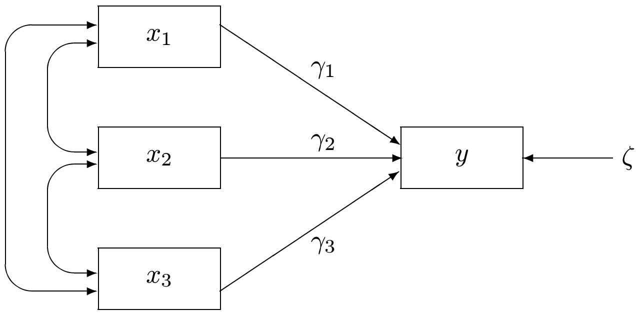
Slide 1

Slides for Examples

Slide 2

Regression-Type Models



A Single Regression Equation

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 3

Example: Regression of GNP

Goldberger (1964, p. 187) presented raw data on gross national product in billions of dollars (y = GNP), labor inputs in millions of man-years (x_1 = LABOR), real capital in billions of dollars (x_2 = CAPITAL), and the time in years measured from 1928 (x_3 = TIME). A path diagram for the regression of y on x_1, x_2, and x_3 is shown in Figure 1 (Slide 4). The data consists of 23 annual observations for the United States during 1929–1941 and 1946–1955. The covariance matrix of the variables is given in Table 1 (Slide 6).

Slide 4

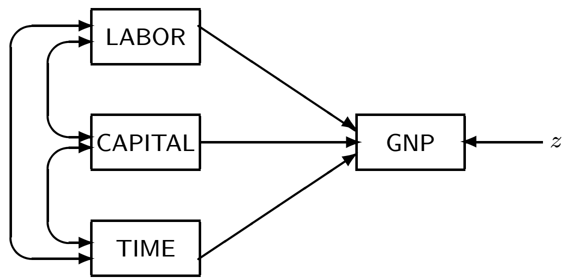


Figure 1: Path Diagram for Regression of GNP

Slide 5

As the path diagram indicates, LABOR, CAPITAL, and TIME are supposed to influence GNP. This is indicated by the three one-way (unidirected) arrows pointing towards GNP. The three two-way arrows on the left indicate that the three independent variables may be correlated. The one-way arrow on the right represents the effect of the error term z.

Slide 6

Table 1: Covariance Matrix for GNP Data

	y	x_1	x_2	x_3
GNP	4256.530			
LABOR	449.016	52.984		
CAPITAL	1535.097	139.449	1114.447	
TIME	537.482	53.291	170.024	73.747
Means	180.435	45.565	50.087	13.739

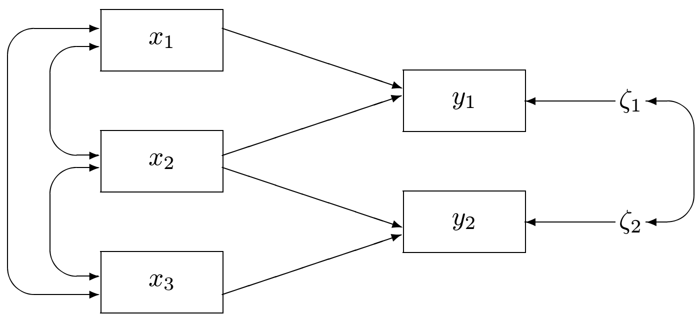
Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 7

The SIMPLIS input file (`EX1A.SPL`):

```
Regression of GNP
Observed variables: GNP LABOR CAPITAL TIME
Covariance Matrix
 4256.530
  449.016     52.984
 1535.097    139.449   1114.447
  537.482     53.291    170.024     73.747
Sample Size: 23
Equation: GNP = LABOR CAPITAL TIME
End of Problem
```

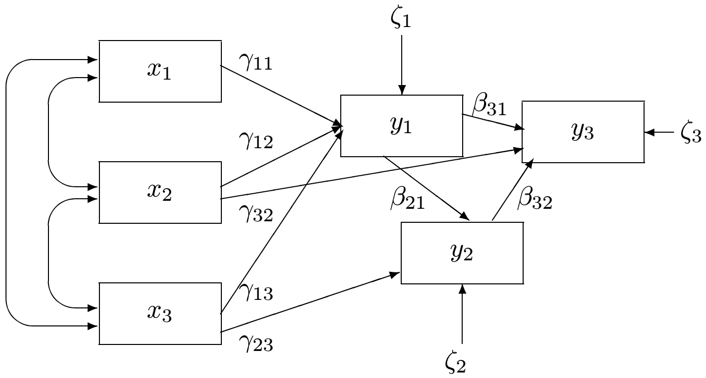
Slide 8



Bivariate Regression

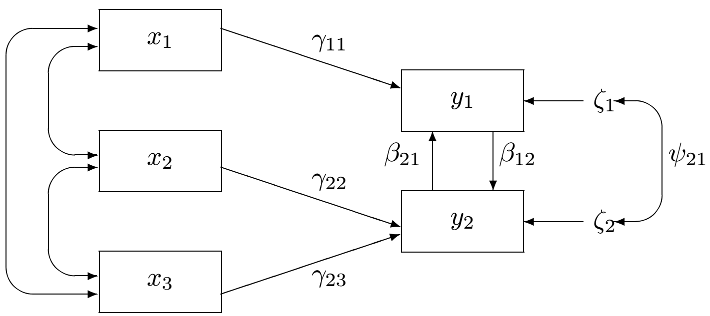
Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 9



