Structural Equation Models Using Stata

Rose Medeiros

Stata Webinar
March 18, 2020

Contents

1 Introduction

1.1 Goals

Goals

- Learn a bit about structural equation modeling (SEM) and generalized structural equation modeling (GSEM)
- Learn about Stata's facilities for fitting SEM and GSEM

1.2 Stata Background

Stata's Tenets

- "Stata" rhymes with "data"
- Type a little, get a little
  - Click a little, get a little works fine, also
- Simple reproducibility

- Easy and complete extensibility
- Easy sharing

Notes and Slides

- Every slide in the presentations is in the notes
- The output from commands is only in the notes

Typography

- For commands, there are various fonts which can be used
  - `Items which must be typed as shown` will be in a `monospaced font`
  - *Items for which a substitution is needed* will be in *italics*
  - [*Optional items*] will be [*in square brackets*], though the brackets do not get typed
- Example Stata commands will often be preceded by a `.`
  - The `.` is a prompt and does not get typed—it is for distinguishing input from output in the notes
    In the handouts, the commands are both boldfaced and slanted. This is done so that they are easier to see on the page (even though it conflicts with the above rules).

1.3 What is SEM?

Descriptions of Linear SEM

- SEM is a class of statistical techniques that allows us to test hypotheses about relationships among variables
- SEM may also be referred to as Analysis of Covariance Structures
  - SEM fits models using the observed covariances and, possibly, means
- SEM encompasses other statistical methods such as correlation, linear regression, and factor analysis

---

Descriptions of Linear SEM (Continued)

- SEM is a multivariate technique that allows us to estimate a system of equations
  - Variables in these equations may be measured with error
  - There may be variables in the model that cannot be measured directly

---

SEM in Stata

- The `sem` command is used to fit standard linear SEM models
- The `gsem` command is used to fit generalizations of linear SEM
  - Generalized linear models including continuous, binary, ordinal, nominal, count, and survival outcomes
  - Multilevel models including random intercepts, random slopes, and crossed effects
- Both types of models can be specified using either commands or path diagrams via the SEM Builder

---

Types of Variables

- Observed variables (manifest variables)
  - ◇ We have observed values of these variables in our dataset
  - ◇ Or we have variances, covariances, and possibly means based on observed values
- Latent variables (unobserved variables, factors)
  - ◇ May represent the following
    - ⋆ Hypothetical constructs
    - ⋆ Variables that cannot be directly measured
    - ⋆ True values of variables measured with error
    - ⋆ Unobserved heterogeneity
    - ⋆ Errors or disturbances
  - ◇ Latent variables are measured by observed variables known as measurements or indicators

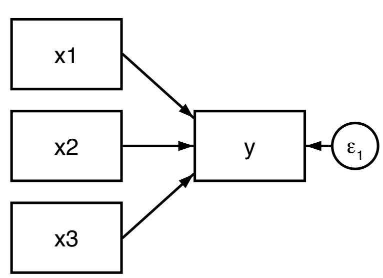
---

Types of Variables (Continued)

- Endogenous variables
  - ◇ Also known as y, dependent, or response variables
  - ◇ Predicted by one or more other variables
  - ◇ May predict other endogenous variables
- Exogenous variables
  - ◇ Also known as x, independent, or explanatory variables
  - ◇ Variables that are not predicted by any other variables in the model
  - ◇ May be correlated with other variables in the model

Path Diagrams

- Observed variables are represented by rectangles
- Latent variables are represented by ovals or circles
- Paths are represented by arrows
- Covariances are represented by curved lines with arrows at each end
- The path diagram below corresponds to a linear regression of y on x1, x2, and x3



---

2 Linear SEM

2.1 Introduction

Data

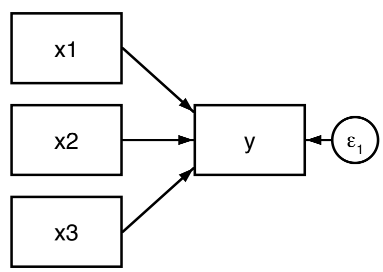
- sem allows two types of data
- Datasets with individual observations
- Datasets made up of summary statistics, specifically covariance or correlation matrices; and possibly means
  - See help ssd for more information about working with summary statistics data

---

Basic Syntax

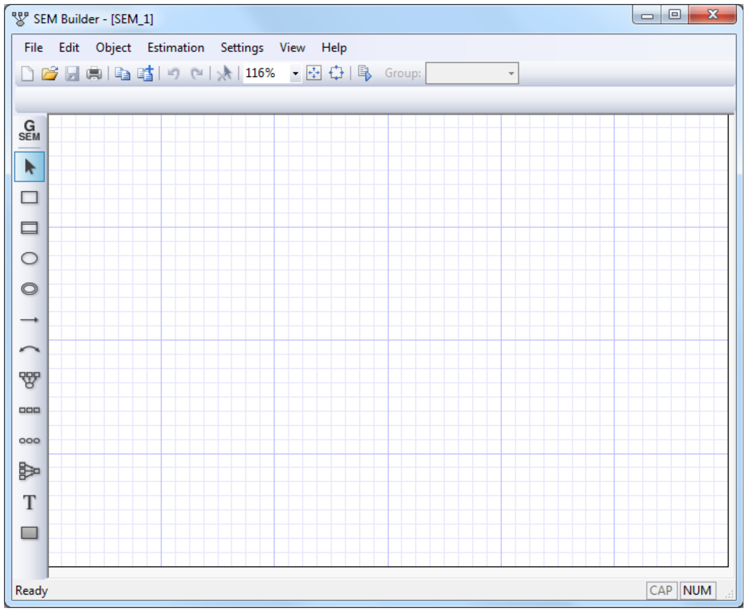
- sem *paths* [if] [in] [weight] [, options]
- The basic rules are
  - All paths are placed inside parentheses
  - Arrows point towards dependent variables
- Beyond that the *paths* specifications are flexible
- The three commands below all fit the path model shown here

```
. sem (y <- x1 x2 x3)
. sem (x1 x2 x3 -> y)
. sem (x1 -> y) (x2 -> y) (x3 -> y)
```



The SEM Builder

- To open the SEM Builder, type sembuilder or click on **Statistics** > **SEM (structural equation modeling)** > **Model building and estimation**



- The tools along the left-hand side allow us to draw the path diagram
- Use the menus at the top to customize the appearance of the path diagram, fit the model, customize the appearance of the results, and more

---

2.2 Path Models

Path Analysis

- We will begin by looking at some examples of path analysis
- Path models include only observed variables and their error terms
- These models can be simple or they may include many observed variables with intricate relationships

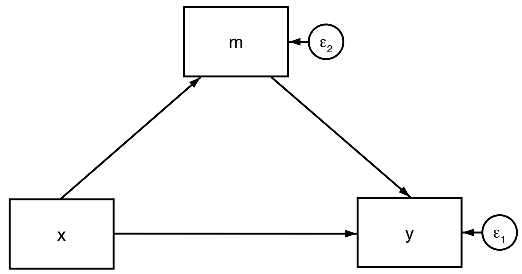
---

Mediation Models

- In a mediation model, a variable `x` is hypothesized to predict `y` in two ways
  - ◇ Directly
  - ◇ Indirectly because `x` predicts a third variable `m` which in turn predicts `y`
- Using `sem` we can fit all of the equations in the mediation model simultaneously

