ===> Legis60,
      Dem65 ===> Press65,
      Dem65 ===> Freop65,
      Dem65 ===> Fair65,
      Dem65 ===> Legis65;
   pvar
      Dem60 = 1, Dem65 = 1;
run;
```

36

Estimates of Path Coefficients (Loadings) for the Political Democracy Data

PATH List

----------Path----------			Parameter	Estimate	Standard Error	t Value	Pr > |t|
Dem60	===>	Press60	_Parm01	2.20567	0.25122	8.7800	<.0001
Dem60	===>	Freop60	_Parm02	3.00132	0.39735	7.5533	<.0001
Dem60	===>	Fair60	_Parm03	2.31033	0.34026	6.7899	<.0001
Dem60	===>	Legis60	_Parm04	2.89483	0.31582	9.1662	<.0001
Dem65	===>	Press65	_Parm05	2.04790	0.25930	7.8977	<.0001
Dem65	===>	Freop65	_Parm06	2.68003	0.33258	8.0583	<.0001
Dem65	===>	Fair65	_Parm07	2.70879	0.31804	8.5171	<.0001
Dem65	===>	Legis65	_Parm08	2.76604	0.30830	8.9719	<.0001

All path estimates are significant.

37

Estimates of Variances for the Political Democracy Data

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Exogenous	Dem60		1.00000			
	Dem65		1.00000			
Error	Press60	_Parm09	2.01359	0.41048	4.9055	<.0001
	Freop60	_Parm10	6.57189	1.20964	5.4329	<.0001
	Fair60	_Parm11	5.42661	0.96546	5.6208	<.0001
	Legis60	_Parm12	2.83887	0.61417	4.6223	<.0001
	Press65	_Parm13	2.63180	0.49311	5.3371	<.0001
	Freop65	_Parm14	4.19276	0.79422	5.2791	<.0001
	Fair65	_Parm15	3.46180	0.68155	5.0793	<.0001
	Legis65	_Parm16	2.88292	0.59927	4.8107	<.0001

All error variance estimates are significant.

38

Estimate of Covariance for the Political Democracy Data

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Dem60	Dem65	_Parm17	0.97528	0.02656	36.7232	<.0001

High and significant correlation between the Democracy factors in 1960 and 1965.

39

How Is the Model Fit?

Fit Summary

Modeling Info	N Observations	75
	N Variables	8
	N Moments	36
	N Parameters	17
	N Active Constraints	0
	Baseline Model Function Value	6.1482
	Baseline Model Chi-Square	454.9633
	Baseline Model Chi-Square DF	28
	Pr > Baseline Model Chi-Square	<.0001
Absolute Index	Fit Function	0.6009
	Chi-Square	44.4686
	Chi-Square DF	19
	Pr > Chi-Square	0.0008
	Z-Test of Wilson & Hilferty	3.1383
	Hoelter Critical N	51
	Root Mean Square Residual (RMR)	0.5388
	Standardized RMR (SRMR)	0.0494
	Goodness of Fit Index (GFI)	0.8658
Parsimony Index	Adjusted GFI (AGFI)	0.7457
	Parsimonious GFI	0.5875
	RMSEA Estimate	0.1346
	RMSEA Lower 90% Confidence Limit	0.0833
	RMSEA Upper 90% Confidence Limit	0.1865
	Probability of Close Fit	0.0062
	ECVI Estimate	1.1240
	ECVI Lower 90% Confidence Limit	0.9065
	ECVI Upper 90% Confidence Limit	1.4608
	Akaike Information Criterion	78.4686
	Bozdogan CAIC	134.8659
	Schwarz Bayesian Criterion	117.8659
	McDonald Centrality	0.8438
Incremental Index	Bentler Comparative Fit Index	0.9403
	Bentler-Bonett NFI	0.9023
	Bentler-Bonett Non-normed Index	0.9121
	Bollen Normed Index Rho1	0.8560
	Bollen Non-normed Index Delta2	0.9416
	James et al. Parsimonious NFI	0.6122

A lot of fit indices, but researchers usually report just a few of them.

40

Using the FITINDEX Statement to Customize the Fit Summary Output

```
proc calis data=polidem;
   path
      Dem60 ===> Press60,
      Dem60 ===> Freop60,
      Dem60 ===> Fair60,
      Dem60 ===> Legis60,
      Dem65 ===> Press65,
      Dem65 ===> Freop65,
      Dem65 ===> Fair65,
      Dem65 ===> Legis65;
   pvar
      Dem60 = 1, Dem65 = 1;
   fitindex on(only) = [chisq df probchi rmsea cn srmsr
           bentlercfi agfi] noindextype;
run;
```

ON(ONLY)= selects the set of fit indices to display.
NOINDEXTYPE suppresses the printing of index types.

41

Customized Fit Summary Table

Fit Summary

Chi-Square	44.4686
Chi-Square DF	19
Pr > Chi-Square	0.0008
Hoelter Critical N	51
Standardized RMR (SRMR)	0.0494
Adjusted GFI (AGFI)	0.7457
RMSEA Estimate	0.1346
Bentler Comparative Fit Index	0.9403

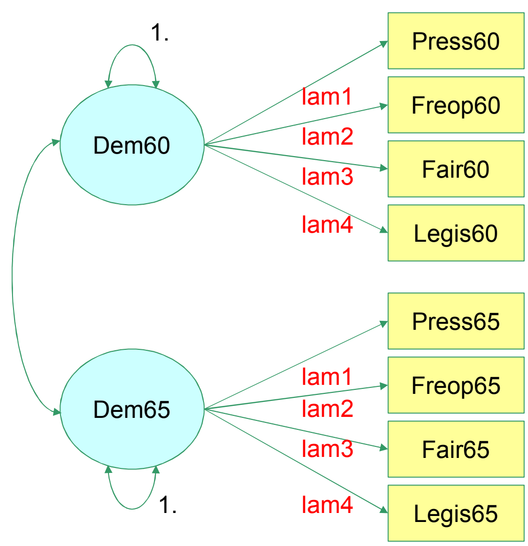
"Good" SRMR and Bentler's CFI. "Bad" chi-square, AGFI, RMSEA.

42

A Confirmatory Factor Model with Loading Constraints

43

Constraining the Path Coefficients (Loadings)



Note: Error variances are not represented because they are default parameters.

44

Fitting a CFA Model with Constraints on the Loadings

```
proc calis data=polidem;
   path
      Dem60 ===> Press60    = lam1,
      Dem60 ===> Freop60    = lam2,
      Dem60 ===> Fair60     = lam3,
      Dem60 ===> Legis60    = lam4,
      Dem65 ===> Press65    = lam1,
      Dem65 ===> Freop65    = lam2,
      Dem65 ===> Fair65     = lam3,
      Dem65 ===> Legis65    = lam4;
   pvar
      Dem60 = 1, Dem65 = 1;
run;
```

These constrain the path coefficients.

Fit Summary Table for the Political Democracy Data with Loading Constraints

Fit Summary

Chi-Square	46.8893
Chi-Square DF	23
Pr > Chi-Square	0.0023
Hoelter Critical N	56
Standardized RMR (SRMR)	0.0714
Adjusted GFI (AGFI)	0.7844
RMSEA Estimate	0.1185
Bentler Comparative Fit Index	0.9440

Only Bentler's CFI indicates a good model fit.

Estimates of the Constrained Loadings for the Political Democracy Data

PATH List

----------Path----------			Parameter	Estimate	Standard Error	t Value	Pr > |t|
Dem60	===>	Press60	lam1	2.13970	0.21716	9.8532	<.0001
Dem60	===>	Freop60	lam2	2.80116	0.29976	9.3447	<.0001
Dem60	===>	Fair60	lam3	2.54987	0.27316	9.3346	<.0001
Dem60	===>	Legis60	lam4	2.82969	0.27285	10.3708	<.0001
Dem65	===>	Press65	lam1	2.13970	0.21716	9.8532	<.0001
Dem65	===>	Freop65	lam2	2.80116	0.29976	9.3447	<.0001
Dem65	===>	Fair65	lam3	2.54987	0.27316	9.3346	<.0001
Dem65	===>	Legis65	lam4	2.82969	0.27285	10.3708	<.0001

All path coefficients are significant.

47

Estimates of Variances and Covariances for the Political Democracy Data with Loading Constraints

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Exogenous	Dem60		1.00000			
	Dem65		1.00000			
Error	Press60	_Add1	2.01017	0.40312	4.9865	<.0001
	Freop60	_Add2	6.72037	1.21196	5.5450	<.0001
	Fair60	_Add3	5.40824	0.97833	5.5280	<.0001
	Legis60	_Add4	2.88468	0.60956	4.7324	<.0001
	Press65	_Add5	2.61966	0.49456	5.2970	<.0001
	Freop65	_Add6	4.16958	0.79818	5.2238	<.0001
	Fair65	_Add7	3.55382	0.67700	5.2494	<.0001
	Legis65	_Add8	2.85029	0.59556	4.7859	<.0001

Covariances Among Exogenous Variables

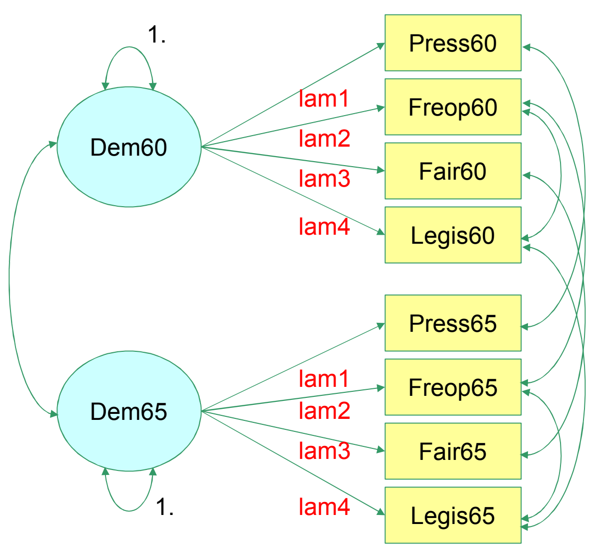
Var1	Var2	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Dem65	Dem60	_Add9	0.97480	0.02682	36.3466	<.0001

48

A Confirmatory Factor Model with Correlated Errors

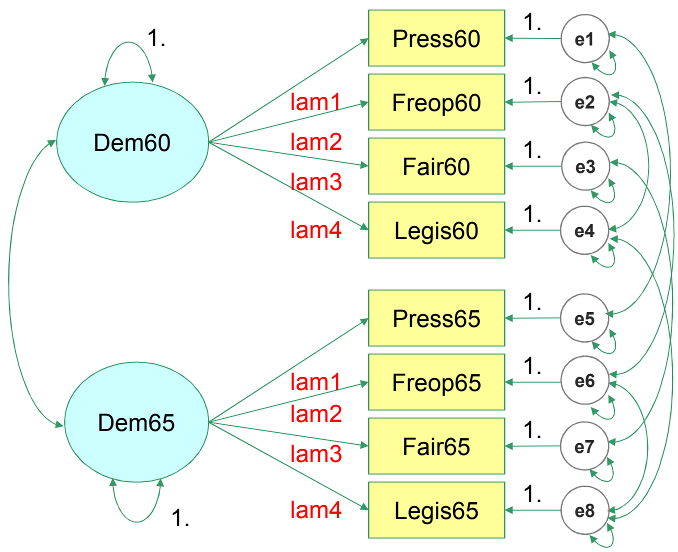
49

Adding Error Covariances



50

Adding Error Covariances (With Explicit Error Terms)



51

Fitting a CFA Model with Loading Constraints and Correlated Errors

```
proc calis data=polidem;
   path
      Dem60 ===> Press60   = lam1,
      Dem60 ===> Freop60   = lam2,
      Dem60 ===> Fair60    = lam3,
      Dem60 ===> Legis60   = lam4,
      Dem65 ===> Press65   = lam1,
      Dem65 ===> Freop65   = lam2,
      Dem65 ===> Fair65    = lam3,
      Dem65 ===> Legis65   = lam4;
   pvar
      Dem60 = 1, Dem65 = 1;
   pcov
      Freop60 Legis60, Freop65 Legis65,
      Press60 Press65, Freop60 Freop65,
      Fair60 Fair65, Legis60 Legis65;
run;
```

Use the PCOV statement to specify error covariances.

52

Fit Summary Table for the CFA Model with Loading Constraints and Correlated Errors

Fit Summary

Chi-Square	15.1946
Chi-Square DF	17
Pr > Chi-Square	0.5815
Hoelter Critical N	135
Standardized RMR (SRMR)	0.0590
Adjusted GFI (AGFI)	0.9043
RMSEA Estimate	0.0000
Bentler Comparative Fit Index	1.0000

All indices indicate a good model fit.

Estimates of the Loadings for the CFA Model with Constrained Loadings and Correlated Errors

PATH List

----------Path----------			Parameter	Estimate	Standard Error	t Value	Pr > |t|
Dem60	===>	Press60	lam1	2.16450	0.23009	9.4074	<.0001
Dem60	===>	Freop60	lam2	2.61630	0.32500	8.0502	<.0001
Dem60	===>	Fair60	lam3	2.61693	0.28700	9.1183	<.0001
Dem60	===>	Legis60	lam4	2.75291	0.28312	9.7236	<.0001
Dem65	===>	Press65	lam1	2.16450	0.23009	9.4074	<.0001
Dem65	===>	Freop65	lam2	2.61630	0.32500	8.0502	<.0001
Dem65	===>	Fair65	lam3	2.61693	0.28700	9.1183	<.0001
Dem65	===>	Legis65	lam4	2.75291	0.28312	9.7236	<.0001

All path coefficients are significant.

Estimates of the Variances for the CFA Model with Constrained Loadings and Correlated Errors

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Exogenous	Dem60		1.00000			
	Dem65		1.00000			
Error	Press60	_Add1	1.91664	0.43982	4.3578	<.0001
	Freop60	_Add2	7.65544	1.39023	5.5066	<.0001
	Fair60	_Add3	5.03798	0.98299	5.1251	<.0001
	Legis60	_Add4	3.27028	0.73387	4.4562	<.0001
	Press65	_Add5	2.52969	0.52882	4.7836	<.0001
	Freop65	_Add6	4.87208	0.94384	5.1620	<.0001
	Fair65	_Add7	3.32508	0.71220	4.6687	<.0001
	Legis65	_Add8	3.25392	0.73319	4.4380	<.0001

All error variance estimates are significant.

55

Estimates of the Covariances for the CFA Model with Constrained Loadings and Correlated Errors

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Dem65	Dem60	_Add9	0.96603	0.02928	32.9904	<.0001

Covariances Among Errors

Error of	Error of	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Freop60	Legis60	_Parm1	1.42826	0.69666	2.0502	0.0403
Freop65	Legis65	_Parm2	1.26677	0.59365	2.1339	0.0329
Press60	Press65	_Parm3	0.58548	0.37178	1.5748	0.1153
Freop60	Freop65	_Parm4	2.09624	0.74763	2.8039	0.0050
Fair60	Fair65	_Parm5	0.74805	0.62336	1.2000	0.2301
Legis60	Legis65	_Parm6	0.47686	0.46214	1.0319	0.3021

Bad news: Some error covariance estimates are not significant.

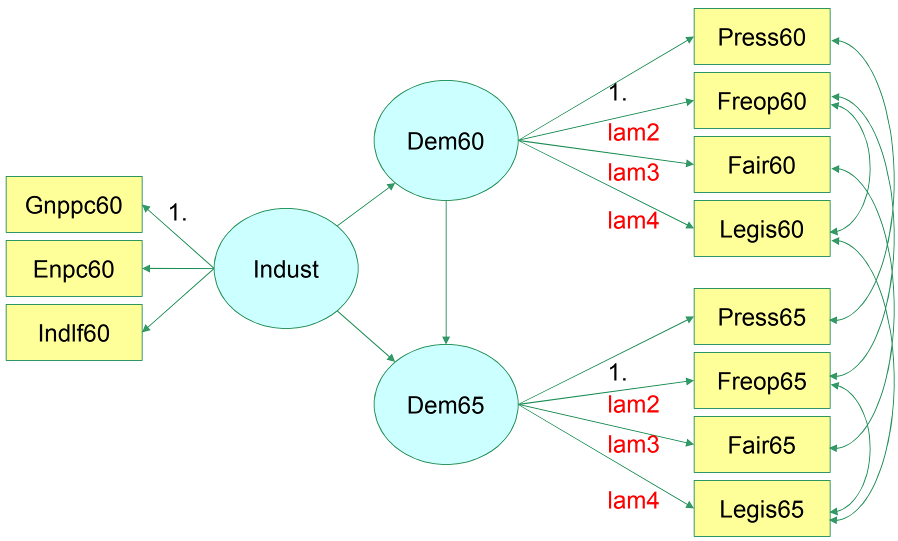
56

Political Democracy and Industrialization: A Full Structural Equation Model

Political Democracy and Industrialization

- Bollen (1989) Chapter 8
- A full structural equation model (a full LISREL model)
- Additional variables for measuring industrialization (Indust) in 1960
  - Gross national product per capita (Gnppc60)
  - Energy consumption per capita (Enpc60)
  - Percent of labor force in industrial occupations (Indlf60)
- Purposes: Validate the measurement model and the structural relationships

Political Democracy and Industrialization: The Path Diagram



59

Fitting the Structural Equation Model for the Political Democracy and Industrialization Data

Multiple-path specifications