Recursive System

Slide 10



Non-recursive System

Slide 11

Example: Union Sentiment of Textile Workers

McDonald & Clelland (1984) analyzed data on union sentiment of southern nonunion textile workers. After transformation of one variable and treatment of outliers, Bollen (1989) reanalyzed a subset of the variables according to the model shown in Figure 3 (Slide 12). The variables are:

$y_1 =$ *deference (submissiveness) to managers*

$y_2 =$ *support for labor activism*

$y_3 =$ *sentiment towards unions*

$x_1 =$ *logarithm of years in textile mill*

$x_2 =$ *age*

Slide 12

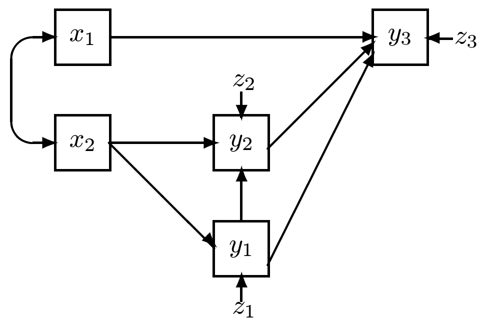


Figure 2: Path Diagram for Union Sentiment Model

Slide 13

Two ways of specifying the model

```
Relationships
Y1 = X2
Y2 = X2 Y1
Y3 = X1 Y1 Y2

Paths
X1 -> Y3
X2 -> Y1 Y2
Y1 -> Y2 Y3
Y2 -> Y3
```

Slide 14

Censored Variable

A censored variable has a large fraction of observations at the minimum or maximum. Ordinary estimates of the mean and variance of a censored variable are biased. Ordinary least squares (OLS) estimates of its regression on a set of explanatory variables are also biased.

A censored normal variable can be defined as follows. Let $y^\star$ be normally distributed with mean μ and variance σ^2. An observed variable y is censored below if

$$\begin{aligned} y &= c \text{ if } y^\star \leq c \\ &= y^\star \text{ otherwise ,} \end{aligned}$$

where c is a given constant. This is illustrated in Figures 15 and 16.

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 15

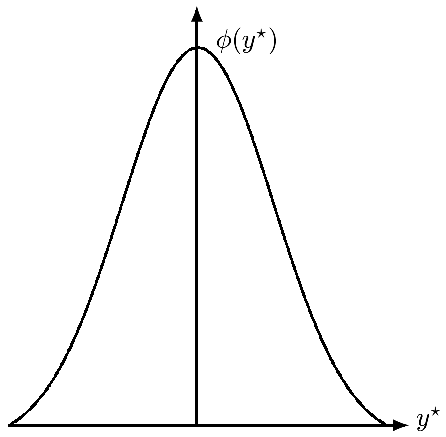


Figure 3: Normal Variable $y^\star$

Slide 16

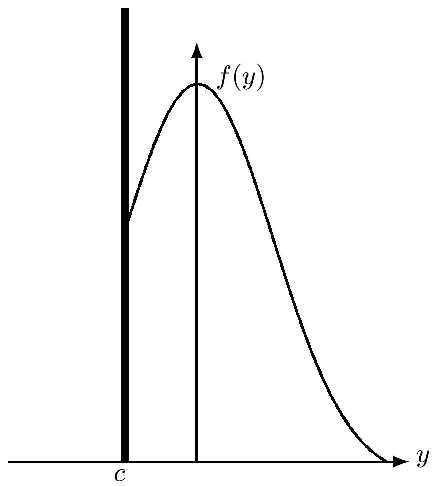


Figure 4: Censored Variable y

Slide 17

Censored Normal Regression

The regression equation to be estimated is

$$y^\star = \alpha + \boldsymbol{\gamma}'\mathbf{x} + z \, , \tag{1}$$

where α is the intercept term, $\boldsymbol{\gamma}$ is the vector of regression coefficients, and $\mathbf{x}$ the regressors. The error term z is assumed to be normally distributed with mean 0 and variance ψ^2.

For a variable that is censored both below and above, the observed variable is

$$\begin{aligned} y &= c_1 \text{ if } y^\star \le c_1 \\ &= y^\star \text{ if } c_1 < y^\star < c_2 \\ &= c_2 \text{ if } y^\star \ge c_2 \, , \end{aligned}$$

where c_1 and c_2 are constants.

Slide 18

Let $(y_i, \mathbf{x}_i)$ be the observed values of y and $\mathbf{x}$ of case i in a random sample of N independent observations. The likelihood of $(y_i, \mathbf{x}_i)$ is

$$L_i = \left[\Phi(\frac{c_1 - \alpha - \boldsymbol{\gamma}'\mathbf{x}_i}{\psi})\right]^{j_{1i}} \tag{2}$$

$$\times \left[\frac{1}{\sqrt{2\pi}\psi} e^{-\frac{1}{2}(\frac{y_i - \alpha - \boldsymbol{\gamma}'\mathbf{x}_i}{\psi})^2}\right]^{1-j_{1i}-j_{2i}} \tag{3}$$

$$\times \left[1 - \Phi(\frac{c_2 - \alpha - \boldsymbol{\gamma}'\mathbf{x}_i}{\psi})\right]^{j_{2i}} \, , \tag{4}$$

where $j_{i1} = 1$ if $y = c_1$ and $j_{i1} = 0$ otherwise and $j_{i2} = 1$ if $y = c_2$ and $j_{i2} = 0$ otherwise. Note that j_{i1} and j_{i2} cannot be 1 simultaneously.

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 19

Example: Affairs

Fair (1978) published an example of censored regression. His study concerns the number of extramarital affairs and its determinants. From a large data set, the results of which were published in the July 1970 issue of *Psychology Today*, Fair extracted 601 observations on men and women who were then currently married for the first time. His data set consisting of 15 variables is available on the Internet at *http://fairmodel.econ.yale.edu/rayfair/workss.htm*. For present purposes the following nine variables are used.