---

A Simple Mediation Model



- The command for the above model is

```
. sem (m <- x) (y <- x m)
```

Job Satisfaction Data

- Fogarty et al. (1999) fit a variety of models that examine relationships among positive and negative affectivity, stress, coping, strain, and job satisfaction
  - ◇ We'll fit a much simpler model
- The data for this example are stored as summary statistics in `jobsat.dta`

```
. use jobsat

(Data from Fogarty et al. (1999))
```

- We can learn about the variables

```
. ssd describe

  Summary statistics data from jobsat.dta
    obs:            114                Data from Fogarty et al. (1999)
   vars:              6                16 Sep 2015 19:24
  ------------------------------------------------------------------------
  variable name                variable label
  ------------------------------------------------------------------------
  stress                       sum of environmental state items
  coping                       sum of resources for dealing with stre..
  strain                       sum of personal reaction to stress items
  na                           sum of negative affectivity items (neg..
  pa                           sum of positive affectivity items (pos..
  satisfaction                 sum of job satisfaction items
  ------------------------------------------------------------------------
```

- We can also look at their summary statistics

```
. ssd list

  Observations = 114

  Means:
        stress        coping        strain            na            pa  satisfaction
        136.11        127.72         80.91         18.57          35.4         60.81
```

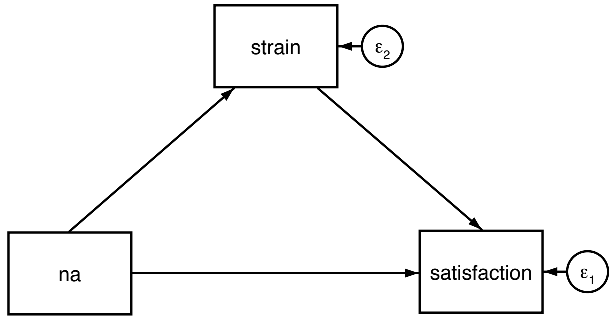
```
  Standard deviations:
        stress        coping        strain            na            pa  satisfaction
         22.48         18.13         19.65             7          6.22         10.84

  Correlations:
        stress        coping        strain            na            pa  satisfaction
             1
          -.38             1
           .59          -.58             1
           .26          -.42           .58             1
          -.29           .49          -.46          -.28             1
          -.46           .27          -.51          -.21           .42             1
```

Fitting a Mediation Model

- As an example, we can fit a model where strain mediates the effect of negative affectivity on job satisfaction

```
. sem (satisfaction <- strain na) (strain <- na)

Endogenous variables

Observed:  satisfaction strain

Exogenous variables

Observed:  na

Fitting target model:

Iteration 0:   log likelihood = -1275.3885
Iteration 1:   log likelihood = -1275.3885

Structural equation model                       Number of obs     =        114
Estimation method  = ml
Log likelihood     = -1275.3885

-----------------------------------------------------------------------------------
                   |                 OIM
                   |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+---------------------------------------------------------------
Structural         |
  satisfaction     |
            strain |  -.3227126   .0541461    -5.96   0.000     -.428837   -.2165882
                na |   .2002222   .1519959     1.32   0.188    -.0976842    .4981286
             _cons |   83.20255    3.68243    22.59   0.000     75.98512    90.41998
  -----------------+---------------------------------------------------------------
  strain           |
                na |   1.628143   .2141733     7.60   0.000     1.208371    2.047915
             _cons |   50.67539   4.248061    11.93   0.000     42.34934    59.00143
-------------------+---------------------------------------------------------------
var(e.satisfaction)|   84.88763   11.24364                      65.47869    110.0497
      var(e.strain)|   253.9833    33.6409                      195.9118     329.268
-----------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(0)   =      0.00, Prob > chi2 =      .
```

- In path diagram form the model is



Direct, Indirect, and Total Effects

- Direct effects are the coefficient estimates we see in the output
  - The direct effect of `na` on `satisfaction` is .2
- We can calculate the indirect effect of negative affectivity on job satisfaction by multiplying the appropriate coefficients
  - The path coefficient from `na` to `strain` is roughly 1.62
  - The path coefficient from `strain` to `satisfaction` is roughly $-.322$
  - $1.62 \times -.322 = -.522$ so the indirect effect is roughly $-.52$
- The total effect is the sum of the direct and indirect effects
  - The total effect of negative affectivity on job satisfaction is $.2 + (-.52) = -.32$

`estat teffects`

- We can obtain direct, indirect, and total effects and their standard errors using `estat teffects`

```
. estat teffects

Direct effects
------------------------------------------------------------------------------
               |                 OIM
               |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+--------------------------------------------------------------
Structural     |
  satisfaction |
        strain |  -.3227126   .0541461    -5.96   0.000     -.428837   -.2165882
            na |   .2002222   .1519959     1.32   0.188    -.0976842    .4981286
  -------------+--------------------------------------------------------------
  strain       |
            na |   1.628143   .2141733     7.60   0.000     1.208371    2.047915
------------------------------------------------------------------------------

Indirect effects
```

```
------------------------------------------------------------------------------
               |                 OIM
               |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+--------------------------------------------------------------
Structural     |
  satisfaction |
        strain |          0  (no path)
            na |  -.5254222   .1120216    -4.69   0.000    -.7449805   -.3058639
  -------------+--------------------------------------------------------------
  strain       |
            na |          0  (no path)
------------------------------------------------------------------------------

Total effects
------------------------------------------------------------------------------
               |                 OIM
               |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+--------------------------------------------------------------
Structural     |
  satisfaction |
        strain |  -.3227126   .0541461    -5.96   0.000     -.428837   -.2165882
            na |     -.3252   .1418029    -2.29   0.022    -.6031285   -.0472715
  -------------+--------------------------------------------------------------
  strain       |
            na |   1.628143   .2141733     7.60   0.000     1.208371    2.047915
------------------------------------------------------------------------------
```

- What do we see?
  - The direct effect of `na` on `satisfaction` is not significantly different from 0
  - The direct effect of `strain` on `satisfaction` is significant different from 0
  - The indirect effect of `na` on `satisfaction` is also significantly different from 0
  - The direct effect of `na` on `satisfaction` can be said to be fully mediated by `strain`

Standardized coefficients

- We can replay the model with standardized coefficients using

```
. sem, standardized

Structural equation model                       Number of obs     =        114
Estimation method  = ml
Log likelihood     = -1275.3885

----------------------------------------------------------------------------------
                   |                 OIM
      Standardized |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+--------------------------------------------------------------
Structural         |
  satisfaction     |
            strain |   -.584991   .0857388    -6.82   0.000     -.753036   -.4169459
                na |   .1292948   .0974336     1.33   0.185    -.0616716    .3202611
             _cons |   7.709399   .4148133    18.59   0.000      6.89638    8.522419
  -----------------+--------------------------------------------------------------
  strain           |
                na |        .58   .0566844    10.23   0.000     .4689007    .6910993
             _cons |   2.590286   .3461981     7.48   0.000      1.91175    3.268822
-------------------+--------------------------------------------------------------
```

```
var(e.satisfaction)|   .7288065   .0709659                       .602183    .8820557
      var(e.strain)|      .6636   .0657539                      .5464667    .8058404
-------------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(0)   =      0.00, Prob > chi2 =      .
```