```
proc calis data=polidem;
   path
      Dem60   ===> Press60 Freop60 Fair60 Legis60   =  1.  lam2  lam3  lam4,
      Dem65   ===> Press65 Freop65 Fair65 Legis65   =  1.  lam2  lam3  lam4,
      Indust  ===> Gnppc60 Enpc60 Indlf60           =  1.,
      Indust  ===> Dem60 Dem65,
      Dem60   ===> Dem65;
   pcov
      Freop60 Legis60, Freop65 Legis65,
      Press60 Press65, Freop60 Freop65,
      Fair60 Fair65, Legis60 Legis65;
run;
```

60

Political Democracy and Industrialization: Fit Summary Table

Fit Summary	
Chi-Square	39.6438
Chi-Square DF	38
Pr > Chi-Square	0.3966
Hoelter Critical N	100
Standardized RMR (SRMR)	0.0558
Adjusted GFI (AGFI)	0.8606
RMSEA Estimate	0.0242
Bentler Comparative Fit Index	0.9975

Not a bad fit for the data.

61

Political Democracy and Industrialization: Estimates of Path Coefficients

PATH List

----------Path-----------			Parameter	Estimate	Standard Error	t Value	Pr > |t|
Dem60	===>	Press60		1.00000			
Dem60	===>	Freop60	lam2	1.19079	0.14020	8.4934	<.0001
Dem60	===>	Fair60	lam3	1.17454	0.12121	9.6899	<.0001
Dem60	===>	Legis60	lam4	1.25099	0.11757	10.6401	<.0001
Dem65	===>	Press65		1.00000			
Dem65	===>	Freop65	lam2	1.19079	0.14020	8.4934	<.0001
Dem65	===>	Fair65	lam3	1.17454	0.12121	9.6899	<.0001
Dem65	===>	Legis65	lam4	1.25099	0.11757	10.6401	<.0001
Indust	===>	Gnppc60		1.00000			
Indust	===>	Enpc60	_Parm01	2.17966	0.13932	15.6453	<.0001
Indust	===>	Indlf60	_Parm02	1.81821	0.15290	11.8913	<.0001
Indust	===>	Dem60	_Parm03	1.47133	0.39496	3.7253	0.0002
Indust	===>	Dem65	_Parm04	0.60046	0.22722	2.6427	0.0082
Dem60	===>	Dem65	_Parm05	0.86504	0.07538	11.4765	<.0001

62

Political Democracy and Industrialization: Estimates of Variances

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Exogenous	Indust	_Add01	0.45466	0.08846	5.1399	<.0001
Error	Press60	_Add02	1.87973	0.44229	4.2500	<.0001
	Freop60	_Add03	7.68378	1.39404	5.5119	<.0001
	Fair60	_Add04	5.02270	0.97587	5.1469	<.0001
	Legis60	_Add05	3.26801	0.73807	4.4278	<.0001
	Press65	_Add06	2.34432	0.48851	4.7990	<.0001
	Freop65	_Add07	5.03534	0.93993	5.3572	<.0001
	Fair65	_Add08	3.60813	0.72394	4.9840	<.0001
	Legis65	_Add09	3.35236	0.71788	4.6698	<.0001
	Gnppc60	_Add10	0.08249	0.01986	4.1538	<.0001
	Enpc60	_Add11	0.12206	0.07105	1.7178	0.0858
	Indlf60	_Add12	0.47297	0.09197	5.1427	<.0001
	Dem60	_Add13	3.92767	0.88311	4.4475	<.0001
	Dem65	_Add14	0.16668	0.23158	0.7197	0.4717

63

Political Democracy and Industrialization: Estimates of Covariances

Covariances Among Errors

Error of	Error of	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Freop60	Legis60	_Parm06	1.45956	0.70251	2.0776	0.0377
Freop65	Legis65	_Parm07	1.39032	0.58859	2.3621	0.0182
Press60	Press65	_Parm08	0.59042	0.36307	1.6262	0.1039
Freop60	Freop65	_Parm09	2.21252	0.75242	2.9405	0.0033
Fair60	Fair65	_Parm10	0.72123	0.62333	1.1571	0.2472
Legis60	Legis65	_Parm11	0.36769	0.45324	0.8112	0.4172

64

Political Democracy and Industrialization: Squared Multiple Correlations

Squared Multiple Correlations

Variable	Error Variance	Total Variance	R-Square
Enpc60	0.12206	2.28211	0.9465
Fair60	5.02270	11.79895	0.5743
Fair65	3.60813	10.09361	0.6425
Freop60	7.68378	14.64875	0.4755
Freop65	5.03534	11.70144	0.5697
Gnppc60	0.08249	0.53715	0.8464
Indlf60	0.47297	1.97602	0.7606
Legis60	3.26801	10.95502	0.7017
Legis65	3.35236	10.70953	0.6870
Press60	1.87973	6.79166	0.7232
Press65	2.34432	7.04548	0.6673
Dem60	3.92767	4.91193	0.2004
Dem65	0.16668	4.70116	0.9645

65

The Extended PATH Modeling Language

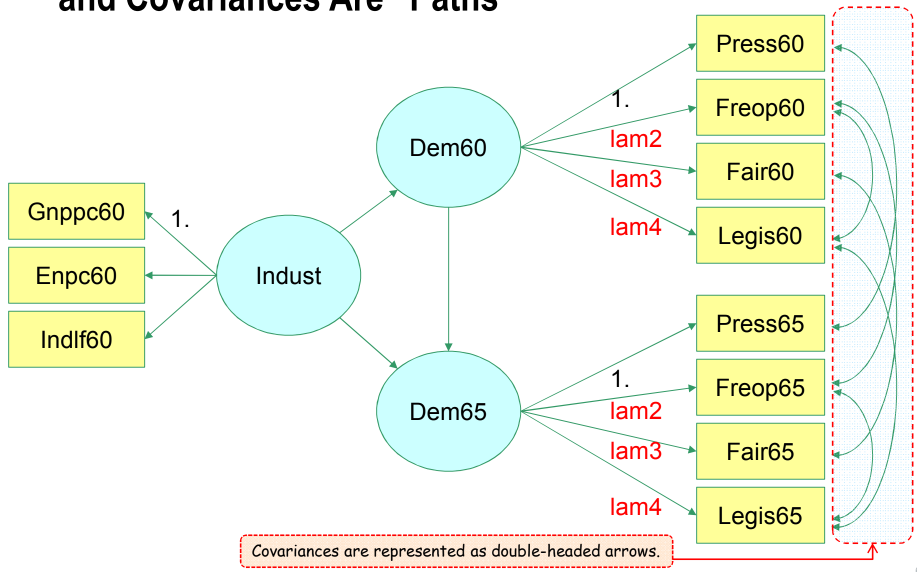
66

Features of the Generalized PATH Modeling Language

- Extension of the PATH modeling language
- Represents all generalized paths in the PATH statement
- Variance-path: Y <==> Y
- Covariance-path: X <==> Y
- Mean or intercept (one-path): 1 ===> Y

67

Political Democracy and Industrialization: Variances and Covariances Are "Paths"



68

An Example of the Generalized PATH Modeling Language

```
PROC CALIS DATA=polidem;
   PATH
      Dem60  ===> Press60 Freop60 Fair60 Legis60    =  1.  lam2  lam3  lam4,
      Dem65  ===> Press65 Freop65 Fair65 Legis65    =  1.  lam2  lam3  lam4,
      Indust ===> Gnppc60 Enpc60 Indlf60            =  1.,
      Indust ===> Dem60 Dem65,
      Dem60  ===> Dem65;
   PCOV
      Freop60 Legis60, Freop65 Legis65,
      Press60 Press65, Freop60 Freop65,
      Fair60 Fair65, Legis60 Legis65;
```

Use the PCOV statement to specify the covariances

```
proc calis data=polidem;
   path
      Dem60  ===> Press60 Freop60 Fair60 Legis60    =  1.  lam2  lam3  lam4,
      Dem65  ===> Press65 Freop65 Fair65 Legis65    =  1.  lam2  lam3  lam4,
      Indust ===> Gnppc60 Enpc60 Indlf60            =  1.,
      Indust ===> Dem60 Dem65,
      Dem60  ===> Dem65,
      Freop60 <==> Legis60, Freop65 <==> Legis65,
      Press60 <==> Press65, Freop60 <==> Freop65,
      Fair60  <==> Fair65,  Legis60 <==> Legis65;
```

Use the generalized path to specify covariances (and variances)

69

Political Democracy and Industrialization: Output with Generalized Paths

PATH List

-----------Path-----------			Parameter	Estimate	Standard Error	t Value	Pr > |t|
Dem60	===>	Press60		1.00000			
Dem60	===>	Freop60	lam2	1.19079	0.14020	8.4934	<.0001
Dem60	===>	Fair60	lam3	1.17454	0.12121	9.6899	<.0001
Dem60	===>	Legis60	lam4	1.25099	0.11757	10.6401	<.0001
Dem65	===>	Press65		1.00000			
Dem65	===>	Freop65	lam2	1.19079	0.14020	8.4934	<.0001
Dem65	===>	Fair65	lam3	1.17454	0.12121	9.6899	<.0001
Dem65	===>	Legis65	lam4	1.25099	0.11757	10.6401	<.0001
Indust	===>	Gnppc60		1.00000			
Indust	===>	Enpc60	_Parm01	2.17966	0.13932	15.6453	<.0001
Indust	===>	Indlf60	_Parm02	1.81821	0.15290	11.8913	<.0001
Indust	===>	Dem60	_Parm03	1.47133	0.39496	3.7253	0.0002
Indust	===>	Dem65	_Parm04	0.60046	0.22722	2.6427	0.0082
Dem60	===>	Dem65	_Parm05	0.86504	0.07538	11.4765	<.0001
Freop60	<==>	Legis60	_Parm06	1.45956	0.70251	2.0776	0.0377
Freop65	<==>	Legis65	_Parm07	1.39032	0.58859	2.3621	0.0182
Press60	<==>	Press65	_Parm08	0.59042	0.36307	1.6262	0.1039
Freop60	<==>	Freop65	_Parm09	2.21252	0.75242	2.9405	0.0033
Fair60	<==>	Fair65	_Parm10	0.72123	0.62333	1.1571	0.2472
Legis60	<==>	Legis65	_Parm11	0.36769	0.45324	0.8112	0.4172

70

Default Free and Fixed Parameters in PROC CALIS

- Default free parameters
  - Variances and covariances among all **exogenous** (independent) variables (observed or latent, except for error terms)
  - Error variances for all **endogenous** (dependent) variables
  - Means or intercepts of all **observed** variables
- Default fixed zeros
  - Unspecified paths and error covariances
  - Means or intercepts of all **latent** variables

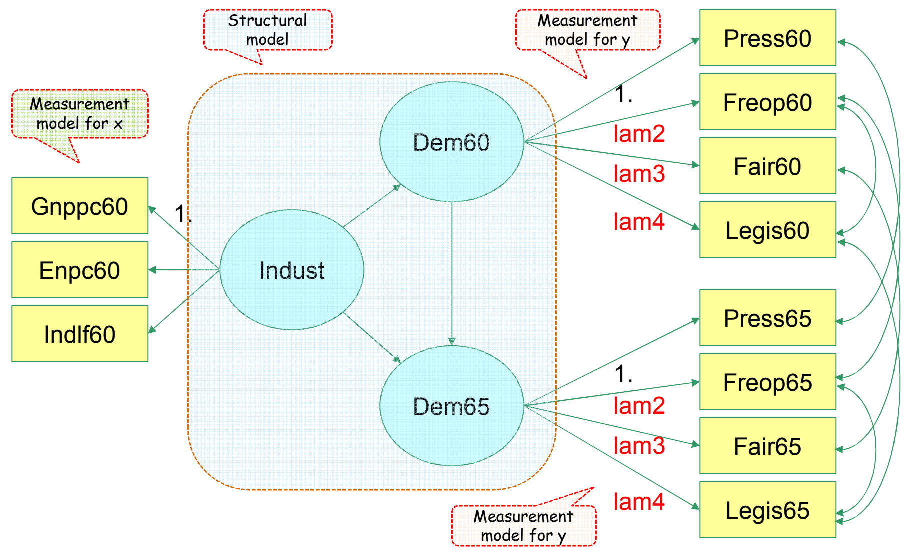
The main purpose of setting default parameters is to enable you to specify only the functional relationships among variables in most practical applications.

71

LISREL Models

72

A LISREL Model for the Political Democracy and Industrialization Data



73

Fitting the Structural Equation Model for the Political Democracy and Industrialization Data

```
proc calis data=polidem;
   path
      Dem60   ===> Press60 Freop60 Fair60 Legis60   =  1.  lam2  lam3  lam4,
      Dem65   ===> Press65 Freop65 Fair65 Legis65   =  1.  lam2  lam3  lam4,
      Indust ===> Gnppc60 Enpc60 Indlf60            =  1.,
      Indust ===> Dem60 Dem65,
      Dem60   ===> Dem65;
   pcov
      Freop60 Legis60, Freop65 Legis65,
      Press60 Press65, Freop60 Freop65,
      Fair60 Fair65, Legis60 Legis65;
run;
```

Measurement model for y

Measurement model for x

Structural model

Measurement model for y

74

A LISREL Model Specified by the LISMOD Modeling Language of PROC CALIS