Slide 20

GENDER 0 = female, 1 = male

AGE in years

YEARS number of years married

CHILDREN 0 = no, 1 = yes

RELIGIOUS 1 = anti, ..., 5 = very religious

EDUCATION number of years of schooling, 9 = grade school, 12 = high school, 20 = PhD

OCCUPATION Hollingshead scale of 1 to 7

HAPPINESS self rating of quality of marriage, 1 = very unhappy, ..., 5 = very happy

AFFAIRS number of affairs in the past year, 1, 2, 3, 4–10 coded as 7, monthly, weekly, and daily coded as 12

Slide 21

Exploratory Factor Analysis

Exploratory factor analysis is a technique often used to detect and assess latent sources of variation and covariation in observed measurements. It is widely recognized that exploratory factor analysis can be quite useful in the early stages of experimentation or test development. Thurstone's (1938) primary mental abilities, French's (1951) factors in aptitude and achievement tests and Guilford's (1956) structure of intelligence are good examples of this. The results of an exploratory factor analysis may have heuristic and suggestive value and may generate hypotheses which are capable of more objective testing by other multivariate methods.

Slide 22

As more knowledge is gained about the nature of social and psychological measurements, however, exploratory factor analysis may not be a useful tool and may even become a hindrance.

Most studies are to some extent both exploratory and confirmatory since they involve some variables of known and other variables of unknown composition. The former should be chosen with great care in order that as much information as possible about the latter may be extracted. It is highly desirable that a hypothesis which has been suggested by mainly exploratory procedures should subsequently be confirmed, or disproved, by obtaining new data and subjecting these to more rigorous statistical techniques.

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 23

The basic idea of factor analysis is the following. For a given set of response variables $x_1, \ldots, x_p$ one wants to find a set of underlying latent factors $\xi_1, \ldots, \xi_k$, fewer in number than the observed variables. These latent factors are supposed to account for the intercorrelations of the response variables in the sense that when the factors are partialed out from the observed variables, there should no longer remain any correlations between these. If both the observed response variables and the latent facors are measured in deviations from the mean, this leads to the model:

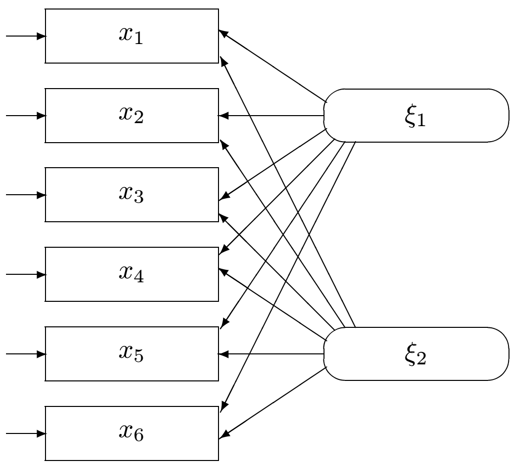
$$x_i = \lambda_{i1}\xi_1 + \lambda_{i2}\xi_2 + \cdots + \lambda_{in}\xi_k + \delta_i \; , \tag{5}$$

where δ_i, the unique part of x_i, is assumed to be uncorrelated with $\xi_1, \xi_2, \ldots, \xi_k$ and with δ_j for $j \neq i$.

Slide 24

The unique part δ_i consists of two components: a specific factor s_i and a pure random measurement error e_i. These are indistinguishable, unless the measurements x_i are designed in such a way that they can be separately identified (panel designs and multitrait-multimethod designs). The term δ_i is often called the *measurement error* in x_i even though it is widely recognized that this term may also contain a specific factor as stated above.

Slide 25



Exploratory Factor Analysis

Slide 26

Example: Exploratory Factor Analysis

Suppose we compute $\mathbf{\Sigma}$ by the formula

$$\mathbf{\Sigma} = \mathbf{\Lambda}_x \mathbf{\Phi} \mathbf{\Lambda}_x' + \mathbf{\Theta}_\delta \ ,$$

where

$$\mathbf{\Lambda}_x = \begin{pmatrix} 0.9 & 0 \\ 0.8 & 0 \\ 0 & 0.7 \\ 0 & 0.6 \\ 0 & 0.5 \\ 0.3 & 0 \end{pmatrix} \quad \mathbf{\Phi} = \begin{pmatrix} 1 & \\ 0.6 & 1 \end{pmatrix} \quad \mathbf{\Theta}_\delta = \mathsf{diag} \begin{pmatrix} 0.19 \\ 0.36 \\ 0.51 \\ 0.64 \\ 0.75 \\ 0.91 \end{pmatrix}$$

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

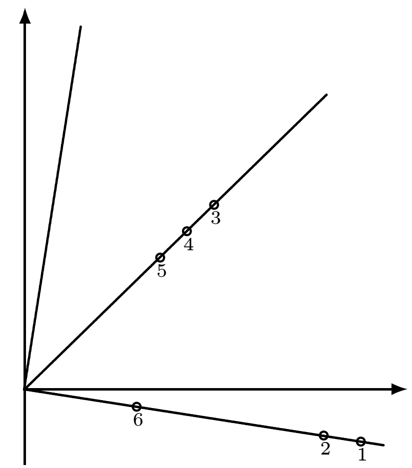
Slide 27

The resulting $\mathbf{\Sigma}$ is

$$\mathbf{\Sigma} = \begin{pmatrix} 1.000 & & & & & \\ 0.720 & 1.000 & & & & \\ 0.378 & 0.336 & 1.000 & & & \\ 0.324 & 0.288 & 0.420 & 1.000 & & \\ 0.270 & 0.240 & 0.350 & 0.300 & 1.000 & \\ 0.270 & 0.240 & 0.126 & 0.108 & 0.090 & 1.000 \end{pmatrix} .$$

Can an exploratory factor analysis of $\mathbf{\Sigma}$ recover the known factor structure? Assume that the sample size is 1000.