- We expect that a one standard deviation increase in strain would produce a .58 standard deviation decrease in satisfaction, holding negative affectivity constant
- We can also obtain standardized indirect and total effects

```
. estat teffects, standardized nodirect

Indirect effects
---------------------------------------------------------------------------------
               |                OIM
               |      Coef.   Std. Err.      z    P>|z|              Std. Coef.
---------------+-----------------------------------------------------------------
Structural     |
  satisfaction |
        strain |          0  (no path)                                        0
            na |  -.5254222   .1120216    -4.69   0.000               -.3392948
  -------------+-----------------------------------------------------------------
  strain       |
            na |          0  (no path)                                        0
---------------------------------------------------------------------------------

Total effects
---------------------------------------------------------------------------------
               |                OIM
               |      Coef.   Std. Err.      z    P>|z|              Std. Coef.
---------------+-----------------------------------------------------------------
Structural     |
  satisfaction |
        strain |  -.3227126   .0541461    -5.96   0.000                -.584991
            na |     -.3252   .1418029    -2.29   0.022                    -.21
  -------------+-----------------------------------------------------------------
  strain       |
            na |   1.628143   .2141733     7.60   0.000                     .58
---------------------------------------------------------------------------------
```

  - ◇ The standardized indirect effects are the products of standardized coefficients

---

Equation Level Goodness-of-fit

- We can also obtain variance decomposition and R-squared for each of the endogenous variables in the model

```
. estat eqgof

Equation-level goodness of fit

-------------------------------------------------------------------------------
             |            Variance             |
     depvars |    fitted  predicted   residual | R-squared        mc       mc2
-------------+---------------------------------+-------------------------------
observed     |                                 |
satisfaction |  116.4748   31.58722   84.88763 |  .2711935  .5207624  .2711935
      strain |  382.7355   128.7522   253.9833 |     .3364       .58     .3364
-------------+---------------------------------+-------------------------------
     overall |                                 |  .3463495
-------------------------------------------------------------------------------
```

```
mc  = correlation between depvar and its prediction
mc2 = mc^2 is the Bentler-Raykov squared multiple correlation coefficient
```

- About 27% of variation in `satisfaction` is explained by the model
- To obtain a Wald test for the null hypothesis that all coefficients in an equation are 0 we can use

```
. estat eqtest

Wald tests for equations
--------------------------------------
             |     chi2   df         p
-------------+------------------------
observed     |
satisfaction |    42.42    2    0.0000
      strain |    57.79    1    0.0000
--------------------------------------
```

Nonrecursive Path Models

- The examples so far have been of recursive models
- Models that contain feedback loops or correlated error terms are said to be nonrecursive
- `sem` can be used to fit nonrecursive models
- Checking whether a nonrecursive model is identified can be simple or complex, depending on the model
- Stata provides tools for evaluating identification of nonrecursive models, for more information see `help estat stable`

2.3 Models with Latent Variables

Models with Latent Variables

- Models with latent variables can estimate and accommodate measurement error
- Some variables cannot be directly measured without error
  - ◇ But we may be able to collect multiple measurements each of which contains some error
- Confirmatory factor analysis (CFA) allows us to evaluate how well the items we collect measure the corresponding concept
- CFA models are also called measurement models
- Examples: Personality features, depression, attitudes

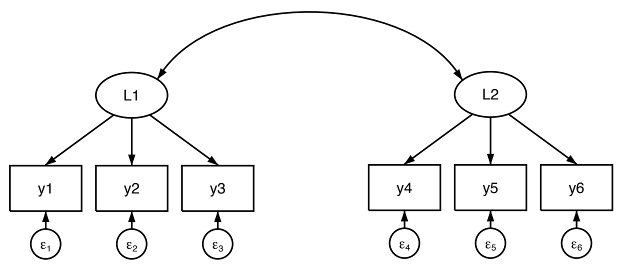
Confirmatory Factor Analysis

- In a confirmatory factor analysis (CFA) model, one or more latent variables is measured by a series of observed variables
  - The latent variables may be correlated, but no structural paths are specified
- Each latent variable is associated with a set of observed variables
- These models are confirmatory in the sense that we specify them based on prior knowledge or theory about
  - What the latent variables represent
  - Which observed variables are associated with each latent variable
- This is unlike exploratory factor analysis where all observed variables are allowed to measure each of the latent variables

---

Confirmatory Factor Analysis (continued)

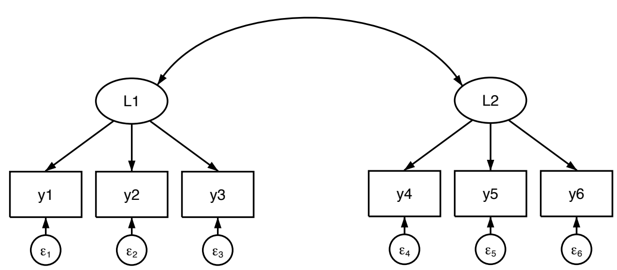
- At least 3 measurement variables are required to identify a model that contains only a single latent variable
- Sometimes also called a measurement model
- Here is an example of a path diagram for a CFA model



Syntax for a CFA Model

- By default, variables with names beginning with a capital letter are assumed to be latent variables
- Three equivalent methods of fitting the CFA model shown below are

```
. sem (L1 -> y1 y2 y3) (L2 -> y4 y5 y6)
. sem (y1 y2 y3 <- L1) (y4 y5 y6 <- L2)
. sem (L1 -> y1) (L1 -> y2) (L1 -> y3) ///
.        (L2 -> y4) (L2 -> y5) (L2 -> y6)
```



Education Data

- Holzinger and Swineford (1939) measured the abilities of students across a variety of areas
- Five of the observed variables measure verbal skills
- Let's open the data and take a look at these variables

```
. use hsdata
. codebook general paragraph sentence wordc wordm

(Data from Holzinger and Swineford (1939))

-------------------------------------------------------------------------------------
general                                                           general information
-------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [8,84]                       units:  1
         unique values:  57                       missing .:  0/301

                  mean:   40.5914
              std. dev:   12.3807

           percentiles:        10%       25%       50%       75%       90%
                                24        31        41        49        56

-------------------------------------------------------------------------------------
paragraph                                                     paragraph comprehension
-------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,19]                       units:  1
         unique values:  20                       missing .:  0/301

                  mean:   9.18272
              std. dev:   3.49235

           percentiles:        10%       25%       50%       75%       90%
                                 5         7         9        11        14

-------------------------------------------------------------------------------------
sentence                                                          sentence completion
-------------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [4,28]                       units:  1
         unique values:  25                       missing .:  0/301

                  mean:   17.3621
              std. dev:   5.16189

           percentiles:        10%       25%       50%       75%       90%
                                10        14        18        21        24

-------------------------------------------------------------------------------------
wordc                                                             word classification
-------------------------------------------------------------------------------------
```