```
proc calis data=polidem nose noparmname;
  lismod
     xvar = Gnppc60 Enpc60 Indlf60,
     yvar = Press60 Freop60 Fair60 Legis60 Press65 Freop65 Fair65 Legis65,
     xi   = Indust,
     eta  = Dem60 Dem65;
  matrix
     _LambdaY_  [ 1, @1] = 1.  lam2  lam3  lam4,    /* Paths from Dem60 and Dem65 to yvar */
                [ 5, @2] = 1.  lam2  lam3  lam4;
  matrix
     _ThetaY_   [ 4, 2], [ 8, 6],                   /* pcov statement in the path model */
                [ 5, 1], [ 6, 2], [ 7, 3], [ 8, 4];
  matrix
     _LambdaX_  [ 1, 1]   = 1.,                     /* Path from Indust to xvar*/
                [ 2 to 3, 1];
  matrix
     _Gamma_    [ 1 to 2, 1];
  matrix
     _Beta_     [2,1];
run;
```

Measurement model for y

Measurement model for x

Structural model

75

LISMOD Output for the Political Democracy and Industrialization Data (Measurement for y)

LAMBDAY Matrix

	Dem60	Dem65
Press60	1.0000	0
Freop60	1.1908	0
Fair60	1.1745	0
Legis60	1.2510	0
Press65	0	1.0000
Freop65	0	1.1908
Fair65	0	1.1745
Legis65	0	1.2510

Note: The NOSE and NOPARMNAME options suppress the printing of the standard error estimates and parameter names.

76

LISMOD Output for the Political Democracy and Industrialization Data (Measurement for y)

THETAY Matrix

	Press60	Freop60	Fair60	Legis60	Press65	Freop65	Fair65	Legis65
Press60	1.8797	0	0	0	0.5904	0	0	0
Freop60	0	7.6838	0	1.4596	0	2.2125	0	0
Fair60	0	0	5.0227	0	0	0	0.7212	0
Legis60	0	1.4596	0	3.2680	0	0	0	0.3677
Press65	0.5904	0	0	0	2.3443	0	0	0
Freop65	0	2.2125	0	0	0	5.0353	0	1.3903
Fair65	0	0	0.7212	0	0	0	3.6081	0
Legis65	0	0	0	0.3677	0	1.3903	0	3.3524

Note: Error variances (diagonal elements) were set by default.

77

LISMOD Output for the Political Democracy and Industrialization Data (Measurement for x)

LAMBDAX Matrix

	Indust
Gnppc60	1.0000
Enpc60	2.1797
Indlf60	1.8182

THETAX Matrix

	Gnppc60	Enpc60	Indlf60
Gnppc60	0.0825	0	0
Enpc60	0	0.1221	0
Indlf60	0	0	0.4730

Note: Error variances (diagonal elements in _THETAX_) were set by default.

78

LISMOD Output for the Political Democracy and Industrialization Data (Structural Model)

BETA Matrix

	Dem60	Dem65
Dem60	0	0
Dem65	0.8650	0

GAMMA Matrix

	Indust
Dem60	1.4713
Dem65	0.6005

79

LISMOD Output for the Political Democracy and Industrialization Data (Structural Covariances)

PSI Matrix

	Dem60	Dem65
Dem60	3.9277	0
Dem65	0	0.1667

PHI Matrix

	Indust
Indust	0.4547

Note: All variances (diagonal elements in _PSI_ and _PHI_) were set by default.

80

Features of the LISMOD Modeling Language

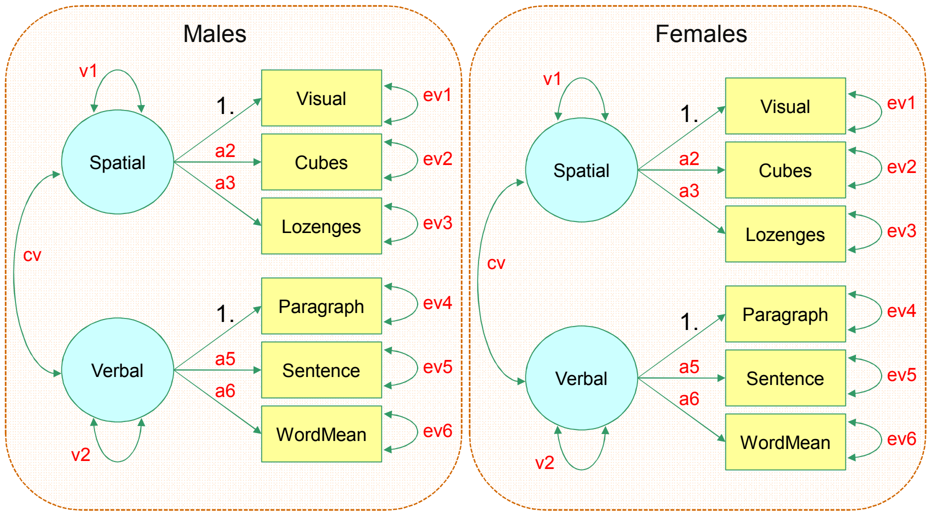
- Supports the JKW (LISREL) models (not the LISREL program)
- Supports mean structure analysis
- Users input:
  - The ordered lists of x, y, ξ, and η variables
  - MATRIX statements to define free and fixed parameters
  - Names for parameters (not required for free parameters)
- Default covariance structure parameters of the LISMOD language:
  - Diagonal elements of all covariance matrices (all variances)
  - Lower triangular elements of the _PHI_ matrix (covariances of the ξ-variables)

81

Multiple-Group Analysis

82

A Completely-Constrained CFA Multiple-Group Model for Males and Females



Data source: Holzinger and Swineford (1939)

83

Different Methods to Specify the Completely-Constrained Multiple-Group Model

1. Two models with the same specifications and same parameters (including the default parameters) for the two groups
2. One model for the two groups
3. Two models for the two groups; one model makes reference to the specification of the other model (REFMODEL statement specification)

84

Method 1: Specify Two Models With Complete Parameter Constraints for the Groups

```
proc calis;
   group 1 / label='Males' data=males;
   group 2 / label='Females' data=females;
   model 1 / group = 1;
      path
         Spatial  ===> Visual Cubes Lozenges       =  1. a2 a3,
         Verbal   ===> Paragraph Sentence Wordmean =  1. a5 a6;
      pvar
         Visual Cubes Lozenges Paragraph Sentence Wordmean = ev1-ev6,
         Spatial = v1, Verbal = v2;
      pcov
         Spatial Verbal = cv;
   model 2 / group = 2;
      path
         Spatial  ===> Visual Cubes Lozenges       =  1. a2 a3,
         Verbal   ===> Paragraph Sentence Wordmean =  1. a5 a6;
      pvar
         Visual Cubes Lozenges Paragraph Sentence Wordmean = ev1-ev6,
         Spatial = v1, Verbal = v2;
      pcov
         Spatial Verbal = cv;
run;
```

Models are constrained through the use of the same parameter names.

85

Comments on Method 1

- Constraints on parameters are set by using the same names.
- You have to enumerate all parameters in the models in order to constrain the two models completely.

86

Method 2: Two Groups Fitted by the Same Model

```
proc calis;
   group 1 / label='Males' data=males;
   group 2 / label='Females' data=females;
   model 1 / group = 1, 2;
      path
         Spatial  ===> Visual Cubes Lozenges       =  1. ,
         Verbal   ===> Paragraph Sentence Wordmean =  1. ;
run;
```

Groups 1 and 2 are fitted by the same model.

Comments on Method 2

- The easiest and quickest way to specify completely-constrained multiple-group models.
- Not applicable to partially-constrained multiple-group models.

Method 3: Model Referencing With the REFMODEL Statement

```
proc calis;
   group 1 / label='Males' data=males;
   group 2 / label='Females' data=females;
   model 1 / group = 1;
      path
         Spatial  ===> Visual Cubes Lozenges       =  1. ,
         Verbal   ===> Paragraph Sentence Wordmean =  1. ;
      pvar
         Visual Cubes Lozenges Paragraph Sentence Wordmean
         Spatial Verbal;
      pcov
         Spatial Verbal;
   model 2 / group = 2;
      refmodel 1;
run;
```

The REFMODEL statement makes reference to all explicit specifications in Model 1.

Note: This method is used for the completely constrained model and the subsequent models with parameter constraints.

89

Comments on Method 3

- All ***explicitly*** specified parameters in the reference model are applied to the model that refers to it.
- For completely-constrained multiple-group models, you still need to specify all parameters in the reference model. However, parameter names are not necessary.
- It is the most convenient method to set up partially-constrained multiple-group models.

90

Fit Summary of the Completely Constrained Multiple-Group Model

Fit Summary

Chi-Square	26.0154
Chi-Square DF	29
Pr > Chi-Square	0.6247
Standardized RMR (SRMR)	0.0968
Adjusted GFI (AGFI)	0.9235
RMSEA Estimate	0.0000
Akaike Information Criterion	52.0154
Bozdogan CAIC	103.7130
Schwarz Bayesian Criterion	90.7130
Bentler Comparative Fit Index	1.0000

Not a bad fit.

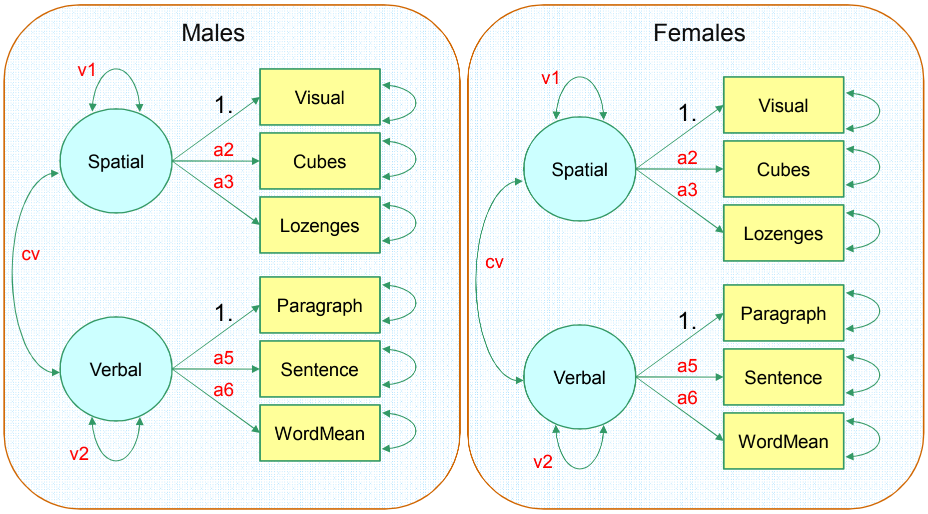
91

Fitting Less Restrictive Multiple-Group Models

- Completely-constrained multiple-group model: Error variances, structural covariances, and loadings are all constrained
- Release the constraints on error variances
- Release the constraints on structural covariances
- Release the constraints on the loadings – Completely unconstrained

92

Release the Constraints on Error Variances



Common parameter names for the error variances are removed.

93

Releasing the Constraints on the Error Variances

```
proc calis;
   group 1 / label='Males' data=males;
   group 2 / label='Females' data=females;
   model 1 / group = 1;
      path
         Spatial  ===> Visual Cubes Lozenges       =  1. ,
         Verbal   ===> Paragraph Sentence Wordmean =  1. ;
      pvar
         /* Visual Cubes Lozenges Paragraph Sentence Wordmean */
         Spatial Verbal;
      pcov
         Spatial Verbal;
   model 2 / group = 2;
      refmodel 1;
run;
```

Comment out the error variance specifications in the PVAR statement, and let PROC CALIS set two distinct sets of default error variances for the two models.

94

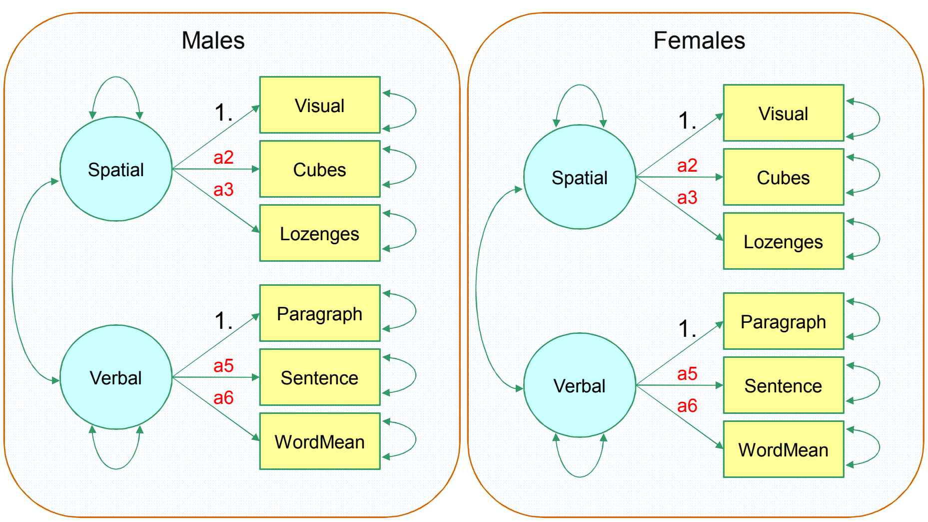
Fit Summary of the Multiple-Group Model with Constraints on Loadings and Structural Covariances

Fit Summary

Chi-Square	22.0334
Chi-Square DF	23
Pr > Chi-Square	0.5182
Standardized RMR (SRMR)	0.0903
Adjusted GFI (AGFI)	0.9163
RMSEA Estimate	0.0000
Akaike Information Criterion	60.0334
Bozdogan CAIC	135.5913
Schwarz Bayesian Criterion	116.5913
Bentler Comparative Fit Index	1.0000

95

Release the Constraints on the Structural Variances and Covariances



Common parameter names for the structural variances and covariance are removed.

96

Releasing the Constraints on the Error Variances and Structural Covariances