Slide 28



Factor Rotation

Slide 29

Example HOLZSWIN

Holzinger & Swineford (1939) collected data on twenty-six psychological tests administered to seventh- and eighth-grade children in two schools in Chicago: the Pasteur School and the Grant-White School. Nine of these tests were selected for this example. The nine tests are (with the original variable number in parenthesis):

VIS PERC Visual Perception (V1)

CUBES Cubes (V2)

LOZENGES Lozenges (V4)

PAR COMP Paragraph Comprehension (V6)

SEN COMP Sentence Completion (V7)

WORDMEAN Word meaning (V9)

Slide 30

ADDITION Addition (V10)

COUNTDOT Counting dots (V12)

S-C CAPS Straight-curved capitals (V13)

The raw data is for the Grant White School is in the file `\LIS850EX\NPV.RAW`. The data is in free format, i.e., there are blank spaces between every entry. There are no missing values.

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 31

Example Second-Order Factor Analysis

McDonald (1985) gives the correlation matrix in Table 2 for three measures of verbal ability, V1, V2, V3, three measures of word fluency, W1, W2, W3, and three measures of reasoning ability, R1, R2, R3. The original data comes from a study by Thurstone. The sample size is 213.

Consider the second-order factor analysis model in Figure 5 (Slide 33). Estimate and test this model in completely standardized form.

Slide 32

Table 2: Correlation Matrix for Nine Psychological Tests

V1	1.000								
V2	0.828	1.000							
V3	0.776	0.779	1.000						
W1	0.439	0.493	0.460	1.000					
W2	0.432	0.464	0.425	0.674	1.000				
W3	0.447	0.489	0.443	0.590	0.541	1.000			
R1	0.447	0.432	0.401	0.381	0.402	0.288	1.000		
R2	0.541	0.537	0.534	0.350	0.367	0.320	0.555	1.000	
R3	0.380	0.358	0.359	0.424	0.446	0.325	0.598	0.452	1.

Slide 33

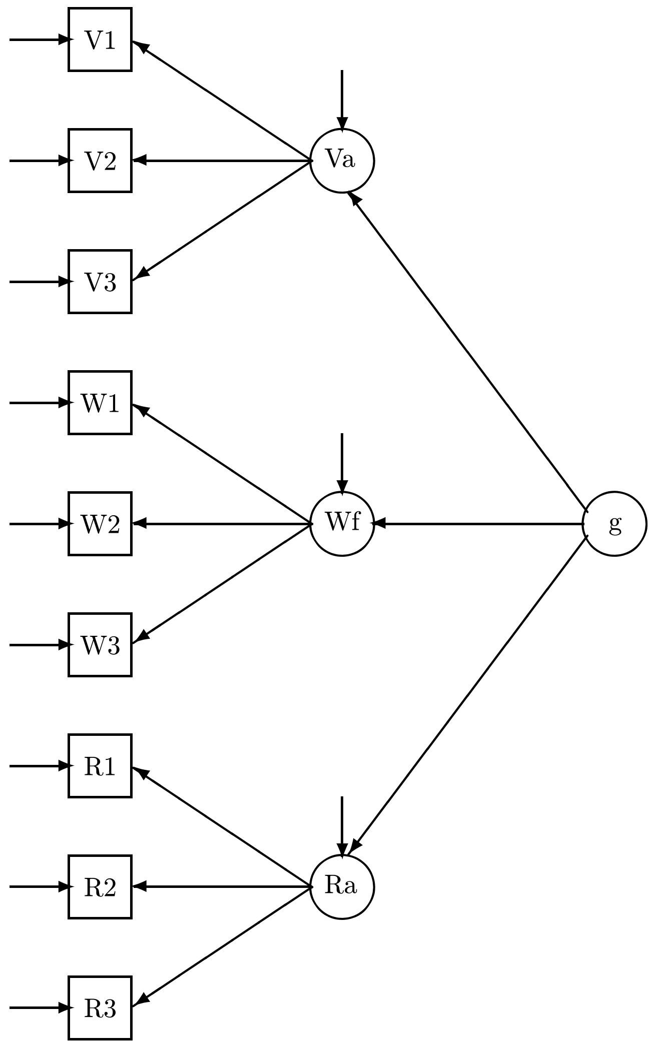


Figure 5: Second-Order Factor Analysis Model for Nine Psychological Tests

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 34

Multitrait-Multimethod Matrices

Wehrle (1982) gives data on several measures supposed to measure two theoretical constructs *Control of Pace* and *Powerlessness* in a sample of 119 white female secretaries working in the United States. Each of the two constructs was measured by three methods: (L) response to a Likert scale composed of five statements expressing progressively greater control over the pace of work or progressively lesser powerlessness at work, (N) a number method where the respondent rated herself on a scale from 1 to 10, and (R) a rating from 1 to 5 made by a trained rater based on the respondent's narrative description of her work.