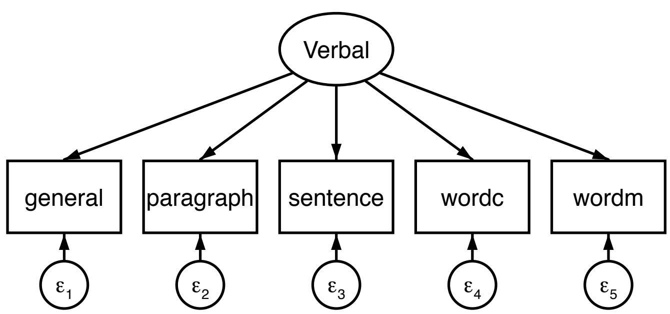
```
                 type:  numeric (float)

                range:  [10,43]                      units:  1
        unique values:  31                       missing .:  0/301

                 mean:   26.1262
             std. dev:   5.67544

          percentiles:        10%       25%       50%       75%       90%
                               19        23        26        30        33

------------------------------------------------------------------------------------------
wordm                                                                         word meaning
------------------------------------------------------------------------------------------

                 type:  numeric (float)

                range:  [1,43]                       units:  1
        unique values:  40                       missing .:  0/301

                 mean:    15.299
             std. dev:   7.66922

          percentiles:        10%       25%       50%       75%       90%
                                7        10        14        19        26
```

- We'll use the observed variables `general`, `paragraph`, `sentence`, `wordc`, and `wordm` as indicators for the latent variable representing verbal abilities

---

Fitting a Single Factor CFA

- The latent variable, Verbal, is assumed to cause the observed variables



- The `sem` command to fit the model is

```
. sem (Verbal -> general paragraph sentence wordc wordm)

Endogenous variables

Measurement:  general paragraph sentence wordc wordm

Exogenous variables
```

```
Latent:       Verbal

Fitting target model:

Iteration 0:   log likelihood = -4403.8152
Iteration 1:   log likelihood = -4403.4283
Iteration 2:   log likelihood = -4403.4268
Iteration 3:   log likelihood = -4403.4268

Structural equation model                       Number of obs     =        301
Estimation method  = ml
Log likelihood     = -4403.4268

 ( 1)  [general]Verbal = 1
--------------------------------------------------------------------------------
                |                 OIM
                |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+---------------------------------------------------------------
Measurement     |
  general       |
         Verbal |          1  (constrained)
          _cons |   40.59136    .712427    56.98   0.000     39.19503    41.98769
  --------------+---------------------------------------------------------------
  paragraph     |
         Verbal |   .2754032   .0165074    16.68   0.000     .2430493    .3077571
          _cons |   9.182724    .200961    45.69   0.000     8.788848    9.576601
  --------------+---------------------------------------------------------------
  sentence      |
         Verbal |   .4349704     .02364    18.40   0.000     .3886368    .4813039
          _cons |   17.36213   .2970317    58.45   0.000     16.77995     17.9443
  --------------+---------------------------------------------------------------
  wordc         |
         Verbal |    .403284   .0277438    14.54   0.000     .3489072    .4576608
          _cons |   26.12625   .3265834    80.00   0.000     25.48615    26.76634
  --------------+---------------------------------------------------------------
  wordm         |
         Verbal |   .6238506   .0348107    17.92   0.000     .5556229    .6920784
          _cons |     15.299   .4413117    34.67   0.000     14.43405    16.16396
----------------+---------------------------------------------------------------
  var(e.general)|   45.59075   4.800626                      37.08914    56.04112
var(e.paragraph)|   4.026519   .4034455                      3.308582    4.900244
 var(e.sentence)|   6.277734   .7412008                      4.980845      7.9123
    var(e.wordc)|   14.67172     1.3443                      12.25997     17.5579
    var(e.wordm)|   16.90726   1.805241                      13.71475    20.84293
     var(Verbal)|   107.1825   12.26806                      85.64373     134.138
--------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(5)   =     18.74, Prob > chi2 = 0.0021
```

Fitting a Single Factor CFA (Continued)

- For each observed variable we obtain an intercept and path coefficient
  - The path from the latent variable to the first observed variable is constrained to 1 for identification
- The error variances for our indicators represent the portion of the indicator's variance that is not explained by the latent variable `Verbal`
- The overall model χ^2 tests the null hypothesis that the covariance matrix implied by our model is equal to the observed covariance matrix in the population

◇ If the model includes means, the respective mean vectors are included in this test

---

Examining Model Fit

- We can examine the fit of this model using

```
. estat gof, stats(all)

----------------------------------------------------------------------------
Fit statistic        |      Value   Description
---------------------+------------------------------------------------------
Likelihood ratio     |
          chi2_ms(5) |     18.739   model vs. saturated
            p > chi2 |      0.002
         chi2_bs(10) |   1005.075   baseline vs. saturated
            p > chi2 |      0.000
---------------------+------------------------------------------------------
Population error     |
               RMSEA |      0.096   Root mean squared error of approximation
 90% CI, lower bound |      0.052
         upper bound |      0.144
              pclose |      0.043   Probability RMSEA <= 0.05
---------------------+------------------------------------------------------
Information criteria |
                 AIC |   8836.854   Akaike's information criterion
                 BIC |   8892.460   Bayesian information criterion
---------------------+------------------------------------------------------
Baseline comparison  |
                 CFI |      0.986   Comparative fit index
                 TLI |      0.972   Tucker-Lewis index
---------------------+------------------------------------------------------
Size of residuals    |
                SRMR |      0.020   Standardized root mean squared residual
                  CD |      0.918   Coefficient of determination
----------------------------------------------------------------------------
```

- The first χ^2 test is the same test reported in the output from `sem`
- The second χ^2 compares the saturated model with a baseline model that includes
  - ◇ The means and variances of all observed variables, and
  - ◇ The covariances of all observed exogenous variables
  - ◇ Different authors define the baseline model differently

---

More Model Fit

- A variety of measures of fit have been proposed for SEM
  - ◇ For many of these measures a variety of standards for what constitutes good or acceptable fit have also been proposed
- `sem` provides the following
  - ◇ RMSEA or root mean square error of approximation
    - ⋆ The p-value labeled pclose corresponds to a test of RMSEA$< .05$
  - ◇ AIC and BIC

◇ The comparative fit index (CFI) and Tucker-Lewis index (TLI)
◇ Standardized root mean squared residual (SRMR)
◇ Coefficient of determination (CD)