```
proc calis;
   group 1 / label='Males' data=males;
   group 2 / label='Females' data=females;
   model 1 / group = 1;
      path
         Spatial  ===> Visual Cubes Lozenges       =  1. ,
         Verbal   ===> Paragraph Sentence Wordmean =  1. ;
      /*
      pvar
         Visual Cubes Lozenges Paragraph Sentence Wordmean
         Spatial Verbal;
      pcov
         Spatial Verbal;
      */
   model 2 / group = 2;
      refmodel 1;
run;
```

Comment out the PVAR and PCOV statements, and let the PROC CALIS set two distinct sets of default variances and covariances for the two models.

97

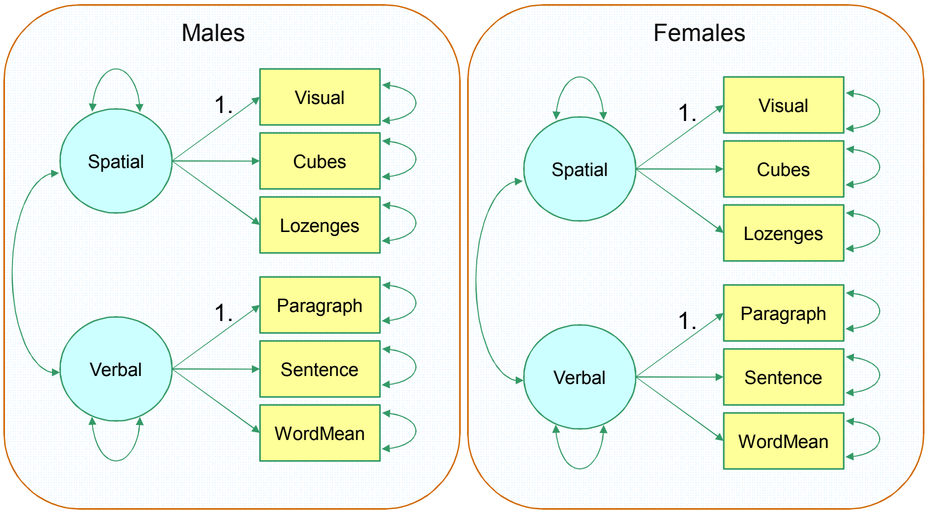
Fit Summary of the Multiple-Group Model with Loading Constraints

Fit Summary

Chi-Square	18.2915
Chi-Square DF	20
Pr > Chi-Square	0.5682
Standardized RMR (SRMR)	0.0539
Adjusted GFI (AGFI)	0.9179
RMSEA Estimate	0.0000
Akaike Information Criterion	62.2915
Bozdogan CAIC	149.7796
Schwarz Bayesian Criterion	127.7796
Bentler Comparative Fit Index	1.0000

98

Release the Constraints on the Loadings



Common parameter names for the loadings are removed.

99

Completely Unconstrained Multiple-Group Model

```
proc calis;
   group 1 / label='Males' data=males;
   group 2 / label='Females' data=females;
   model 1 / group = 1;
      path
         Spatial  ===> Visual Cubes Lozenges        =  1. ,
         Verbal   ===> Paragraph Sentence Wordmean =  1. ;
   model 2 / group = 2;
      path
         Spatial  ===> Visual Cubes Lozenges        =  1. ,
         Verbal   ===> Paragraph Sentence Wordmean =  1. ;
run;
```

The REFMODEL statement is not used here because the parameters in the two models are not constrained with each other.

100

Fit Summary of the Completely Unconstrained Multiple-Group Model

Fit Summary	
Chi-Square	16.4795
Chi-Square DF	16
Pr > Chi-Square	0.4200
Standardized RMR (SRMR)	0.0449
Adjusted GFI (AGFI)	0.9077
RMSEA Estimate	0.0205
Akaike Information Criterion	68.4795
Bozdogan CAIC	171.8746
Schwarz Bayesian Criterion	145.8746
Bentler Comparative Fit Index	0.9984

The unconstrained SEM model for the groups gives you the best fit, but it is also the least interesting multiple-group model.

101

Chi-Square Difference Tests for the Nested Multiple-Group Models

	Completely Constrained	Constrained Loadings and Struct. Cov.	Constrained Loadings
Constrained Loadings and Structural Covariances	3.892 (p=0.32)		
Constrained Loadings	7.724 (p=0.44)	3.742 (p=0.71)	
Completely Unconstrained	9.536 (p=.27)	5.539 (p=.41)	1.182 (p=.23)

There are no significant differences between the multiple-group models.

102

Comparing Model Fits by Using Various Fit Indices

	Completely Constrained	Constrained Loadings and Struct. Cov.	Constrained Loadings	Completely Unconstrained
Chi-Square	26.0154	22.0334	18.2915	16.4795
Chi-Square DF	29	23	20	16
Pr > Chi-Square	0.6247	0.5182	0.5682	0.4200
Standardized RMR (SRMR)	0.0968	0.0903	0.0539	0.0449
Adjusted GFI (AGFI)	0.9235	0.9163	0.9179	0.9077
RMSEA Estimate	0.0000	0.0000	0.0000	0.0205
Akaike Information Criterion	52.0154	60.0334	62.2915	68.4795
Bozdogan CAIC	103.7130	135.5913	149.7796	171.8746
Schwarz Bayesian Criterion	90.7130	116.5913	127.7796	145.8746
Bentler Comparative Fit Index	1.0000	1.0000	1.0000	0.9984

Absolute indices: Chi-square, SRMR (smaller is better)
Parsimonious indices: AGFI (larger is better),
RMSEA, AIC, CAIC, SBC (smaller is better)
Incremental indices: Bentler CFI (larger is better)

103

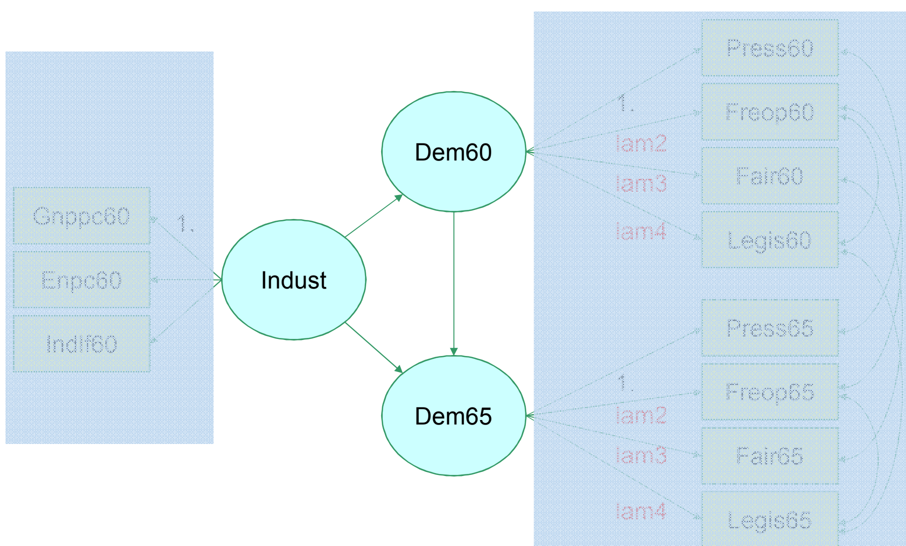
Ideal Characteristics of the "Best" Model Among Competing Models

- Smallest values in the fit indices that takes model parsimony into account. For example, RMSEA, AIC, CAIC, SBC.
- Acceptable absolute and comparative fit statistics. For example, SRMR less than .05 and Bentler's CFI larger than .9.
- Substantively meaningful.

104

Analyzing Direct and Indirect Effects

Democracy and Industrialization Data: Direct and Indirect Effects of Industrialization



Industrialization Effects on Democracy in 1960 and 1965

- On democracy in 1960

  A direct effect:

  Indust ===>Dem60

- On democracy in 1965

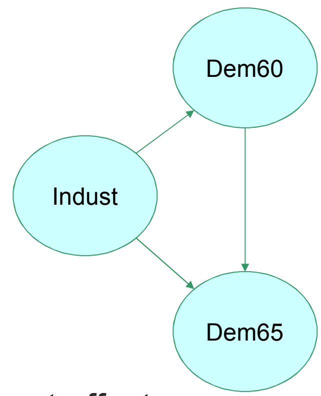
  A direct effect:

  Indust ===>Dem65

  An indirect effect:

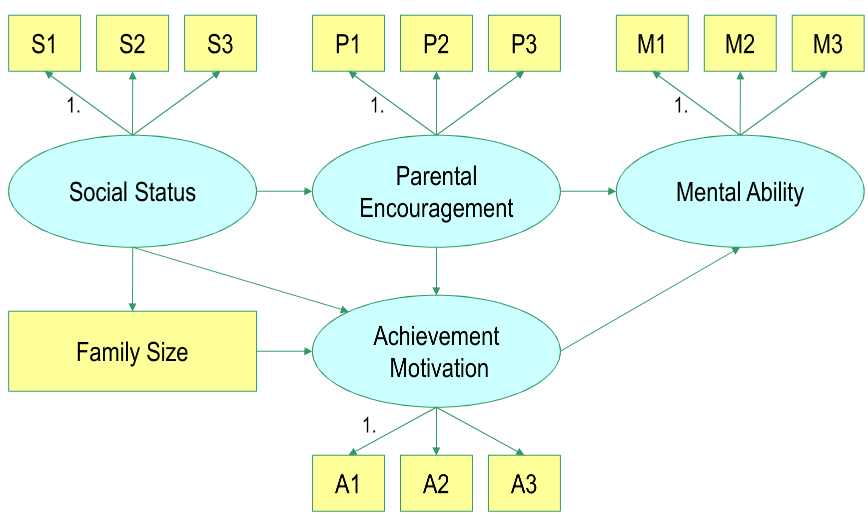
  Indust ===>Dem60 ===>Dem65

- Total effect = direct effect + indirect effect



107

Factors Affecting Mental Abilities: A Model Inspired by Marjoribanks (1974)



108

Factors Affecting Mental Abilities: PROC CALIS Code

```
proc calis data=mental nobs=115 effpart;
   path
      /* Structural Model */
      SocialStatus ===>  ParentalEncouragement FamilySize
                         AchievementMotivation,
      FamilySize    ===>  AchievementMotivation,
      ParentalEncouragement  ===>  AchievementMotivation MentalAbility,
      AchievementMotivation  ===>  MentalAbility,

      /* Measurement Model */
      SocialStatus          ===> S1 S2 S3   = 1.,
      ParentalEncouragement ===> P1 P2 P3   = 1.,
      AchievementMotivation ===> A1 A2 A3   = 1.,
      MentalAbility         ===> M1 M2 M3   = 1.;
run;
```

The EFFPART option analyzes the effect partitioning in the model.

109

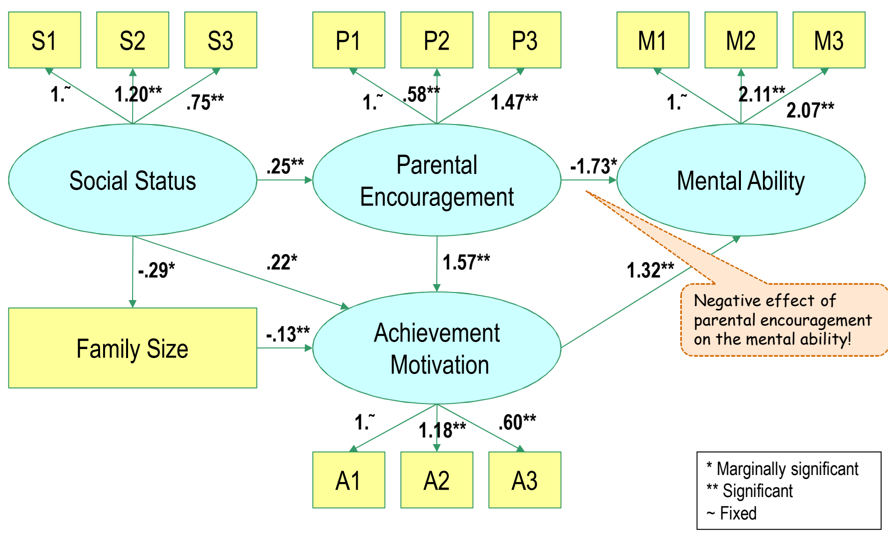
Fit Summary

Fit Summary	
Standardized RMR (SRMR)	0.0936
Adjusted GFI (AGFI)	0.7341
RMSEA Estimate	0.1431
Bentler Comparative Fit Index	0.8087

Not a very good model fit.

110

Factors Affecting Mental Abilities: Estimation Results



111

Partitioning of the Effects: A Prerequisite

```
Stability Coefficient of Reciprocal Causation = 0

        Stability Coefficient < 1

      Total and Indirect Effects Converge
```

```
NOTE: The stability coefficient is 0, which is less
      than one. The condition for converged total and
      indirect effects is satisfied.
```

112

Partitioning of the Effects: Total Effects

Total Effects
Effect / Std Error / t Value / p Value

	FamilySize	Achievement Motivation	Mental Ability	Parental Encouragement	SocialStatus
A1	-0.1287 0.0360 -3.5769 0.000348	1.0000	0	1.5701 0.5176 3.0337 0.002416	0.6523 0.0942 6.9258 <.0001
A2	-0.1522 0.0420 -3.6274 0.000286	1.1821 0.1084 10.9012 <.0001	0	1.8560 0.6051 3.0672 0.002161	0.7710 0.1036 7.4398 <.0001
.	.	.	.	.	.
MentalAbility	-0.1699 0.0572 -2.9696 0.002982	1.3196 0.4124 3.1995 0.001377	0	0.3376 0.4045 0.8346 0.4039	0.4244 0.1280 3.3159 0.000914
ParentalEncouragement	0	0	0	0	0.2516 0.0692 3.6356 0.000277

Details omitted.

113

Partitioning of the Effects: Direct Effects

Direct Effects
Effect / Std Error / t Value / p Value

	FamilySize	Achievement Motivation	Mental Ability	Parental Encouragement	SocialStatus
A1	0	1.0000	0	0	0
A2	0	1.1821 0.1084 10.9012 <.0001	0	0	0
.	.	.	.	.	.
MentalAbility	0	1.3196 0.4124 3.1995 0.001377	0	-1.7343 0.9047 -1.9169 0.0553	0
ParentalEncouragement	0	0	0	0	0.2516 0.0692 3.6356 0.000277

Details omitted.

114

Partitioning of the Effects: Indirect Effects

Indirect Effects
Effect / Std Error / t Value / p Value

	FamilySize	Achievement Motivation	Mental Ability	Parental Encouragement	SocialStatus
A1	-0.1287 0.0360 -3.5769 0.000348	0	0	1.5701 0.5176 3.0337 0.002416	0.6523 0.0942 6.9258 <.0001
A2	-0.1522 0.0420 -3.6274 0.000286	0	0	1.8560 0.6051 3.0672 0.002161	0.7710 0.1036 7.4398 <.0001
.	.	.	.	.	.
MentalAbility	-0.1699 0.0572 -2.9696 0.002982	0	0	2.0719 1.0483 1.9763 0.0481	0.4244 0.1280 3.3159 0.000914
ParentalEncouragement	0	0	0	0	0

Details omitted.

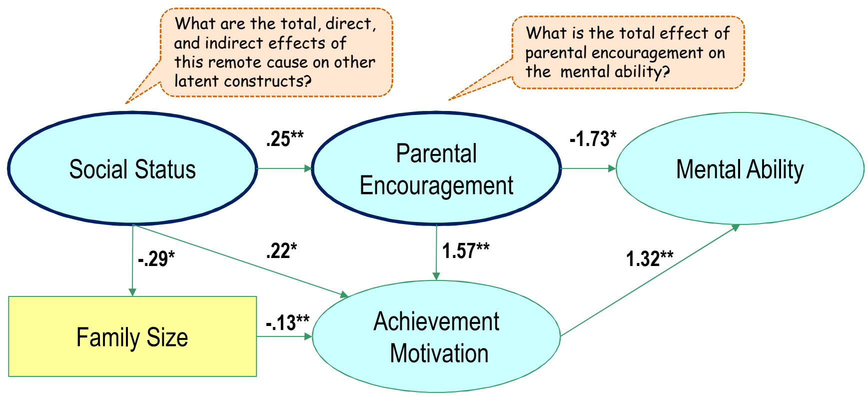
115

Customized Effect Analysis

- The EFFPART option displays all logical possible effects of the variables
- Columns: Five variables, each of which serves as a predictor at least once:
  - FamilySize
  - AchivementMotivation
  - MentalAbility
  - ParentalEncouragement
  - SocialStatus
- Rows: Sixteen variables, each of which serves as an outcome variable at least once (all variables except for SocialStatus)

116

Customized Effect Analysis on the Structural Model



117

Factors Affecting Mental Abilities: Customized Effect Analysis