Slide 35

The variables are

- x_{1L} = Control of pace of work measured by Likert method
- x_{1N} = Control of pace of work measured by Number method
- x_{1R} = Control of pace of work measured by Rating method
- x_{2L} = Powerlessness at work measured by Likert method
- x_{2N} = Powerlessness at work measured by Number method
- x_{2R} = Powerlessness at work measured by Rating method

Slide 36

Covariance Matrix of Control of Pace and Powerlessness
Measured by Three Methods

x_{1L}	x_{1N}	x_{1R}	x_{2L}	x_{2N}	x_{2R}
.976					
1.056	3.480				
.339	.699	.517			
.272	.277	.202	.903		
.756	1.797	.879	1.522	7.126	
.303	.440	.239	.508	1.577	.933

Slide 37

Problems

Analyze the data to answer the questions

1. Do the six variables measure the two constructs they are supposed to measure? If so, how highly correlated are these constructs?
2. Is there a specific method variance in the observed measures attributable to each method? If so what is the relative variance contribution to the total variance due to trait, method, and error.

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 38

Research and Data Collection Designs

- Cross-sectional
- Time-series
- Longitudinal
  - Retrospective
  - Prospective
  - Repeated measures
  - Panel
  - Rotating panel
- Experimental and quasi-experimental

Slide 39

Models for Longitudinal Data

The characteristic feature of a longitudinal research design is that the same measurement instruments are used on the same people at two or more occasions. The purpose of a longitudinal is to assess the changes that occur between the occasions, and to attribute these changes to certain background characteristics and events existing or occurring before the first occasion and/or to various treatments and developments that occur after the first occasion. Often, when the same variables are used repeatedly, there is a tendency for the measurement errors in these variables to correlate over time because of specific factors, memory or other retest effects. Hence there is a need to consider models with correlated measurement errors.

Slide 40

Example: Stability of Alienation

Wheaton, et al. (1977)report on a study concerned with the stability over time of attitudes such as alienation, and the relation to background variables such as education and occupation. Data on attitude scales were collected from 932 persons in two rural regions in Illinois at three points in time: 1966, 1967, and 1971. The variables used for the present example are the Anomia subscale and the Powerlessness subscale, taken to be indicators of Alienation. This example uses data from 1967 and 1971 only. The background variables are the respondent's education (years of schooling completed) and Duncan's Socioeconomic Index (SEI). These are taken to be indicators of the respondent's socioeconomic status (Ses). The sample covariance matrix of the six observed variables is

Slide 41

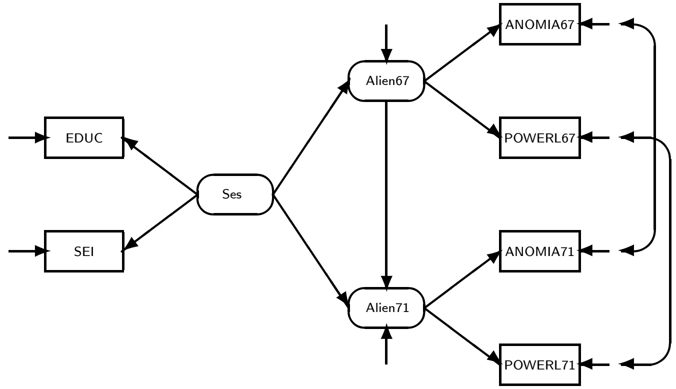
	y_1	y_2	y_3	y_4	x_1	x_2
ANOMIA67	11.834					
POWERL67	6.947	9.364				
ANOMIA71	6.819	5.091	12.532			
POWERL71	4.783	5.028	7.495	9.986		
EDUC	−3.839	−3.889	−3.841	−3.625	9.610	
SEI*	−2.190	−1.883	−2.175	−1.878	3.552	4.503

* The variable SEI has been scaled down by a factor 10.

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 42

Path Diagram for Stability of Alienation



Slide 43

Autocorrelated Measurement Error

Specific Error Variance

$$\text{Time 1}: \ y_1 = \lambda_1\eta_1 + \epsilon_1 = \lambda_1\eta_1 + s + e_1$$

$$\text{Time 2}: \ y_3 = \lambda_3\eta_2 + \epsilon_3 = \lambda_3\eta_2 + s + e_3$$

e_1 and e_3 uncorrelated

$$Cov(\epsilon_1, \epsilon_3) = Var(s)$$

Slide 44

Simplex Models

A simplex model is a type of covariance structure which often occurs in longitudinal studies when the same variable is measured repeatedly on the same people over several occasions. The simplex model is equivalent to the covariance structure generated by a first-order non-stationary autoregressive process. Guttman (1954) used the term simplex also for variables which are not ordered through time but by other criteria. One of his examples concerns tests of verbal ability ordered according to increasing complexity. The typical feature of a simplex correlation structure is that the entries in the correlation matrix decrease as one moves away from the main diagonal.

Slide 45

Jöreskog (1970) formulated various simplex models in terms of the well-known Wiener and Markov stochastic processes. A distinction was made between a perfect simplex and a quasi-simplex. A *perfect simplex* is reasonable only if the measurement errors in the variables are negligible. A *quasi-simplex*, on the other hand, allows for sizable errors of measurement.