Standardized Coefficients

- We can replay the model to obtain standardized coefficients

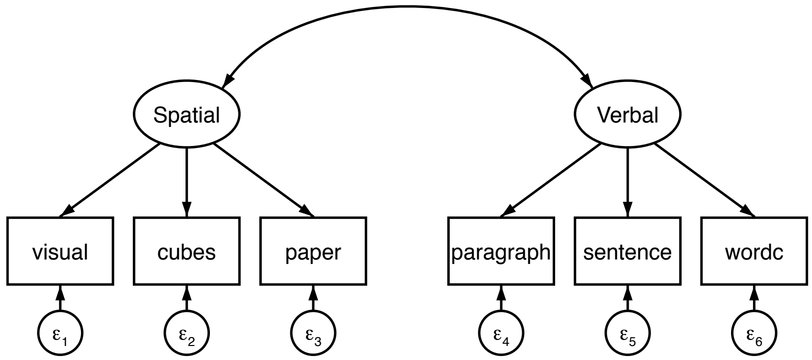
```
. sem, standardized

Structural equation model                       Number of obs     =        301
Estimation method  = ml
Log likelihood     = -4403.4268

 ( 1)  [general]Verbal = 1
--------------------------------------------------------------------------------
                |                 OIM
   Standardized |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+---------------------------------------------------------------
Measurement     |
  general       |
         Verbal |   .8376031   .0209013    40.07   0.000     .7966373    .8785688
          _cons |   3.284052    .145731    22.54   0.000     2.998424     3.56968
  --------------+---------------------------------------------------------------
  paragraph     |
         Verbal |   .8177789   .0224446    36.44   0.000     .7737883    .8617694
          _cons |   2.633762   .1218401    21.62   0.000      2.39496    2.872564
  --------------+---------------------------------------------------------------
  sentence      |
         Verbal |   .8738473    .017864    48.92   0.000     .8388345      .90886
          _cons |   3.369123   .1489219    22.62   0.000     3.077242    3.661005
  --------------+---------------------------------------------------------------
  wordc         |
         Verbal |    .736878   .0293638    25.09   0.000      .679326    .7944299
          _cons |   4.611049   .1965727    23.46   0.000     4.225773    4.996324
  --------------+---------------------------------------------------------------
  wordm         |
         Verbal |   .8435557     .02038    41.39   0.000     .8036117    .8834998
          _cons |   1.998179   .0997732    20.03   0.000     1.802627    2.193731
----------------+---------------------------------------------------------------
  var(e.general)|   .2984211    .035014                      .2371141    .3755793
var(e.paragraph)|   .3312377   .0367094                      .2665665    .4115988
 var(e.sentence)|    .236391   .0312208                      .1824779    .3062326
    var(e.wordc)|   .4570109    .043275                       .379599    .5502094
    var(e.wordm)|   .2884137   .0343833                      .2283178    .3643277
     var(Verbal)|          1          .                             .           .
--------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(5)   =     18.74, Prob > chi2 = 0.0021
```

- The standardized loading is the correlation between the latent variable and the observed variable when each indicator measures only a single latent variable
- The standardized error variances are the proportion of variation not explained by the latent variable

---

Two Factor CFA

- Now we'll add a second latent variable called `Spatial` which represents students' spacial abilities, using the indicators `visual`, `cube`, and `paper`



  ◇ Read nothing into the fact that we have reduced the number of items for the variable `Verbal` from five to three

---

Fitting a Two Factor CFA

- We can fit this model using

```
. sem (Spatial -> visual cubes paper) ///
      (Verbal -> paragraph sentence wordc)

Endogenous variables

Measurement:  visual cubes paper paragraph sentence wordc

Exogenous variables

Latent:       Spatial Verbal

Fitting target model:

Iteration 0:   log likelihood = -5046.1909
Iteration 1:   log likelihood = -5046.0419
Iteration 2:   log likelihood = -5046.0415

Structural equation model                       Number of obs     =        301
Estimation method  = ml
Log likelihood     = -5046.0415

 ( 1)  [visual]Spatial = 1
 ( 2)  [paragraph]Verbal = 1
------------------------------------------------------------------------------------
                   |                 OIM
                   |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
Measurement        |
  visual           |
           Spatial |          1  (constrained)
             _cons |   29.61462   .4030662    73.47   0.000     28.82462    30.40461
  -----------------+----------------------------------------------------------------
```

```
  cubes          |
         Spatial |    .364576   .0793793     4.59   0.000     .2089954    .5201566
           _cons |   24.35216   .2710169    89.85   0.000     23.82098    24.88334
  ---------------+----------------------------------------------------------------
  paper          |
         Spatial |   .2662524   .0525865     5.06   0.000     .1631847    .3693201
           _cons |   14.22924   .1628644    87.37   0.000     13.91003    14.54844
  ---------------+----------------------------------------------------------------
  paragraph      |
          Verbal |          1  (constrained)
           _cons |   9.182724   .2009609    45.69   0.000     8.788848      9.5766
  ---------------+----------------------------------------------------------------
  sentence       |
          Verbal |   1.618814   .1038495    15.59   0.000     1.415273    1.822355
           _cons |   17.36213   .2970316    58.45   0.000     16.77995     17.9443
  ---------------+----------------------------------------------------------------
  wordc          |
          Verbal |   1.484747   .1104262    13.45   0.000     1.268316    1.701179
           _cons |   26.12625   .3265833    80.00   0.000     25.48615    26.76634
-----------------+----------------------------------------------------------------
     var(e.visual)|   21.41672   5.034175                      13.51049    33.94961
      var(e.cubes)|   18.45538   1.700839                      15.40553    22.10901
      var(e.paper)|   6.035589   .6176517                      4.938693    7.376109
  var(e.paragraph)|   4.064789   .4958005                      3.200465    5.162533
   var(e.sentence)|   5.353155   1.049893                      3.644744    7.862355
      var(e.wordc)|   14.26686   1.429614                      11.72286    17.36294
      var(Spatial)|   27.48446   5.927568                      18.00979     41.9436
       var(Verbal)|   8.091177   1.004038                      6.344329      10.319
-----------------+----------------------------------------------------------------
cov(Spatial,Verbal)|   7.390188   1.374653     5.38   0.000     4.695917    10.08446
----------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(8)   =     15.45, Prob > chi2 = 0.0509
```

- We can examine model fit using

```
. estat gof, stats(all)

----------------------------------------------------------------------------
Fit statistic        |      Value   Description
---------------------+------------------------------------------------------
Likelihood ratio     |
         chi2_ms(8) |     15.454   model vs. saturated
           p > chi2 |      0.051
        chi2_bs(15) |    559.669   baseline vs. saturated
           p > chi2 |      0.000
---------------------+------------------------------------------------------
Population error     |
              RMSEA |      0.056   Root mean squared error of approximation
 90% CI, lower bound |      0.000
        upper bound |      0.097
             pclose |      0.360   Probability RMSEA <= 0.05
---------------------+------------------------------------------------------
Information criteria |
                AIC |  10130.083   Akaike's information criterion
                BIC |  10200.518   Bayesian information criterion
---------------------+------------------------------------------------------
Baseline comparison  |
                CFI |      0.986   Comparative fit index
                TLI |      0.974   Tucker-Lewis index
---------------------+------------------------------------------------------
Size of residuals    |
```

```
                 SRMR |      0.029   Standardized root mean squared residual
                   CD |      0.950   Coefficient of determination
----------------------------------------------------------------------------
```

Modification Indices

- MIs are used to check for paths and covariances that could be added to the model to improve model fit
  - ◇ Over-fitting is a serious danger here
- Approximate change in the χ^2 statistic if the parameter is added to the model
- To obtain modification indices for our model we can type

```
. estat mindices

Modification indices

------------------------------------------------------------------------
                         |                                      Standard
                         |        MI     df   P>MI        EPC        EPC
-------------------------+----------------------------------------------
Measurement              |
  paragraph              |
                 Spatial |     4.815      1   0.03   .0915926   .1377239
  -----------------------+----------------------------------------------
  sentence               |
                 Spatial |    12.089      1   0.00   -.215882  -.2196211
-------------------------+----------------------------------------------
cov(e.visual,e.paragraph)|     5.118      1   0.02   1.903032   .2039627
 cov(e.visual,e.sentence)|     5.506      1   0.02  -2.867484  -.2678056
 cov(e.paragraph,e.wordc)|    12.089      1   0.00  -4.494894  -.5902503
  cov(e.sentence,e.wordc)|     4.815      1   0.03   4.997621    .571866
------------------------------------------------------------------------
EPC = expected parameter change
```

- The largest MIs are associated with
  - ◇ Adding a path from `Spatial` to `sentence`
  - ◇ Adding a covariance between the error terms for `paragraph` and `wordc` (word classification)