```
proc calis data=mental nobs=115;
   path
      /* Structural Model */
      SocialStatus ===>  ParentalEncouragement FamilySize AchievementMotivation,
      FamilySize   ===>  AchievementMotivation,
      ParentalEncouragement  ===>  AchievementMotivation MentalAbility,
      AchievementMotivation  ===>  MentalAbility,

      /* Measurement Model */
      SocialStatus           ===> S1  S2  S3   = 1.,
      ParentalEncouragement ===> P1  P2  P3   = 1.,
      AchievementMotivation ===> A1  A2  A3   = 1.,
      MentalAbility          ===> M1  M2  M3   = 1.;
   effpart
      SocialStatus           ===> ParentalEncouragement AchievementMotivation
                                  MentalAbility,
      ParentalEncouragement ===> MentalAbility;
run;
```

The EFFPART statement defines the customized effects of interest.

118

Effects of Social Status

Effects of SocialStatus
Effect / Std Error / t Value / p Value

	Total	Direct	Indirect
ParentalEncouragement	0.2516	0.2516	0
	0.0692	0.0692	
	3.6356	3.6356	
	0.000277	0.000277	
AchievementMotivation	0.6523	0.2193	0.4330
	0.0942	0.1147	0.1203
	6.9258	1.9125	3.5985
	<.0001	0.0558	0.000320
MentalAbility	0.4244	0	0.4244
	0.1280		0.1280
	3.3159		3.3159
	0.000914		0.000914

Direct effect only

Direct and indirect effects

Indirect effect only

119

Effects of Parental Encouragement on Mental Ability

Effects of ParentalEncouragement
Effect / Std Error / t Value / p Value

	Total	Direct	Indirect
MentalAbility	0.3376	-1.7343	2.0719
	0.4045	0.9047	1.0483
	0.8346	-1.9169	1.9763
	0.4039	0.0553	0.0481

Parental encouragement has a positive total effect on the mental ability, although the total effect is not significant.

120

Standardized Effects of Social Status

Standardized Effects of SocialStatus
Effect / Std Error / t Value / p Value

	Total	Direct	Indirect
ParentalEncouragement	0.6675	0.6675	0
	0.0833	0.0833	
	8.0141	8.0141	
	<.0001	<.0001	
AchievementMotivation	0.6990	0.2350	0.4640
	0.0597	0.1207	0.1153
	11.7105	1.9478	4.0231
	<.0001	0.0514	<.0001
MentalAbility	0.4960	0	0.4960
	0.0850		0.0850
	5.8374		5.8374
	<.0001		<.0001

121

Standardized Effects of Parental Encouragement on Mental Ability

Standardized Effects of ParentalEncouragement
Effect / Std Error / t Value / p Value

	Total	Direct	Indirect
MentalAbility	0.1487	-0.7639	0.9126
	0.1705	0.3096	0.3487
	0.8722	-2.4675	2.6171
	0.3831	0.0136	0.008869

Parental encouragement has a positive standardized total effect on the mental ability, although the standardized total effect is not significant.

122

Creating Path Diagrams (SAS/STAT 13.1)

123

PLOTS=PATHDIAGRAM Option

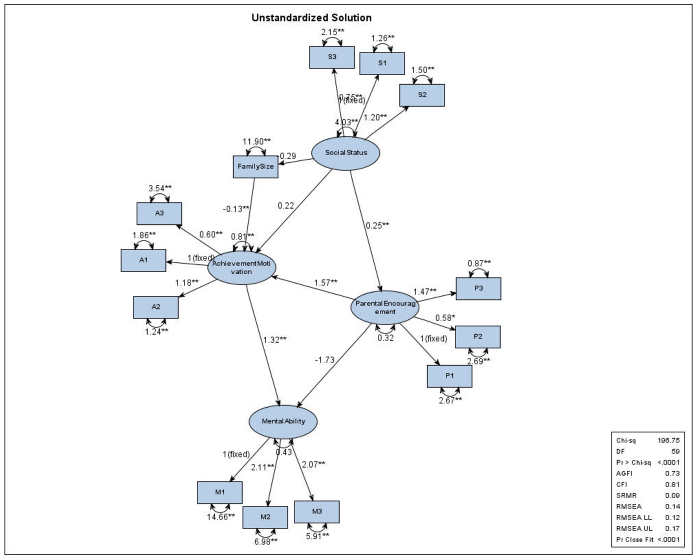
```
ods graphics on;

proc calis data=mental nobs=115 plots=pathdiagram;
   path
      /* Structural Model */
      SocialStatus ===>  ParentalEncouragement FamilySize AchievementMotivation,
      FamilySize   ===>  AchievementMotivation,
      ParentalEncouragement  ===>  AchievementMotivation MentalAbility,
      AchievementMotivation  ===>  MentalAbility,

      /* Measurement Model */
      SocialStatus          ===> S1  S2  S3   = 1.,
      ParentalEncouragement ===> P1  P2  P3   = 1.,
      AchievementMotivation ===> A1  A2  A3   = 1.,
      MentalAbility         ===> M1  M2  M3   = 1.;
run;
```

124

Path Diagram for the Full Model



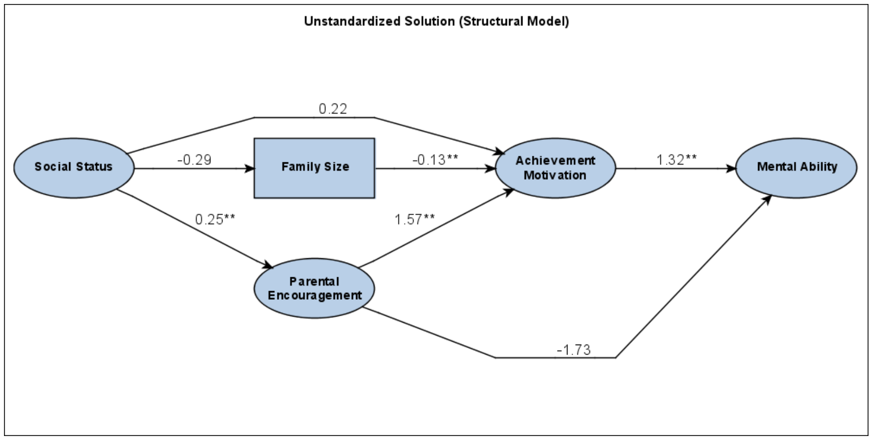
125

PATHDIAGRAM Statement

```
Ods graphics on;
proc calis data=mental nobs=115;
   path
      SocialStatus ===>  ParentalEncouragement FamilySize AchievementMotivation,
      FamilySize   ===>  AchievementMotivation,
      ParentalEncouragement  ===>  AchievementMotivation MentalAbility,
      AchievementMotivation  ===>  MentalAbility,
      SocialStatus          ===> S1  S2  S3    = 1.,
      ParentalEncouragement ===> P1  P2  P3    = 1.,
      AchievementMotivation ===> A1  A2  A3    = 1.,
      MentalAbility         ===> M1  M2  M3    = 1.;
   pathdiagram structural(only) structadd=[FamilySize]
              nofittable arrange=flow novariance
              label=[SocialStatus          ='Social Status'
                     FamilySize            = 'Family Size'
                     ParentalEncouragement = 'Parental Encouragement'
                     AchievementMotivation = 'Achievement Motivation'
                     MentalAbility         = 'Mental Ability'];
run;
```

126

Path Diagram for the Structural Model

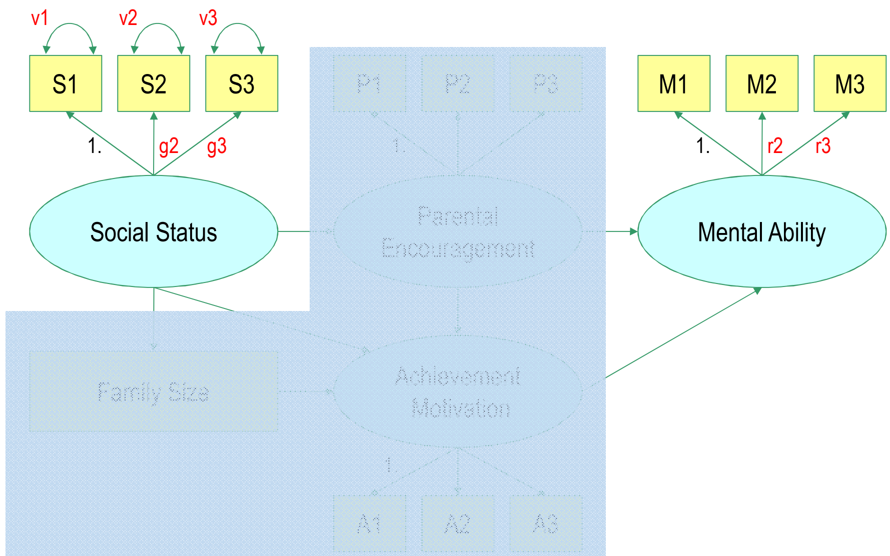


127

Testing Specific Hypotheses

128

Testing Hypotheses about the Measurement Model



129

Specific Hypotheses

- Parallel items for measuring SocialStatus

  H1 : g2 = 1

  H2: g3 = 1

  H3: v1 = v2

  H4: v2 = v3

- Equality of loadings for MentalAbility items M2 and M3

  H5 : r2 = r3

- Sum of the loadings for M2 and M3 is two times as much as the sum of the loadings for S1 and S2

  H6: (r2 + r3) / (g2 + g3) = 2

130

PROC CALIS Hypotheses Testing: h(θ) = 0

- Parallel items for measuring SocialStatus:
  - H1 : h1 = g2 — 1 = 0
  - H2: h2 = g3 — 1 = 0
  - H3 : h3 = v1 — v2 = 0
  - H4: h4 = v2 — v3 = 0
- Equality of loadings for MentalAbility items M2 and M3
  - H5 : h5 = r2 — r3 = 0
- Sum of the loadings for M2 and M3 is two times as much as the sum of the loadings for S1 and S2
  - H6: h6 =2(g2 + g3) — (r2 + r3) = 0

131

Testing Specific Hypotheses about the Measurement Model Using PROC CALIS

```
proc calis data=mental nobs=115;
   path
      SocialStatus ===> ParentalEncouragement FamilySize AchievementMotivation,
      FamilySize   ===> AchievementMotivation,
      ParentalEncouragement  ===>  AchievementMotivation MentalAbility,
      AchievementMotivation  ===>  MentalAbility,
      SocialStatus           ===> S1 S2 S3    = 1. g2 g3,
      ParentalEncouragement  ===> P1 P2 P3    = 1.,
      AchievementMotivation  ===> A1 A2 A3    = 1.,
      MentalAbility          ===> M1 M2 M3    = 1. r2 r3;
   pvar S1-S3 = v1-v3;
   simtest parallel_social_items=[h1 h2 h3 h4];
   testfunc h5_equal_load_m2_m3 h6_proportional_sum;
   h1 = g2 - 1;
   h2 = g3 - 1;
   h3 = v1 - v2;
   h4 = v2 - v3;
   h5_equal_load_m2_m3 = r2 - r3;
   h6_proportional_sum = 2*(g2 + g3) - (r2 + r3);
run;
```

Specify g2, g3, r2, r3, v1, v2, and v3 explicitly.

Use the SIMTEST statement to test simultaneously hypotheses. Use the TESTFUNC statement to test individual hypotheses.

Use the SAS programming statements to define the parametric functions in the tests.

132

Individual Tests of Parametric Functions: TESTFUNC Results

Tests for Parametric Functions

Parametric Function	Estimate	Standard Error	t Value	Pr > |t|
h5_equal_load_M2_M3	0.04147	0.24290	0.1707	0.8644
h6_proportional_sum	-0.27995	1.02816	-0.2723	0.7854

Both individual hypotheses are supported.

133

Tests for Parallel Social Status Items: SIMTESTS Results

Simultaneous Tests

Simultaneous Test	Parametric Function	Function Value	DF	Chi-Square	p Value
parallel_social_items			4	24.23862	<.0001
	h1	0.19873	1	3.57013	0.0588
	h2	-0.25004	1	8.37521	0.0038
	h3	-0.24067	1	0.17774	0.6733
	h4	-0.64108	1	1.49312	0.2217

Overall parallelism hypothesis is not supported for the SocialStatus items, although the equality of error variances is supported.

134

Model Modifications

When the Model Does Not Fit Well ...

Fit Summary

Chi-Square	196.7455
Chi-Square DF	59
Pr > Chi-Square	<.0001
Standardized RMR (SRMR)	0.0936
Adjusted GFI (AGFI)	0.7341
RMSEA Estimate	0.1431
Bentler Comparative Fit Index	0.8087

- Large SRMR and RMSEA
- Small AGFI and CFI
- Model modification: suggests ways to improve the model fit
- Lagrange multiplier (LM) tests: which parameters you can add to significantly decrease the model fit chi-square value

Using the MODIFICATION Option

```
proc calis data=mental nobs=115 modification;
   path
      SocialStatus ===>  ParentalEncouragement FamilySize
                         AchievementMotivation,
      FamilySize   ===>  AchievementMotivation,
      ParentalEncouragement  ===>  AchievementMotivation MentalAbility,
      AchievementMotivation  ===>  MentalAbility,
      SocialStatus           ===> S1  S2  S3    = 1.  ,
      ParentalEncouragement  ===> P1  P2  P3    = 1.,
      AchievementMotivation  ===> A1  A2  A3    = 1.,
      MentalAbility          ===> M1  M2  M3    = 1.;
run;
```

137

LM Tests for Paths

Rank Order of the 10 Largest LM Stat for Path Relations

To	From	LM Stat	Pr > ChiSq	Parm Change
P2	P1	56.19414	<.0001	-0.73639
P1	P2	56.19349	<.0001	-0.72904
A2	M2	19.17647	<.0001	0.22842
A2	ParentalEncouragement	18.57947	<.0001	-2.31463
A2	MentalAbility	17.20340	<.0001	0.95581
ParentalEncouragement	A1	17.04464	<.0001	0.27042
A1	ParentalEncouragement	15.86099	<.0001	1.88904
FamilySize	A2	14.43548	0.0001	-1.14590
A1	MentalAbility	13.88705	0.0002	-0.75314
A2	P3	12.96818	0.0003	-0.57151

Adding the P2 <=== P1 (or P1 <=== P2) path reduces your model fit chi-square by 56 approximately.
Adding the A2 <=== M2 path reduces your model fit chi-square by 19 approximately.

138

LM Tests for Error Variances and Covariances

Rank Order of the 10 Largest LM Stat for Error Variances and Covariances

Error of	Error of	LM Stat	Pr > ChiSq	Parm Change
P2	P1	56.19312	<.0001	-1.96473
ParentalEncouragement	A1	12.26622	0.0005	0.46050
ParentalEncouragement	A2	12.08031	0.0005	-0.48351
FamilySize	A2	11.22650	0.0008	-1.88205
M2	A2	10.26408	0.0014	1.55895
S2	S1	7.78117	0.0053	1.55314
MentalAbility	A2	7.48800	0.0062	0.78161
AchievementMotivation	A1	6.95709	0.0083	-0.52904
P2	A3	6.54315	0.0105	0.76007
A3	A2	6.21429	0.0127	-0.67173

Adding the error covariance (P2 <===> P1) reduces your model fit chi-square by 56 approximately.

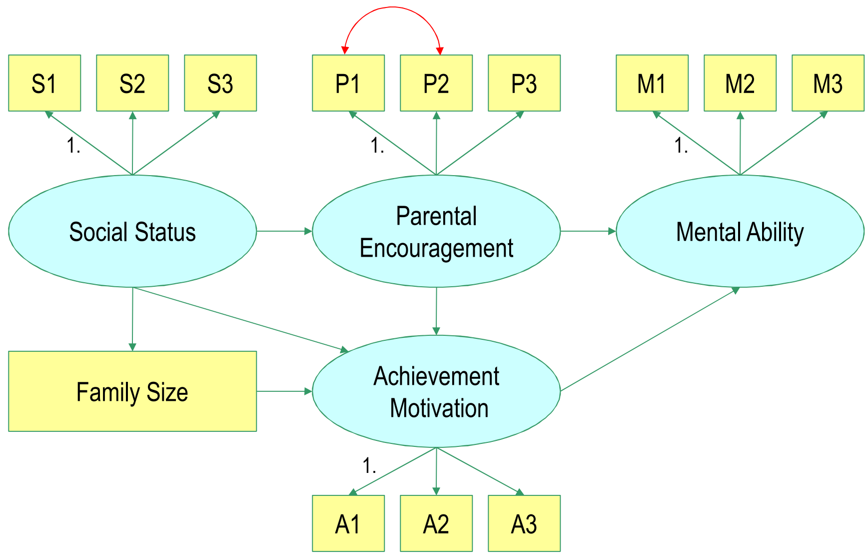
139

Notes on the LM Statistics

- Chi-square reductions are linear approximations
- Chi-square reductions are not additive
- Modifications suggested might not be substantively meaningful

140

Adding Error Covariance between P1 and P2



141

Adding Covariance between the Errors of P1 and P2