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 46

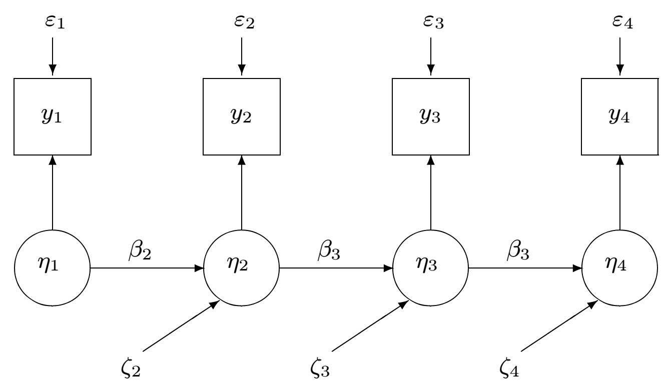
Consider p fallible variables $y_1, y_2, \ldots, y_p$. The unit of measurement in the true variables η_i may be chosen to be the same as in the observed variables y_i. The equations defining the model are then

$$y_i = \eta_i + \epsilon_i, \quad i = 1, 2, \ldots, p \ ,$$

$$\eta_i = \beta_i \eta_{i-1} + \zeta_i, \qquad i = 2, 3, \ldots, p \ ,$$

where the ϵ_i are uncorrelated among themselves and uncorrelated with all the η_i and where ζ_i is uncorrelated with η_{i-1} for $i = 2, 3, \ldots, p$). A path diagram of the simplex model with $p = 4$ is given in the following figure.

Slide 47



A Simplex Model

Slide 48

The parameters of the model are $\psi_1 = \mathsf{Var}(\eta_1)$, $\psi_i = \mathsf{Var}(\zeta_i)$ $(i = 2, 3, \ldots, p)$, $\theta_i = \mathsf{Var}(\epsilon_i)$ $(i = 1, 2, \ldots, p)$ and β_2, β_3, …, β_p. It can be shown that (see LISREL 8: User's Reference Guide, pp 232-233) β_2, $\psi_1 = \psi_1$, ψ_2, ψ_4, θ_1, and θ_4 are not identified whereas β_3, β_4, ψ_3, θ_2, and θ_3 are identified.

Slide 49

Example: Simplex Model for Academic Performance

Humphreys (1968) gives the following correlation[a] *matrix. The variables include eight semesters of grade-point averages,* $y_1, y_2, \ldots, y_8$, *high school rank* y_0 *and a composite score on the American College Testing test* y'_0 *for approximately 1600 undergraduate students at the University of Illinois.*

[a] For standard errors and chi-squares to be correct, the covariance matrix should be analyzed.

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 50

Correlations between Grade Point Averages
High School Rank, and an Aptitude Test

	y_0	y'_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
y_0	1.000									
y'_0	.393	1.000								
y_1	.387	.375	1.000							
y_2	.341	.298	.556	1.000						
y_3	.278	.237	.456	.490	1.000					
y_4	.270	.255	.439	.445	.562	1.000				
y_5	.240	.238	.415	.418	.496	.512	1.000			
y_6	.256	.252	.399	.383	.456	.469	.551	1.000		
y_7	.240	.219	.387	.364	.445	.442	.500	.544	1.000	
y_8	.222	.173	.342	.339	.354	.416	.453	.482	.541	1.000

Slide 51

Multigroup Analysis

LISREL can analyze data from several groups or populations. These may be different nations, states, or regions, culturally or socioeconomically different groups, groups of individuals selected on the basis of some known or unknown selection variables, groups receiving different treatments, and control groups, etc. In fact, they may be any set of mutually exclusive groups of individuals that are clearly defined. It is assumed that a number of variables have been measured on a number of individuals from each population. This approach is particularly useful in comparing a number of treatment and control groups regardless of whether individuals have been assigned to the groups randomly or not.

Slide 52

Any LISREL model may be specified and fitted for each group of data. Let $\bar{\mathbf{z}}_g$ and $\mathbf{S}_g$ be the sample mean vector and covariance matrix in group g, and let $\boldsymbol{\mu}_g(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_g(\boldsymbol{\theta})$ be the corresponding population mean vector and covariance matrix $g = 1, 2, \ldots, G$. The fit function for the multigroup case is defined as

$$\mathsf{F}(\boldsymbol{\theta}) = \sum_{g=1}^{G} \frac{N_g}{N} \mathsf{F}_g(\boldsymbol{\theta}) \ , \qquad (6)$$

where $\mathsf{F}_g(\boldsymbol{\theta}) = \mathsf{F}(\bar{\mathbf{z}}_g, \mathbf{S}_g, \boldsymbol{\mu}_g(\boldsymbol{\theta}), \boldsymbol{\Sigma}_g(\boldsymbol{\theta}))$ is any of the fit functions defined for a single group. Here N_g is the sample size in group g and $N = N_1 + N_2 + \ldots + N_G$ is the total sample size. To test the model, one can again use $c = (N-1)$ times the minimum of F as a χ^2 with degrees of freedom $d = Gk(k+1)/2 - t$, where k is the number of variables.

Slide 53

Equal Factor Structures

Example: Testing Equality of Factor Structures

The table in Slide 54 gives observed covariance matrices for two samples (N_1 = 865, N_2 = 900, respectively) of candidates who took the Scholastic Aptitude Test (SAT) in January 1971. The four measures are, in order, VERBAL40 = a 40-item verbal aptitude section, VERBAL50 = a separately timed 50-item verbal aptitude section, MATH35 = a 35-item math aptitude section, and MATH25 = a separately timed 25-item math aptitude section.