Refitting our Model

- We can use the `var()` option to add the suggested covariance between error terms
- In this case we use `var(e.paragraph*e.wordc)`

```
. sem (Spatial -> visual cubes paper) ///
      (Verbal -> paragraph sentence wordc), ///
      var(e.paragraph*e.wordc)

Endogenous variables

Measurement:  visual cubes paper paragraph sentence wordc

Exogenous variables

Latent:       Spatial Verbal
```

```
Fitting target model:

Iteration 0:   log likelihood = -5046.1909
Iteration 1:   log likelihood = -5042.6382
Iteration 2:   log likelihood = -5039.6371
Iteration 3:   log likelihood = -5039.4847
Iteration 4:   log likelihood = -5039.4837
Iteration 5:   log likelihood = -5039.4837

Structural equation model                       Number of obs     =        301
Estimation method  = ml
Log likelihood     = -5039.4837

 ( 1)  [visual]Spatial = 1
 ( 2)  [paragraph]Verbal = 1
------------------------------------------------------------------------------------------
                        |                 OIM
                        |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------------------+-----------------------------------------------------------------
Measurement             |
  visual                |
                Spatial |          1  (constrained)
                  _cons |   29.61462   .4030668    73.47   0.000     28.82462    30.40461
  ----------------------+-----------------------------------------------------------------
  cubes                 |
                Spatial |   .3567101   .0747435     4.77   0.000     .2102156    .5032046
                  _cons |   24.35216   .2710172    89.85   0.000     23.82098    24.88334
  ----------------------+-----------------------------------------------------------------
  paper                 |
                Spatial |   .2613528   .0489846     5.34   0.000     .1653447    .3573608
                  _cons |   14.22924   .1628647    87.37   0.000     13.91003    14.54844
  ----------------------+-----------------------------------------------------------------
  paragraph             |
                 Verbal |          1  (constrained)
                  _cons |   9.182724    .200961    45.69   0.000     8.788848    9.576601
  ----------------------+-----------------------------------------------------------------
  sentence              |
                 Verbal |   1.180813      .1409     8.38   0.000     .9046541    1.456972
                  _cons |   17.36213   .2970317    58.45   0.000     16.77995     17.9443
  ----------------------+-----------------------------------------------------------------
  wordc                 |
                 Verbal |   1.497145   .1099068    13.62   0.000     1.281732    1.712559
                  _cons |   26.12625   .3265834    80.00   0.000     25.48615    26.76634
------------------------+-----------------------------------------------------------------
          var(e.visual)|    20.8425    4.75466                      13.32818    32.59333
           var(e.cubes)|   18.53828   1.683387                      15.51584    22.14948
           var(e.paper)|   6.067431   .6035889                      4.992604    7.373649
       var(e.paragraph)|   1.006939   1.244935                      .0892514    11.36034
        var(e.sentence)|   11.01125   1.867395                      7.897352    15.35294
           var(e.wordc)|   7.113704   3.023351                      3.092659    16.36288
           var(Spatial)|   28.05882    5.72042                      18.81624    41.84137
            var(Verbal)|   11.14904   1.586895                      8.434941    14.73645
------------------------+-----------------------------------------------------------------
cov(e.paragraph,e.wordc)|  -5.196533   1.796904    -2.89   0.004    -8.718401   -1.674665
     cov(Spatial,Verbal)|   8.726427   1.422593     6.13   0.000     5.938197    11.51466
------------------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(7)   =      2.34, Prob > chi2 = 0.9388
```

- In the SEM Builder, you can add a curved, double-headed arrow between `paragraph` and `wordc`

- After fitting the model we could check model fit to see if our model fits better. We could also check the MIs again.

Full Structural Equation Models

- Combine path analysis and confirmatory factor analysis
- One or more latent variables are included in the model with corresponding observed indicators
- Structural relationships may exist among latent and/or observed variables

---

Data on Alienation

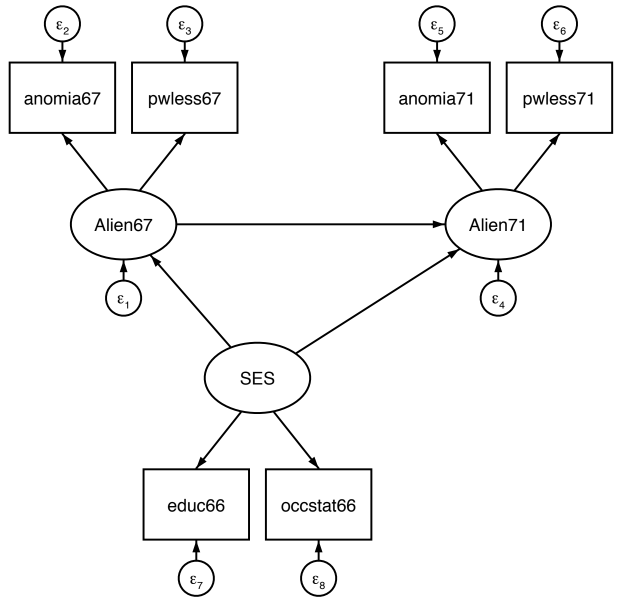
- The data for this set of examples come from Wheaton et al. (1977)

```
. use wheaton
. ssd describe

(Structural model with measurement component)

  Summary statistics data from wheaton.dta
    obs:            932                Structural model with measurem..
   vars:             13                8 Jun 2012 11:28
                                       (_dta has notes)
  -------------------------------------------------------------------------
  variable name                variable label
  -------------------------------------------------------------------------
  educ66                       Education, 1966
  occstat66                    Occupational status, 1966
  anomia66                     Anomia, 1966
  pwless66                     Powerlessness, 1966
  socdist66                    Latin American social distance, 1966
  occstat67                    Occupational status, 1967
  anomia67                     Anomia, 1967
  pwless67                     Powerlessness, 1967
  socdist67                    Latin American social distance, 1967
  occstat71                    Occupational status, 1971
  anomia71                     Anomia, 1971
  pwless71                     Powerlessness, 1971
  socdist71                    Latin American social distance, 1971
  -------------------------------------------------------------------------
```

- The model includes three latent variables with structural paths between them
- Alienation in 1971 is predicted by alienation in 1967 and socioeconomic status in 1966
- In each year, alienation is measured by observed variables measuring feelings of anomia and powerlessness
- Socioeconomic status is measured by education level and occupational status