```
proc calis data=mental nobs=115 modification;
   path
      SocialStatus ===> ParentalEncouragement FamilySize
                        AchievementMotivation,
      FamilySize   ===> AchievementMotivation,
      ParentalEncouragement  ===> AchievementMotivation MentalAbility,
      AchievementMotivation  ===> MentalAbility,
      SocialStatus           ===> S1 S2 S3   = 1. ,
      ParentalEncouragement  ===> P1 P2 P3   = 1.,
      AchievementMotivation  ===> A1 A2 A3   = 1.,
      MentalAbility          ===> M1 M2 M3   = 1.;
   pcov P1 P2;
run;
```

142

Before and After Adding the Error Covariance between P1 and P2

Before …

Fit Summary

Chi-Square	196.7455
Chi-Square DF	59
Pr > Chi-Square	<.0001
Standardized RMR (SRMR)	0.0936
Adjusted GFI (AGFI)	0.7341
RMSEA Estimate	0.1431
Bentler Comparative Fit Index	0.8087

After…

Fit Summary

Chi-Square	110.6388
Chi-Square DF	58
Pr > Chi-Square	<.0001
Standardized RMR (SRMR)	0.0661
Adjusted GFI (AGFI)	0.8062
RMSEA Estimate	0.0892
Bentler Comparative Fit Index	0.9269

Improve the model fit chi-square a lot more than 56.

143

A New Set of LM Tests for Paths

Rank Order of the 10 Largest LM Stat for Path Relations

To	From	LM Stat	Pr > ChiSq	Parm Change
A2	M2	23.93741	<.0001	0.23830
A2	ParentalEncouragement	22.22260	<.0001	-2.30398
A2	MentalAbility	21.70998	<.0001	0.90790
A1	ParentalEncouragement	19.34281	<.0001	1.96513
A1	MentalAbility	18.10545	<.0001	-0.75752
ParentalEncouragement	A1	17.30632	<.0001	0.32504
A1	M2	15.06992	0.0001	-0.17617
A2	FamilySize	15.06786	0.0001	-0.17675
FamilySize	A2	14.29265	0.0002	-0.89658
AchievementMotivation	A1	11.75569	0.0006	-0.35962

Adding the A2 <=== M2 path now reduces your model fit chi-square by 24 (was 19 before adding the error covariance).

144

A New Set of LM Tests for Error Variances and Covariances

Rank Order of the 10 Largest LM Stat for Error Variances and Covariances

Error of	Error of	LM Stat	Pr > ChiSq	Parm Change
ParentalEncouragement	A2	14.11717	0.0002	-0.53966
ParentalEncouragement	A1	14.00823	0.0002	0.51242
FamilySize	A2	13.96827	0.0002	-1.98162
M2	A2	11.86015	0.0006	1.65462
AchievementMotivation	A1	11.75570	0.0006	-0.57402
MentalAbility	A2	10.30441	0.0013	0.86542
P1	A1	9.02987	0.0027	0.53357
S2	S1	8.94243	0.0028	1.66876
MentalAbility	FamilySize	8.05365	0.0045	2.49300
MentalAbility	AchievementMotivation	6.94045	0.0084	0.98947

145

Customized LM Tests

- Principled modification process
- Restrict the set of parameters of interest for the LM tests

146

Customized LM Tests by Using the LMTESTS Statement

```
proc calis data=mental nobs=115;
   path
      SocialStatus ===> ParentalEncouragement FamilySize AchievementMotivation,
      FamilySize   ===> AchievementMotivation,
      ParentalEncouragement  ===> AchievementMotivation MentalAbility,
      AchievementMotivation  ===> MentalAbility,
      SocialStatus           ===> S1 S2 S3   = 1. ,
      ParentalEncouragement  ===> P1 P2 P3   = 1.,
      AchievementMotivation  ===> A1 A2 A3   = 1.,
      MentalAbility          ===> M1 M2 M3   = 1.;
   lmtests corr_err=[coverr] path=[LV->LV LV->MV];
run;
```

Explore the set of LM tests called "corr_err," which contains all the potential error covariance parameters (COVERR) to be freed.

Explore the set of LM tests called "path," which contains all potential latent variable paths (LV-> LV) and measurement paths (LV -> MV) to be freed.

147

Customized LM Tests for Error Covariances

Rank Order of the 10 Largest LM Stat for Set corr_err

Type	Var1	Var2	LM Stat	Pr > ChiSq	Parm Change
COVERR	P2	P1	56.19312	<.0001	-1.96473
COVERR	ParentalEncouragement	A1	12.26622	0.0005	0.46050
COVERR	ParentalEncouragement	A2	12.08031	0.0005	-0.48351
COVERR	FamilySize	A2	11.22650	0.0008	-1.88205
COVERR	M2	A2	10.26408	0.0014	1.55895
COVERR	S2	S1	7.78117	0.0053	1.55314
COVERR	MentalAbility	A2	7.48800	0.0062	0.78161
COVERR	AchievementMotivation	A1	6.95709	0.0083	-0.52904
COVERR	P2	A3	6.54315	0.0105	0.76007
COVERR	A3	A2	6.21429	0.0127	-0.67173

148

Customized LM Tests for Paths

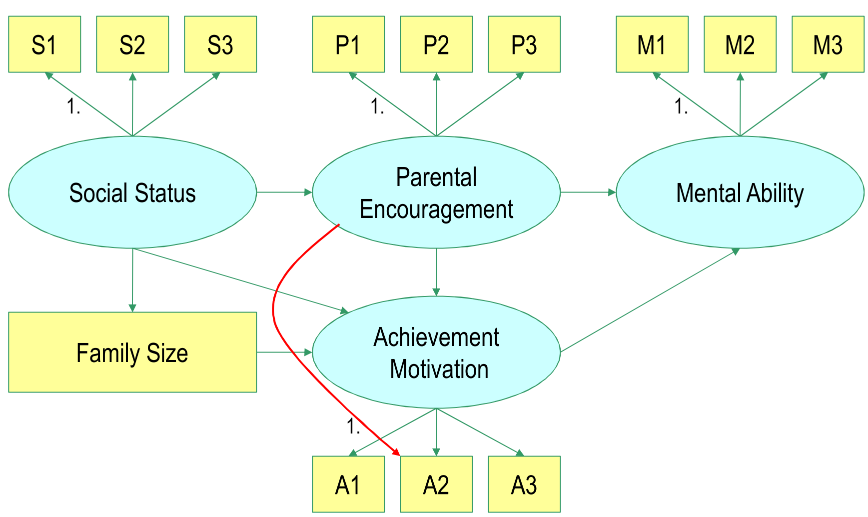
Rank Order of the 10 Largest LM Stat for Set path

Type	Var1	Var2	LM Stat	Pr > ChiSq	Parm Change
DV_DV	A2	ParentalEncouragement	18.57947	<.0001	-2.31463
DV_DV	A2	MentalAbility	17.20340	<.0001	0.95581
DV_DV	A1	ParentalEncouragement	15.86099	<.0001	1.88904
DV_DV	A1	MentalAbility	13.88705	0.0002	-0.75314
DV_DV	S2	MentalAbility	8.97262	0.0027	-0.38910
DV_DV	S3	AchievementMotivation	6.27192	0.0123	0.32436
DV_DV	S2	AchievementMotivation	6.19233	0.0128	-0.40928
DV_DV	ParentalEncouragement	MentalAbility	5.57439	0.0182	0.29376
DV_DV	ParentalEncouragement	AchievementMotivation	5.33788	0.0209	0.39218
DV_DV	FamilySize	AchievementMotivation	5.33730	0.0209	-1.20671

The measurement path A2 <=== ParentalEncouragement reduces the model fit chi-square by 19.

149

Adding a Path from Parental Achievement to A2



150

Adding the ParentalEncouragement ===> A2 path

```
proc calis data=mental nobs=115;
   path
      SocialStatus ===> ParentalEncouragement FamilySize AchievementMotivation,
      FamilySize   ===> AchievementMotivation,
      ParentalEncouragement  ===> AchievementMotivation MentalAbility,
      AchievementMotivation  ===> MentalAbility,
      SocialStatus           ===> S1 S2 S3     = 1. ,
      ParentalEncouragement  ===> P1 P2 P3 A2  = 1.,
      AchievementMotivation  ===> A1 A2 A3     = 1.,
      MentalAbility          ===> M1 M2 M3     = 1.;
run;
```


151

Before and After Adding the ParentalEncouragement ===> A2 Path

Before …

Fit Summary

Chi-Square	196.7455
Chi-Square DF	59
Pr > Chi-Square	<.0001
Standardized RMR (SRMR)	0.0936
Adjusted GFI (AGFI)	0.7341
RMSEA Estimate	0.1431
Bentler Comparative Fit Index	0.8087

After…

Fit Summary

Chi-Square	168.2618
Chi-Square DF	58
Pr > Chi-Square	<.0001
Standardized RMR (SRMR)	0.0896
Adjusted GFI (AGFI)	0.7675
RMSEA Estimate	0.1291
Bentler Comparative Fit Index	0.8468

The model fit improves quite a bit.

152

Full Information Maximum Likelihood (FIML) Method (SAS/STAT 9.3)

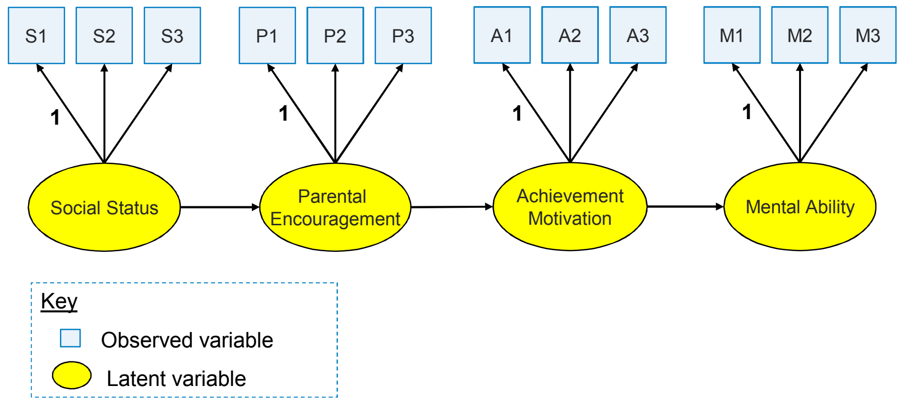
153

Why FIML?

- Utilize incomplete observations in the analysis
- Valid under the missing at random (MAR) condition
- METHOD=FIML in the PROC CALIS statement

154

Factors Affecting Mental Abilities



** This example is inspired by an example in Marjoribanks (1974). In this example, the model is purely hypothetical and the data are simulated. Except for the similarity in the names of the hypothetical constructs, no part of the current analysis represents the original research.

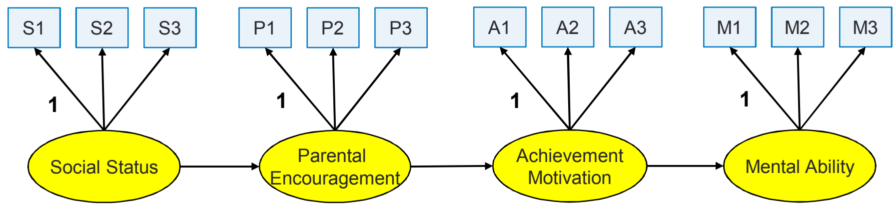
155

Data “miss3”

- Twelve observed variables
- S1, S2, S3: Observed indicators of social status
- P1, P2, P3: Observed indicators of parental encouragement
- A1, A2, A3: Observed indicators of achievement motivation
- M1, M2, M3: Observed indicators of mental ability
- 200 observations are generated.
- 100 incomplete observations with at least one missing value (but not all missing values).

156

PROC CALIS Code for the Path Diagram