Slide 54

Covariance Matrices for SAT Verbal and Math Sections

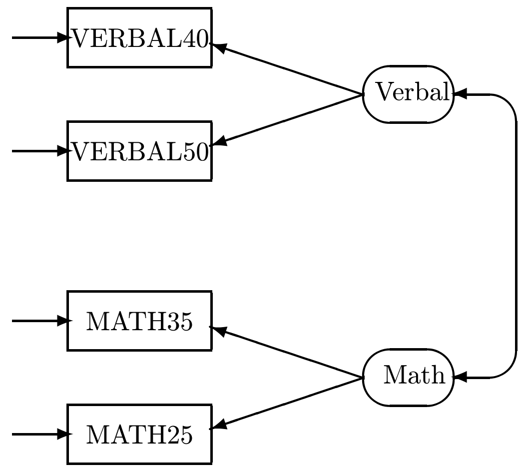
Covariance Matrix for Group 1

Tests	VERBAL40	VERBAL50	MATH35	MATH25
VERBAL40	63.382			
VERBAL50	70.984	110.237		
MATH35	41.710	52.747	60.584	
MATH25	30.218	37.489	36.392	32.295

Covariance Matrix for Group 2

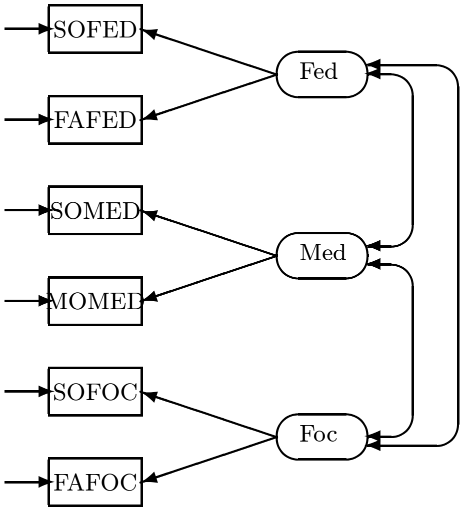
Tests	VERBAL40	VERBAL50	MATH35	MATH25
VERBAL40	67.898			
VERBAL50	72.301	107.330		
MATH35	40.549	55.347	63.203	
MATH25	28.976	38.896	39.261	35.403

Slide 55



Path Diagram for SAT Verbal and Math

Slide 56



Path Diagram for Parental Socioeconomic Characteristics

Slide 57

Equal Regressions

Consider the problem of testing whether a regression equation is the same in several populations. Suppose that a dependent variable y and a number of explanatory variables $x_1, x_2, \ldots, x_q$ are observed in two or more groups. We are interested in determining the extent to which the regression equation

$$y = \alpha + \gamma_1 x_1 + \gamma_2 x_2 + \cdots + \gamma_q x_q + z \tag{7}$$

is the same in different groups. We say that the regressions are:

- equal if $\alpha, \gamma_1, \gamma_2, \ldots, \gamma_q$, are the same in all groups
- parallel if $\gamma_1, \gamma_2, \ldots, \gamma_q$, are the same in all groups

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 58

Normally the covariance matrix of the x-variables is not expected to be the same across groups and often one finds that the intercept terms differ between groups. It may also be the case that *only some* of the regression coefficients are the same across groups.

Slide 59

Testing Equality of Regressions

Sörbom (1976) gave the covariance matrices in Slide 60. These are based on scores on the ETS Sequential Test of Educational Progress (STEP) for two groups of boys who took the test in both Grade 5 and Grade 7. The two groups were defined according to whether or not they were in the academic curriculum in Grade 12. This data will be used to demonstrate how one can test the equality of various regressions.

Slide 60

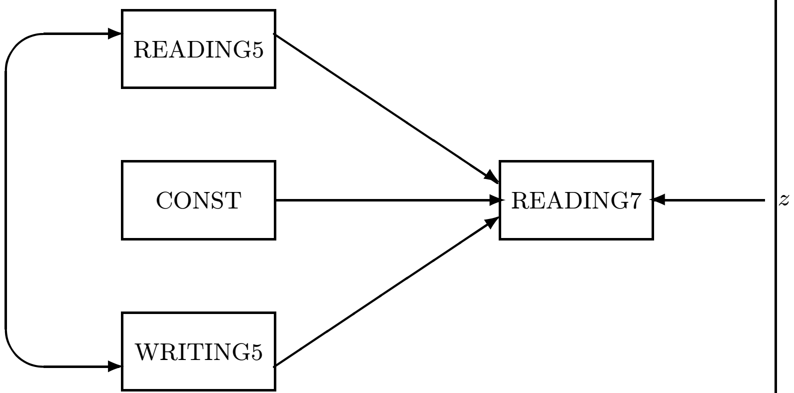
Means and Covariance Matrices for STEP Reading and Writing
Boys Academic (N = 373)

STEP Reading, Grade 5	281.349			
STEP Writing, Grade 5	184.219	182.821		
STEP Reading, Grade 7	216.739	171.699	283.289	
STEP Writing, Grade 7	198.376	153.201	208.837	246.069
Means	262.236	258.788	275.630	269.075

Boys Non-Academic (N = 249)

STEP Reading, Grade 5	174.485			
STEP Writing, Grade 5	134.468	161.869		
STEP Reading, Grade 7	129.840	118.836	228.449	
STEP Writing, Grade 7	102.194	97.767	136.058	180.460
Means	248.675	246.896	258.546	253.349

Slide 61



Path Diagram for Regression of READING7

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 62

Estimation of Means of Latent Variables

Since a latent variable is unobservable, it does not have an intrinsic scale. Neither the origin nor the unit of measurement are defined. In a single population the origin is fixed by assuming that all observed variables are measured in deviations from their means and that the means of all latent variables are zero. The unit of measurement of each latent variable is usually fixed either by assuming that it is a standardized variable with variance 1 or by fixing a non-zero loading for a reference variable.