---

The Alientation Model



Fitting a SEM Model

- Fit the model

```
. sem (Alien67 -> anomia67 pwless67) ///
      (Alien71 -> anomia71 pwless71) ///
      (SES -> educ66 occstat66) ///
      (Alien67 <- SES) ///
      (Alien71 <- Alien67 SES)

Endogenous variables

Measurement:  anomia67 pwless67 anomia71 pwless71 educ66 occstat66
Latent:       Alien67 Alien71

Exogenous variables

Latent:       SES

Fitting target model:

Iteration 0:   log likelihood = -15249.988
Iteration 1:   log likelihood = -15246.584
Iteration 2:   log likelihood = -15246.469
Iteration 3:   log likelihood = -15246.469

Structural equation model                       Number of obs     =        932
Estimation method  = ml
Log likelihood     = -15246.469

 ( 1)  [anomia67]Alien67 = 1
 ( 2)  [anomia71]Alien71 = 1
 ( 3)  [educ66]SES = 1
----------------------------------------------------------------------------------
                |                 OIM
```

```
                |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
Structural      |
  Alien67       |
            SES |  -.6140404   .0562407   -10.92   0.000    -.7242701   -.5038107
  --------------+----------------------------------------------------------------
  Alien71       |
        Alien67 |   .7046342   .0533512    13.21   0.000     .6000678    .8092007
            SES |  -.1744153   .0542489    -3.22   0.001    -.2807413   -.0680894
----------------+----------------------------------------------------------------
Measurement     |
  anomia67      |
        Alien67 |          1  (constrained)
          _cons |      13.61   .1126205   120.85   0.000     13.38927    13.83073
  --------------+----------------------------------------------------------------
  pwless67      |
        Alien67 |   .8884887   .0431565    20.59   0.000     .8039034    .9730739
          _cons |      14.67   .1001798   146.44   0.000     14.47365    14.86635
  --------------+----------------------------------------------------------------
  anomia71      |
        Alien71 |          1  (constrained)
          _cons |      14.13   .1158943   121.92   0.000     13.90285    14.35715
  --------------+----------------------------------------------------------------
  pwless71      |
        Alien71 |   .8486022   .0415205    20.44   0.000     .7672235    .9299808
          _cons |       14.9   .1034537   144.03   0.000     14.69723    15.10277
  --------------+----------------------------------------------------------------
  educ66        |
            SES |          1  (constrained)
          _cons |       10.9   .1014894   107.40   0.000     10.70108    11.09892
  --------------+----------------------------------------------------------------
  occstat66     |
            SES |   5.331259   .4307503    12.38   0.000     4.487004    6.175514
          _cons |      37.49   .6947112    53.96   0.000     36.12839    38.85161
----------------+----------------------------------------------------------------
 var(e.anomia67)|   4.009921   .3582978                      3.365724    4.777416
 var(e.pwless67)|   3.187468    .283374                      2.677762    3.794197
 var(e.anomia71)|   3.695593   .3911512                      3.003245     4.54755
 var(e.pwless71)|   3.621531   .3037908                      3.072483    4.268693
   var(e.educ66)|   2.943819   .5002527                      2.109908    4.107319
var(e.occstat66)|     260.63   18.24572                      227.2139    298.9605
  var(e.Alien67)|   5.301416    .483144                      4.434225    6.338201
  var(e.Alien71)|   3.737286   .3881546                      3.048951    4.581019
        var(SES)|    6.65587   .6409484                      5.511067    8.038482
---------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(6)   =     71.62, Prob > chi2 = 0.0000
```

Revising the Model

- On substantive grounds we could argue that covariances should be added between the errors for
  - ◇ Powerlessness in 1967 and 1971
  - ◇ Anomia in 1967 and 1971
- One method of evaluating whether adding these covariances to the model improves model fit is to perform a likelihood-ratio test

◇ In the SEM literature this is often called the χ^2 difference test

- We can use the `lrtest` command to perform this test
- First we will need to fit both models

The Likelihood-ratio Test

- We'll start by storing the estimates from the model we have already run

```
. estimates store nocov
```

- Now we can run the model with the covariances added

```
. sem (Alien67 -> anomia67 pwless67) ///
      (Alien71 -> anomia71 pwless71) ///
      (SES -> educ66 occstat66) ///
      (Alien67 <- SES) ///
      (Alien71 <- Alien67 SES), ///
          cov(e.anomia67*e.anomia71 ///
             e.pwless67*e.pwless71)

Endogenous variables

Measurement:  anomia67 pwless67 anomia71 pwless71 educ66 occstat66
Latent:       Alien67 Alien71

Exogenous variables

Latent:       SES

Fitting target model:

Iteration 0:   log likelihood = -15249.988
Iteration 1:   log likelihood =  -15217.95
Iteration 2:   log likelihood = -15213.126
Iteration 3:   log likelihood = -15213.046
Iteration 4:   log likelihood = -15213.046

Structural equation model                       Number of obs     =        932
Estimation method  = ml
Log likelihood     = -15213.046

 ( 1)  [anomia67]Alien67 = 1
 ( 2)  [anomia71]Alien71 = 1
 ( 3)  [educ66]SES = 1
------------------------------------------------------------------------------------------
-
                          |                OIM
                          |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval
> ]
--------------------------+---------------------------------------------------------------
-
Structural                |
  Alien67                 |
                      SES |  -.5752228    .057961    -9.92   0.000    -.6888244   -.461621
> 3
  ------------------------+---------------------------------------------------------------
-
  Alien71                 |
```

```
                  Alien67 |    .606954   .0512305    11.85   0.000     .5065439     .70736
> 4
                      SES |  -.2270301   .0530773    -4.28   0.000    -.3310596   -.123000
> 6
-------------------------+----------------------------------------------------------------
-
Measurement              |
  anomia67               |
                 Alien67 |          1  (constrained)
                   _cons |      13.61   .1126143   120.85   0.000     13.38928    13.8307
> 2
  -----------------------+----------------------------------------------------------------
-
  pwless67               |
                 Alien67 |   .9785952   .0619825    15.79   0.000     .8571117    1.10007
> 9
                   _cons |      14.67   .1001814   146.43   0.000     14.47365    14.8663
> 5
  -----------------------+----------------------------------------------------------------
-
  anomia71               |
                 Alien71 |          1  (constrained)
                   _cons |      14.13   .1159036   121.91   0.000     13.90283    14.3571
> 7
  -----------------------+----------------------------------------------------------------
-
  pwless71               |
                 Alien71 |   .9217508   .0597225    15.43   0.000     .8046968    1.03880
> 5
                   _cons |       14.9   .1034517   144.03   0.000     14.69724    15.1027
> 6
  -----------------------+----------------------------------------------------------------
-
  educ66                 |
                     SES |          1  (constrained)
                   _cons |       10.9   .1014894   107.40   0.000     10.70108    11.0989
> 2
  -----------------------+----------------------------------------------------------------
-
  occstat66              |
                     SES |    5.22132    .425595    12.27   0.000     4.387169    6.05547
> 1
                   _cons |      37.49   .6947112    53.96   0.000     36.12839    38.8516
> 1
-------------------------+----------------------------------------------------------------
-
         var(e.anomia67)|   4.728874    .456299                       3.914024    5.71336
> 5
         var(e.pwless67)|   2.563413   .4060733                       1.879225      3.496
> 7
         var(e.anomia71)|   4.396081   .5171156                       3.490904    5.53596
> 6
         var(e.pwless71)|   3.072085   .4360333                       2.326049    4.05739
> 8
           var(e.educ66)|   2.803674   .5115854                       1.960691    4.00909
> 1
        var(e.occstat66)|   264.5311   18.22483                       231.1177    302.775
> 1
          var(e.Alien67)|   4.842059   .4622537                       4.015771    5.83836
> 4
```

```
            var(e.Alien71)|   4.084249   .4038995                        3.364613    4.95780
> 2
                  var(SES)|   6.796014   .6524866                        5.630283    8.20310
> 5
--------------------------+-------------------------------------------------------------
-
cov(e.anomia67,e.anomia71)|   1.622024   .3154267     5.14   0.000     1.003799    2.24024
> 9
cov(e.pwless67,e.pwless71)|   .3399961   .2627541     1.29   0.196    -.1749925    .854984
> 7
----------------------------------------------------------------------------------------
-
LR test of model vs. saturated: chi2(4)   =      4.78, Prob > chi2 = 0.3111
```