```
proc calis data=miss3 method=fiml;
   path
      S1-S3   <===   SocialStatus           = 1,
      P1-P3   <===   ParentalEncouragement = 1,
      A1-A3   <===   AchievementMotivation = 1,
      M1-M3   <===   MentalAbility          = 1,
      SocialStatus           ===> ParentalEncouragement,
      ParentalEncouragement ===> AchievementMotivation,
      AchievementMotivation ===> MentalAbility;
run;
```

157

Modeling Information

Modeling Information

Data Set	WORK.MISS3
N Records Read	200
N Complete Records	100
N Incomplete Records	100
N Complete Obs	100
N Incomplete Obs	100
Model Type	PATH
Analysis	Means and Covariances

158

Data Coverage

Proportions of Data Present for Means (Diagonal) and Covariances (Off-Diagonal)

	S1	S2	S3	M1	M2	M3	A1	A2	A3	P1	P2	P3
S1	0.9250											
S2	0.8900	0.9250										
S3	0.8800	0.8900	0.9100									
M1	0.8750	0.8800	0.8800	0.8900								
M2	0.8950	0.8950	0.8800	0.8750	0.9150							
M3	0.9000	0.8950	0.8850	0.8850	0.8900	0.9200						
A1	0.8900	0.8850	0.8850	0.8800	0.8900	0.8850	0.9200					
A2	0.8950	0.8800	0.8750	0.8750	0.8850	0.8850	0.8850	0.9000				
A3	0.8900	0.8900	0.8800	0.8800	0.8800	0.8800	0.8850	0.8800	0.9100			
P1	0.5150	0.5100	0.5000	0.5050	0.5100	0.5100	0.5150	0.5100	0.5150	0.5350		
P2	0.8900	0.9100	0.8900	0.8800	0.9050	0.8850	0.9000	0.8900	0.8950	0.5150	0.9350	
P3	0.8850	0.8850	0.8900	0.8800	0.8800	0.8900	0.8900	0.8850	0.8950	0.5100	0.8900	0.9200

Problematic coverages

Average Proportion Coverage of Means	0.883750
Average Proportion Coverage of Covariances	0.824015

159

Rank Order of Mean Coverages

Rank Order of the 6 Smallest Variable (Mean) Coverages

Variable	Coverage
P1	0.5350
M1	0.8900
A2	0.9000
S3	0.9100
A3	0.9100
M2	0.9150

160

Rank Order of Covariance Coverages

Rank Order of the 10 Smallest Covariance Coverages

Var1	Var2	Coverage
P1	S3	0.5000
P1	M1	0.5050
P1	S2	0.5100
P1	M2	0.5100
P1	M3	0.5100
P1	A2	0.5100
P3	P1	0.5100
P1	S1	0.5150
P1	A1	0.5150
P1	A3	0.5150

161

Why Does P1 Have a Lot of Missing Values?

- P1: "My parents set consistent goals for me to achieve."
- The data coverage and the missing pattern analyses are useful for locating problematic items (variables).

162

Dominant Missing Patterns

Rank Order of the 5 Most Frequent Missing Patterns

Total Number of Distinct Patterns with Missing Values = 26

	Pattern	NVar Miss	Freq	Proportion	Cumulative
1	xxxxxxxxx.xx	1	75	0.3750	0.3750
2	x...xx..x..x	7	1	0.0050	0.3800
3	.x.x....xxx.	7	1	0.0050	0.3850
4	...x.xx.....	9	1	0.0050	0.3900
5	..xx.x.....x	8	1	0.0050	0.3950

NOTE: Nonmissing Pattern Proportion = 0.5000 (N=100)

163

Mean Profiles of the Dominant Missing Patterns

Means of the Nonmissing and the Most Frequent Missing Patterns

		Missing Pattern				
	Nonmissing	1	2	3	4	5
Variable	(N=100)	(N=75)	(N=1)	(N=1)	(N=1)	(N=1)
S1	4.04000	3.58667	4.00000	.	.	.
S2	3.91000	3.65333	.	3.00000	.	.
S3	4.00000	3.52000	.	.	.	7.00000
M1	4.02000	3.56000	.	1.00000	2.00000	6.00000
M2	4.04000	3.56000	6.00000	.	.	.
M3	4.01000	3.42667	3.00000	.	4.00000	6.00000
A1	4.18000	3.66667	.	.	4.00000	.
A2	4.29000	3.50667	.	.	.	.
A3	4.30000	3.46667	4.00000	2.00000	.	.
P1	4.08000	.	.	3.00000	.	.
P2	4.15000	3.73333	.	3.00000	.	.
P3	4.06000	3.70667	6.00000	.	.	6.00000

164

Model Fit Summary of FIML

Chi-Square	58.6765
Chi-Square DF	51
Pr > Chi-Square	0.2147
Standardized RMR (SRMR)	0.0403
RMSEA Estimate	0.0274
Bentler Comparative Fit Index	0.9958

165

Path Effect Estimates

Path			Parameter	Estimate	Standard Error	t Value
S1	<===	SocialStatus		1.00000		
S2	<===	SocialStatus	_Parm01	0.97633	0.05143	18.98353
S3	<===	SocialStatus	_Parm02	0.93340	0.05879	15.87763
P1	<===	ParentalEncouragement		1.00000		
P2	<===	ParentalEncouragement	_Parm03	1.02929	0.08125	12.66804
P3	<===	ParentalEncouragement	_Parm04	1.01757	0.08056	12.63135
A1	<===	AchievementMotivation		1.00000		
A2	<===	AchievementMotivation	_Parm05	1.06612	0.07252	14.70005
A3	<===	AchievementMotivation	_Parm06	1.04898	0.06795	15.43673
M1	<===	MentalAbility		1.00000		
M2	<===	MentalAbility	_Parm07	1.02426	0.06116	16.74674
M3	<===	MentalAbility	_Parm08	1.04717	0.06530	16.03538
SocialStatus	===>	ParentalEncouragement	_Parm09	0.70886	0.06736	10.52383
ParentalEncouragement	===>	AchievementMotivation	_Parm10	0.77686	0.07893	9.84259
AchievementMotivation	===>	MentalAbility	_Parm11	0.86009	0.07966	10.79700

All effect estimates are statistically significant.

166

Regular ML Estimation – Complete Case Analysis

```
proc calis data=miss3 /* method=fiml */;
   path
      S1-S3    <===    SocialStatus           = 1,
      P1-P3    <===    ParentalEncouragement = 1,
      A1-A3    <===    AchievementMotivation = 1,
      M1-M3    <===    MentalAbility          = 1,
      SocialStatus           ===> ParentalEncouragement,
      ParentalEncouragement ===> AchievementMotivation,
      AchievementMotivation ===> MentalAbility;
   run;
```

Only the 100 complete observations are used in the default ML estimation.

167

Comparing ML and FIML Fit

	Regular ML	FIML
Chi-Square	70.0624	58.6765
Chi-Square DF	51	51
Pr > Chi-Square	0.0394	0.2147
Standardized RMR (SRMR)	0.0647	0.0403
RMSEA Estimate	0.0614	0.0274
Bentler Comparative Fit Index	0.9831	0.9958

The complete-case analysis by regular ML estimation does not support a good model fit according to the chi-square test, SRMR, and RMSEA.

168

Case-Level Residual Diagnostics (SAS/STAT 12.1)

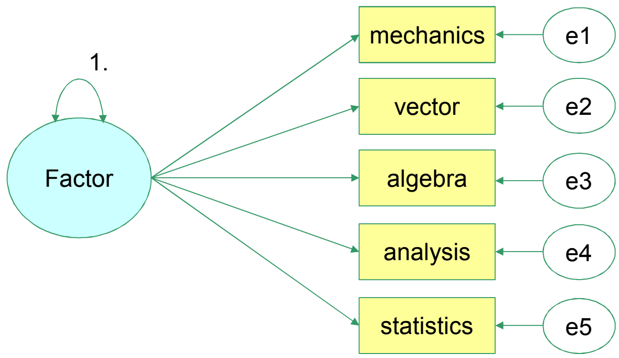
169

Case-Level Residual Diagnostics in Structural Equation Modeling

- Traditionally, residuals in SEM are those for covariances and means
- Difficulties of case-level residual diagnostics in SEM
  1. Multivariate responses
  2. Latent variable values not observed
- Yuan and Hayashi (2010)
  1. Residuals of multivariate responses --- Mahalanobis distance (M-distance) of multivariate residuals
  2. Estimation of latent factors --- generalization of Bartlett's formula for factor scores to structural equation modeling

170

A One-Factor Model for Mardia Data (N=88)



171

Residual Analysis

```
data mardia;
   input mechanics vector algebra analysis statistics;
   datalines;
77.000     82.000     67.000     67.000     81.000
63.000     78.000     80.000     70.000     81.000
75.000     73.000     71.000     66.000     81.000
 .          .          .          .          .
 .          .          .          .          .
    { More Data }
 .          .          .          .          .
 .          .          .          .          .
 .          .          .          .          .
 .          .          .          .          .
;
ods graphics on;
proc calis data=mardia residual plots=all;
   path   Factor ===> mechanics vectors algebra analysis statistics;
   pvar   Factor = 1;
run;
ods graphics off;
```

172

Covariance Residuals in SEM Output

Raw Residual Matrix

	mechanics	vectors	algebra	analysis	statistics
mechanics	-0.00001	35.32758	-0.50719	-13.81981	-13.38132
vectors	35.32758	-0.00000	-0.35983	-5.92758	-10.54651
algebra	-0.50719	-0.35983	0.00000	0.35575	0.16143
analysis	-13.81981	-5.92758	0.35575	0.00000	12.35947
statistics	-13.38132	-10.54651	0.16143	12.35947	0.00000

Average Absolute Residual	6.183100
Average Off-diagonal Absolute Residual	9.274649

These are "traditional" variance and covariance residuals in structural equation modeling.

173

Case-Level Diagnostics

- Which *observations* are outliers?
  - Observations that have large residuals in response variables mechanics, vector, algebra, analysis, and statistics --- large M-distances in residuals e1, e2, e3, e4, and e5.
  - If individual residual $e_i = \{e_{1i}, e_{2i}, e_{3i}, e_{4i}, e_{5i}\}$ and Cov(e) are known, residual M-distance is:

$$d_{ri} = \sqrt{e_i \text{Cov}^{-1}(e) e_i'}$$

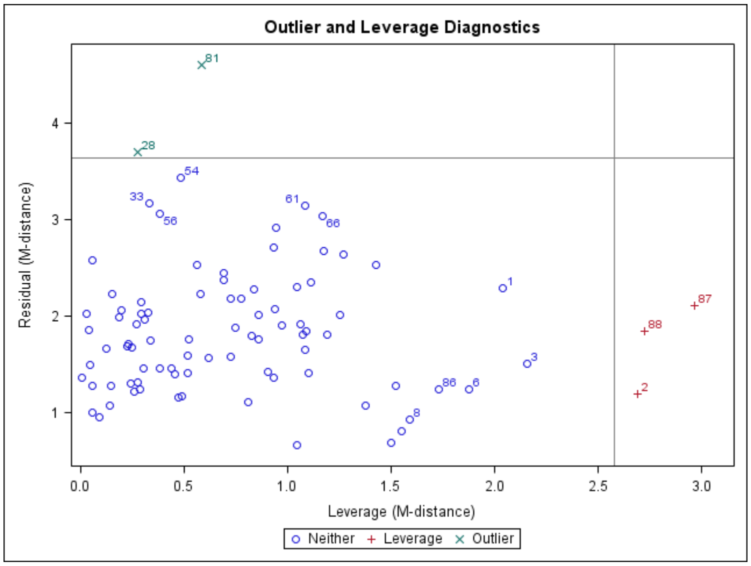
- Which *observations* are leverage points?
  - Observations that have large Factor scores --- large M-distances in factor scores.
  - If individual factor score f_i and var(f) are known, leverage M-distance is:

$$d_{fi} = \sqrt{f_i \text{var}^{-1}(f) f_i'}$$

In practice, d_{ri} and d_{fi} are estimated from the data given the model estimates.

174

Outlier and Leverage Point Detection



Outlier Detection: Numerical Output

Diagnostics of the 7 Largest Case-Level Residuals (alpha=0.01)

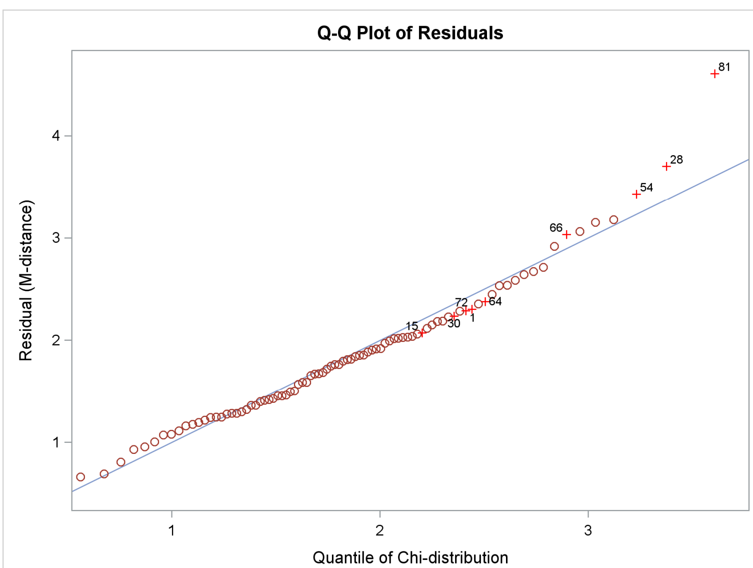
Case Number	Residual (M-Distance)	----Diagnostics----- Outlier	Leverage
81	4.60517	*	
28	3.69951	*	
54	3.43066		
33	3.17665		
61	3.15110		
56	3.06087		
66	3.03273		

Leverage Points: Numerical Output

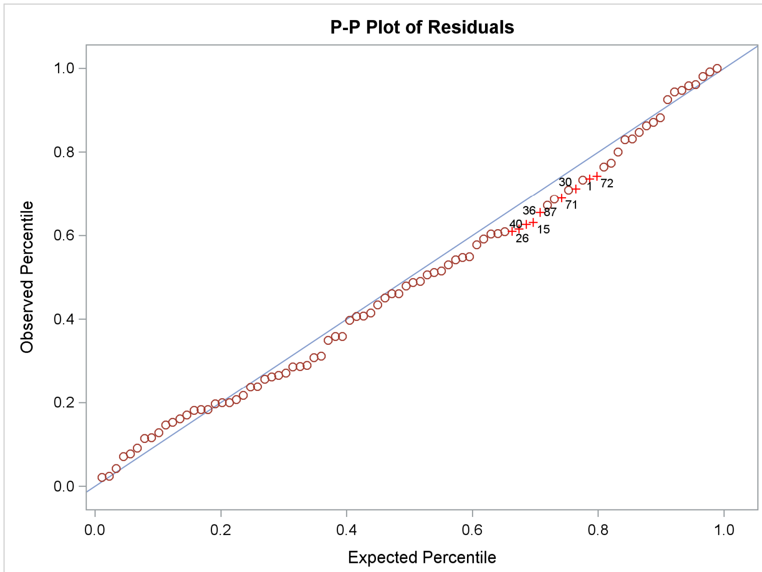
Diagnostics of the 8 Largest Case-Level Leverages (alpha=0.01)

Case Number	Leverage (M-Distance)	----Diagnostics----- Leverage	Outlier
87	2.96439	*	
88	2.72440	*	
2	2.69293	*	
3	2.15520		
1	2.04016		
6	1.87893		
86	1.72951		
8	1.59244		

Q-Q Plot of Residual M-Distances



P-P Plot of Residual M-Distances



179

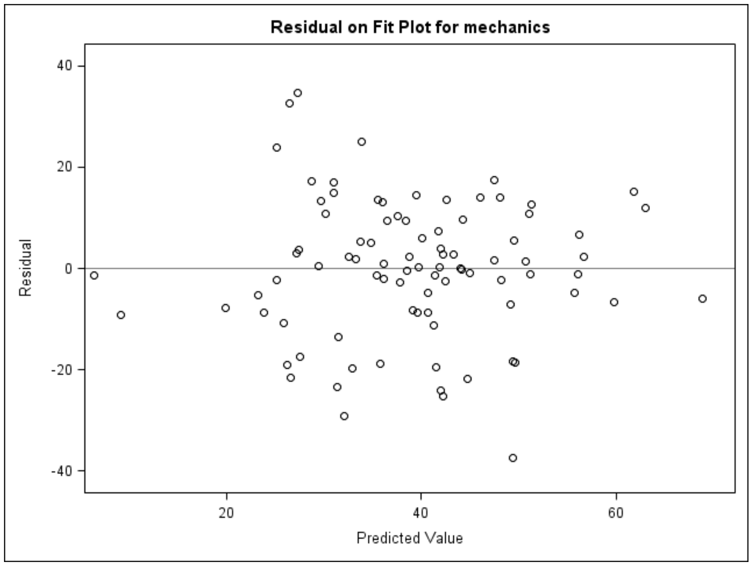
Departures From the Theoretical Distribution

Largest Departures From the Theoretical Residual Distribution

Percentile Region	Case Number	Quantile Deviation	Percentile Deviation
(.65,.90)	26		-0.05303
	36		-0.05835
	40		-0.05864
	15	-0.13083	-0.06507
	87		-0.05262
	71		-0.05145
	30	-0.11975	-0.05199
	1	-0.12349	-0.05067
	72	-0.13743	-0.05500
	64	-0.12554	
(.90,.95)	66	0.13814	
>.95	54	0.20037	
	28	0.32396	
	81	0.99849	

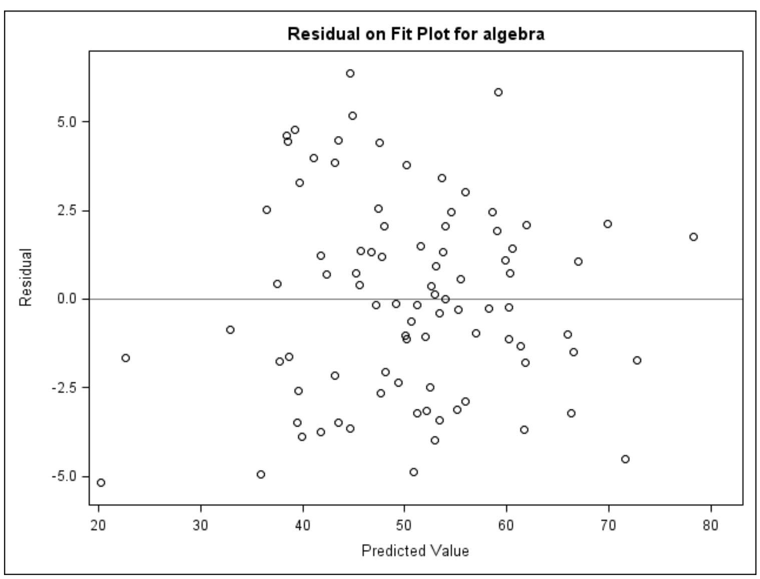
180

Residual on Fit Plot - Mechanics



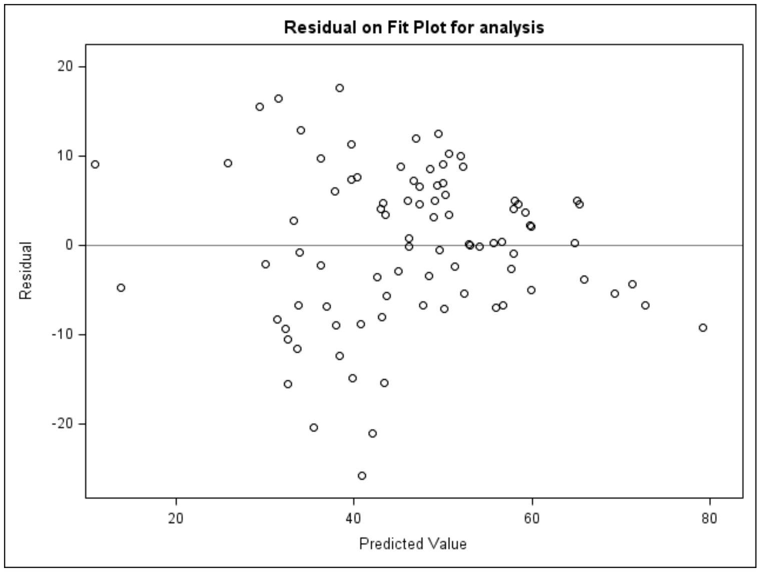
181

Residual on Fit Plot - Algebra



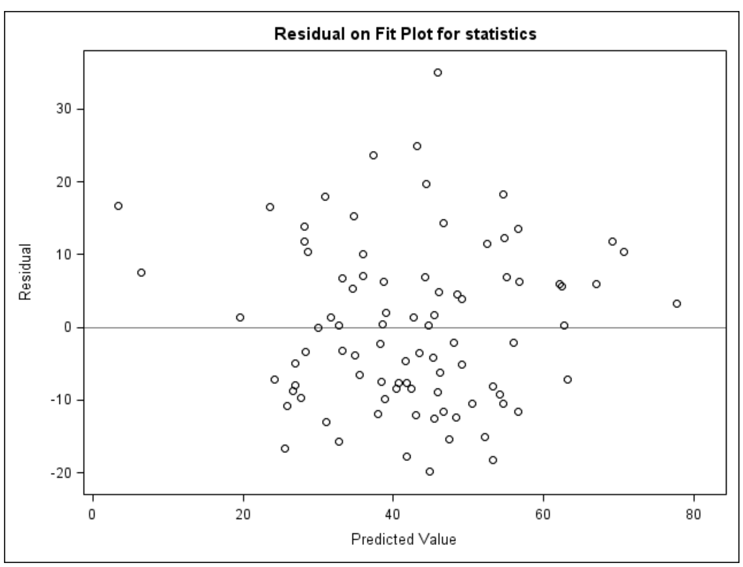
182

Residual on Fit Plot - Analysis



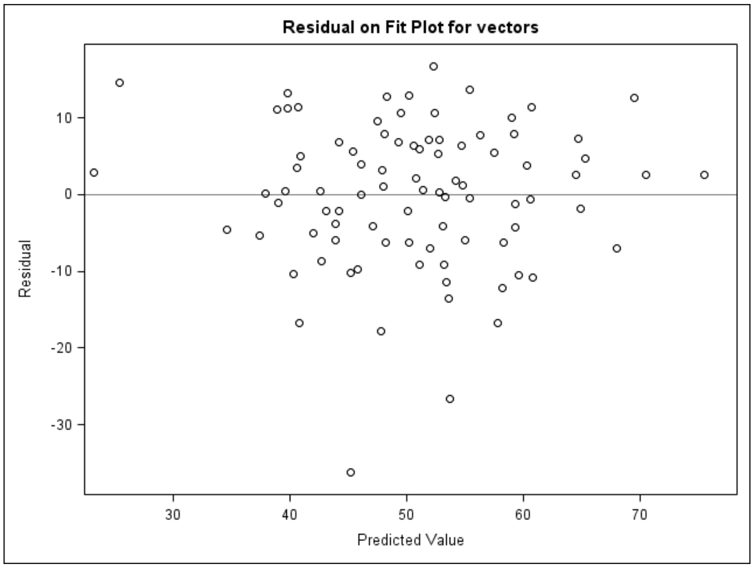
183

Residual on Fit Plot - Statistics



184

Residual on Fit Plot - Vector



185

Robust Estimation (SAS/STAT 12.1)

186

The ROBUST Option