Slide 63

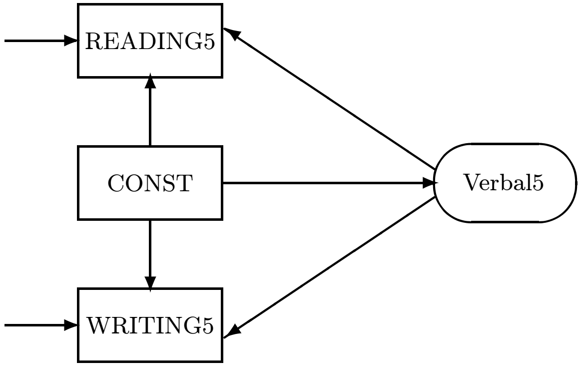
In multi-group studies, these restrictions can be relaxed somewhat by assuming that the latent variables are on the same scale in all groups. The common scale may be defined by assuming that the means of the latent variables are zero in one group and that the loadings of the observed variables on the latent variables are invariant over groups, with one loading for each latent variable fixed for a reference variable. Under these assumptions it is possible to estimate the means and covariance matrices of the latent variables relative to this common scale.

Slide 64

Mean Difference in Verbal Ability

Using the variables and data in Slide 60, take READING5 and WRITING5 to be indicators of a latent variable Verbal5 (Verbal Ability at Grade 5) and estimate the mean difference in Verbal5 between groups.

Slide 65



Path Diagram for Estimating Mean of Verbal5

Karl G Jöreskog LISREL Workshop B, Jena, Spring 2003

Slide 66

The measurement model for READING5 and WRITING5 is:

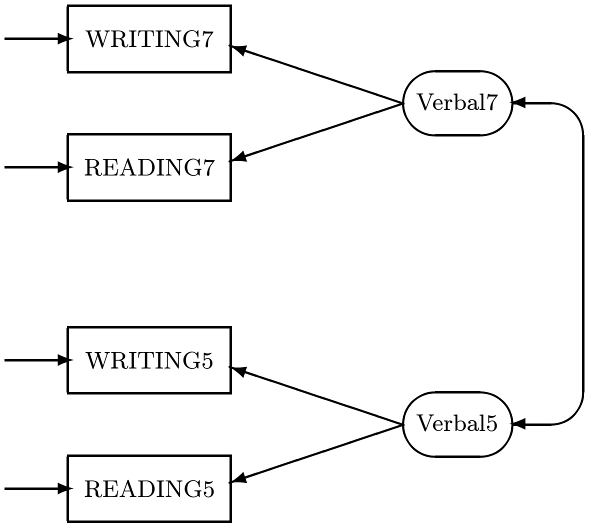
$$\texttt{READING5} = \nu_1 + \lambda_1 \texttt{Verbal5} + D1$$

$$\texttt{WRITING5} = \nu_2 + \lambda_2 \texttt{Verbal5} + D2$$

Note that these relationships now have intercept terms ν_1 and ν_2. The scale for Verbal5 is fixed by assuming that the mean is zero in Group BA and that $\lambda_1 = 1$. It is further assumed that $\nu_1, \nu_2, \lambda_1, \lambda_2$ are invariant over groups. We can then estimate the mean of Verbal5 in Group BNA as well as all the other parameters. The mean of Verbal5 is interpreted as the mean difference in verbal ability between the groups.

In the path diagram in Slide 65, the path from CONST to READING5 corresponds to ν_1, the path from CONST to WRITING5 corresponds to ν_2, and the path from CONST to Verbal5 corresponds to the mean of Verbal5.

Slide 67



Path Diagram for Estimating the Mean of Verbal5 and Verbal7

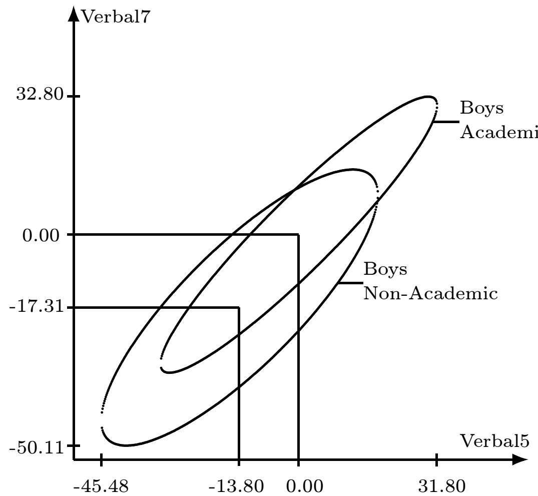
Slide 68

Estimated Means and Covariance Matrices of Verbal5 and Verbal7

Boys Academic ($N = 373$)		
	Verbal5	Verbal7
Verbal5	220.06	
Verbal7	212.11	233.59
Means	0.00	0.00

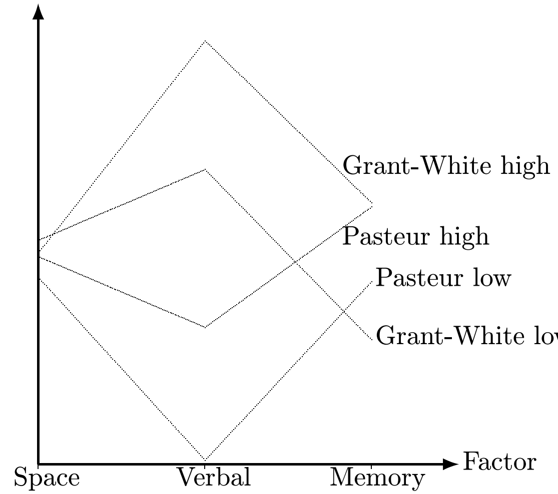
Boys Non-Academic ($N = 249$)		
	Verbal5	Verbal7
Verbal5	156.34	
Verbal7	126.96	153.73
Means	−13.80	−17.31

Slide 69



90% Confidence Ellipses for Verbal5 and Verbal7

Slide 70



Factor Mean Profiles

Slide 71

This is the last slide