```
ods graphics on;
proc calis data=mardia robust residual plots=all;
   path   Factor ===> mechanics vectors algebra analysis statistics;
   pvar   Factor = 1;
run;
ods graphics off;
```

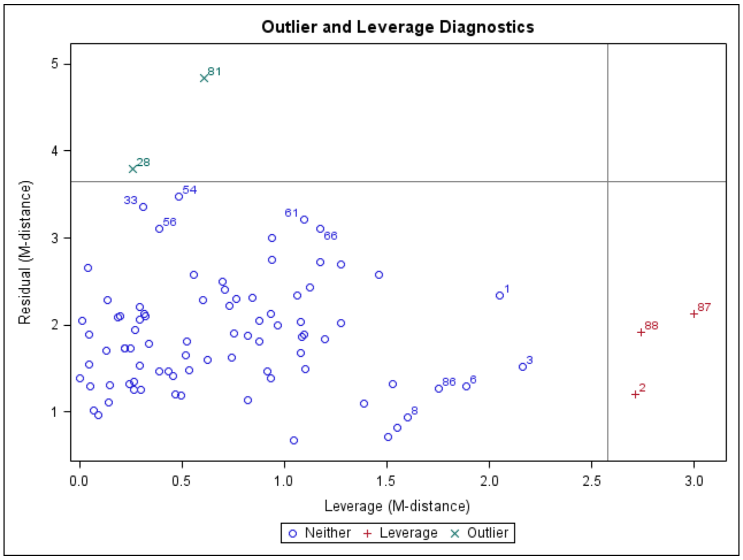
187

Robust Estimation Technique in PROC CALIS

- Yuan and Hayashi (2010)
- Iteratively-reweighted Least Squares (IRLS)
- Observations are re-weighted during the estimation
- Huber-type weights:
  - weight = 1 for "normal" observations
  - weight < 1 for outlying observations with large residuals

188

Outlier and Leverage Point Detection With Robust Estimation



189

Masking Effects and Unmasking

Robust Estimation

Case Number	Residual (M-Distance)	Outlier
81	4.84496	*
28	3.78838	*
54	3.47349	
33	3.35818	
61	3.20898	
56	3.10657	
66	3.10424	

Regular ML Estimation

Case Number	Residual (M-Distance)	Outlier
81	4.60517	*
28	3.69951	*
54	3.43066	
33	3.17665	
61	3.15110	
56	3.06087	
66	3.03273	

Masking effects is minimal with the regular ML estimation.

190

Model Fit: Robust and Regular ML Estimations

Robust Estimation

Chi-Square	6.0031
Chi-Square DF	5
Pr > Chi-Square	0.3059
Standardized RMR (SRMR)	0.0370
Goodness of Fit Index (GFI)	0.9716
RMSEA Estimate	0.0480
Bentler Comparative Fit Index	0.9949

Regular ML Estimation

Chi-Square	8.9782
Chi-Square DF	5
Pr > Chi-Square	0.1099
Standardized RMR (SRMR)	0.0475
Goodness of Fit Index (GFI)	0.9584
RMSEA Estimate	0.0956
Bentler Comparative Fit Index	0.9791

With the outliers being downweighted in estimation, the robust estimation indicates a better model fit .

191

More About PROC CALIS ....

- Many other different modeling languages: COSAN, FACTOR, LINEQS, MSTRUCT, and RAM – All support multiple-group analysis and mean structures
- Estimation methods: ML (default), FIML, GLS, WLS (ADF), ULS, DWLS, ROBUST (robust ML)
- Standardized solutions with standard error estimates

192

Glossary

Manifest – Observed variables (measured variables) in the data set.
Latent – Unobserved variables.
Endogenous – Dependent /mediating variables; at least one single-headed arrow points to it; used as an outcome variable in an equation; can also be a predictor variable in other equations.
Exogenous – Independent variables; no single-headed arrows point to it; never used as an outcome variable in the model; used only as a predictor in the model.
Factor – A latent (unmeasured) variable that is treated as a hypothetical construct (systematic source) in the model.
Error – An exogenous term for uncertainty (unsystematic source) associated with an endogenous *manifest* variable (or any endogenous variable, in a more general definition).
Disturbance – An exogenous term for uncertainty (unsystematic source) associated with an endogenous *latent* variable.
Path diagram representation
– Rectangles: Observed / manifest variables.
– Ovals / circles : Latent variables (factors, errors, and disturbances). Errors and disturbances are not necessarily put into ovals/circles.

193

Glossary

– Single-headed arrows: Directed paths, direct effects, path coefficients; specified in the PATH statement.
– Double-headed arrows that point to individual variables: Variance parameters of exogenous variables or error variance parameters of endogenous variables; specified in the PVAR statement.
– Double-headed arrows that point to two distinct variables: Covariance parameters between exogenous variables or error covariance parameters between endogenous variables; specified in the PCOV statement.
Fit assessment
– model fit chi-square statistic: Nonsignificance means that the theoretical model is supported; not a very practical index because it almost always rejects all approximating models that are practically useful.
– AGFI (adjusted goodness-of-fit index) and Bentler's CFI (comparative fit index): Two popular fit indices that indicate good model fit when their values are above 0.9.
– SRMR (standardized root mean square residual) and RMSEA (root mean squared error approximation): Two popular fit indices that indicate good model fit when their values are below 0.05.
– AIC, CAIC, and SBC: Information criteria for comparing competing models. The smaller the better.

194

References

Arbuckle, J. (2008). *AMOS 17.0 User's Guide*. Crawfordville, FL: Amos Development Corporation.

Bentler, P. M. (1985). *Theory and Implementation of EQS: A Structural Equations Program, Manual for Program Version 2.0*. Los Angeles: BMDP Statistical Software.

Bentler, P. M. (1995). *EQS, Structural Equations Program Manual*, *Program Version 5.0*. Encino, CA: Multivariate Software.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

Fuller, W. A. (1987). *Measurement Error Models*. New York: John Wiley and Sons.

Holzinger, K. J., & Swineford, F. A. (1939). A Study in Factor Analysis: The Stability of a Bi-Factor Solution. *Supplementary Educational Monographs, No. 48*. Chicago: University of Chicago, Dept. of Education.

Jöreskog, K. G. (1973). A General Method for Estimating a Linear Structural Equation System. In A. S. Goldberger and O. D. Duncan (eds.): *Structural Equation Models in the Social Sciences*. New York: Academic Press.

Keesling, J. W. (1972). *Maximum Likelihood Approaches to Causal Analysis*. Ph.D. thesis, University of Chicago.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

Marjoribanks, K., ed. (1974). *Environments for Learning*. London: National Foundation for Educational Research Publications.

McDonald, R. P. (1978). A Simple Comprehensive Model for the Analysis of Covariance Structures. *British Journal of Mathematical and Statistical Psychology*, 31, 59–72.

195

References

McDonald, R. P. (1980). A Simple Comprehensive Model for the Analysis of Covariance Structures: Some Remarks on Applications. *British Journal of Mathematical and Statistical Psychology*, 33, 161–183.

Nowak, T. P., Hoffman, D. L., & Yung, Y. F. (2000). Measuring the Customer Experience in Online Environments: A Structural Modeling Approach. *Marketing Science*, *19(1)*, 22–42.

Wiley, D. E. (1973). The Identification Problem for Structural Equation Models with Unmeasured Variables. In A. S. Goldberger and O. D. Duncan (eds.): *Structural Equation Models in the Social Sciences*. New York: Academic Press.

Yung, Y.-F. (2014). Creating Path Diagrams That Impress: A New Graphical Capability of the CALIS Procedure. http://support.sas.com/rnd/app/stat/papers/2014/yungpd2014.pdf

Yung, Y.-F. and Zhang, W. (2011). Making Use of Incomplete Observations in the Analysis of Structural Equation Models: The CALIS Procedure's Full Information Maximum Likelihood Method in SAS/STAT 9.3. In *Proceedings of the SAS Global Forum 2011 Conference*, Cary, NC: SAS Institute Inc. http://support.sas.com/resources/papers/proceedings11/333-2011.pdf

Yuan, K.-H., & Hayashi, K. (2010). Fitting Data to Model: Structural Equation Modeling Diagnosis Using Two Scatter Plots. *Psychological Methods*, 15, 335–351.

Zhang, W. and Yung, Y.-F. (2011). A Tutorial on Structural Equation Modeling with Incomplete Observations: Multiple Imputation and FIML Methods Using SAS. http://support.sas.com/rnd/app/stat/papers/imps2011_FIML.pdf

196