- First we store the estimates

```
. estimates store withcov
```

The Likliehood-ratio Test (Continued)

- Then we can run the likelihood ratio test

```
. lrtest nocov withcov

Likelihood-ratio test                                 LR chi2(2)  =     66.85
(Assumption: nocov nested in withcov)                 Prob > chi2 =    0.0000
```

- The likelihood-ratio test indicates significantly better fit with the two covariances
- We could also have run only the model with the covariances and used the `test` command to perform a Wald test for joint significance of the covariances

---

2.4 Multiple Group Models

Comparing Groups

- Multiple group SEM allows for estimating parameters of a model separately for across groups
  - ◇ All parameters may be estimated separately, or
  - ◇ Some or all parameters can be constrained to equality across groups
- This allows us to examine whether parameters vary across groups
- We can use the `group()` option of `sem` to fit a model for two or more groups
- The `ginvariant()` option allows you to specify what parameters should be constrained across groups
  - ◇ By default measurement coefficients and measurement intercepts are constrained across groups

---

Multiple Group CFA

- To demonstrate we will return to the two-factor CFA model we fit earlier

```
. use hsdata, clear

(Data from Holzinger and Swineford (1939))
```

- The students in this dataset come from two different schools

```
. tab school

     school |      Freq.     Percent        Cum.
------------+-----------------------------------
    Pasteur |        156       51.83       51.83
Grant-White |        145       48.17      100.00
------------+-----------------------------------
      Total |        301      100.00
```

A Model with No Cross-group Constraints

- We will begin by estimating all parameters separately for each group to do this we will
  - ◇ Specify `ginvariant(none)`
  - ◇ Set the means of the latent variables to 0 using the `mean()` option
- Our command is

```
. sem (Spatial -> visual cubes paper) ///
      (Verbal -> paragraph sentence wordc), ///
      mean(Spatial@0 Verbal@0) ///
      group(school) ginvariant(none)

Endogenous variables

Measurement:  visual cubes paper paragraph sentence wordc

Exogenous variables

Latent:       Spatial Verbal

Fitting target model:

Iteration 0:   log likelihood = -5018.5723
Iteration 1:   log likelihood = -5017.9684
Iteration 2:   log likelihood = -5017.9585
Iteration 3:   log likelihood = -5017.9585

Structural equation model                       Number of obs     =        301
Grouping variable  = school                     Number of groups  =          2
Estimation method  = ml
Log likelihood     = -5017.9585

 ( 1)  [visual]1bn.school#c.Spatial = 1
 ( 2)  [paragraph]1bn.school#c.Verbal = 1
 ( 3)  [/]mean(Spatial)#1bn.school = 0
 ( 4)  [/]mean(Verbal)#1bn.school = 0
 ( 5)  [visual]2.school#c.Spatial = 1
 ( 6)  [paragraph]2.school#c.Verbal = 1
 ( 7)  [/]mean(Spatial)#2.school = 0
```

```
 ( 8)  [/]mean(Verbal)#2.school = 0

Group              : Pasteur                   Number of obs     =        156

------------------------------------------------------------------------------------
                   |                 OIM
                   |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
Measurement        |
  visual           |
           Spatial |          1  (constrained)
             _cons |   29.64744   .5674292    52.25   0.000      28.5353    30.75958
  -----------------+----------------------------------------------------------------
  cubes            |
           Spatial |   .2712034   .1027351     2.64   0.008     .0698463    .4725605
             _cons |    23.9359   .3927222    60.95   0.000     23.16618    24.70562
  -----------------+----------------------------------------------------------------
  paper            |
           Spatial |   .1973023   .0641721     3.07   0.002     .0715272    .3230774
             _cons |   14.16026    .227089    62.36   0.000     13.71517    14.60534
  -----------------+----------------------------------------------------------------
  paragraph        |
            Verbal |          1  (constrained)
             _cons |   8.467949   .2759271    30.69   0.000     7.927142    9.008756
  -----------------+----------------------------------------------------------------
  sentence         |
            Verbal |   1.574508   .1457788    10.80   0.000     1.288787    1.860229
             _cons |   15.98077   .4185334    38.18   0.000     15.16046    16.80108
  -----------------+----------------------------------------------------------------
  wordc            |
            Verbal |   1.365996   .1459215     9.36   0.000     1.079996    1.651997
             _cons |   24.19872   .4216584    57.39   0.000     23.37228    25.02515
-------------------+----------------------------------------------------------------
     mean(Spatial)|          0  (constrained)
      mean(Verbal)|          0  (constrained)
-------------------+----------------------------------------------------------------
     var(e.visual)|   13.00047   10.01906                         2.870547    58.87808
      var(e.cubes)|   21.32184   2.625654                          16.7496    27.14219
      var(e.paper)|   6.595619   .8718128                         5.090299    8.546098
  var(e.paragraph)|   3.715095   .6950598                         2.574649    5.360704
   var(e.sentence)|   7.092144   1.564653                         4.602432    10.92868
      var(e.wordc)|   12.50614   1.783697                         9.456279    16.53966
      var(Spatial)|   37.22777   11.33104                         20.50153    67.60016
       var(Verbal)|   8.162082   1.392037                         5.842909    11.40178
-------------------+----------------------------------------------------------------
cov(Spatial,Verbal)|   8.594239   2.055358     4.18   0.000     4.565812    12.62267
------------------------------------------------------------------------------------

Group              : Grant-White               Number of obs     =        145

------------------------------------------------------------------------------------
                   |                 OIM
                   |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
Measurement        |
  visual           |
           Spatial |          1  (constrained)
             _cons |   29.57931   .5721785    51.70   0.000     28.45786    30.70076
  -----------------+----------------------------------------------------------------
  cubes            |
```

```
         Spatial |   .3769229   .1031886     3.65   0.000     .1746769    .5791688
           _cons |       24.8   .3678649    67.42   0.000       24.079      25.521
  ---------------+----------------------------------------------------------------
  paper          |
         Spatial |   .3004087   .0787246     3.82   0.000     .1461113    .4547062
           _cons |   14.30345   .2335324    61.25   0.000     13.84573    14.76116
  ---------------+----------------------------------------------------------------
  paragraph      |
          Verbal |          1  (constrained)
           _cons |   9.951724   .2793454    35.63   0.000     9.404217    10.49923
  ---------------+----------------------------------------------------------------
  sentence       |
          Verbal |   1.550538   .1540779    10.06   0.000     1.248551    1.852525
           _cons |   18.84828   .3847671    48.99   0.000     18.09415    19.60241
  ---------------+----------------------------------------------------------------
  wordc          |
          Verbal |   1.392518   .1663741     8.37   0.000     1.066431    1.718605
           _cons |       28.2   .4433791    63.60   0.000     27.33099    29.06901
-----------------+----------------------------------------------------------------
    mean(Spatial)|          0  (constrained)
     mean(Verbal)|          0  (constrained)
-----------------+----------------------------------------------------------------
    var(e.visual)|   23.07059   6.254186                      13.56149    39.24734
     var(e.cubes)|   16.15544   2.130434                      12.47585    20.92029
     var(e.paper)|   5.705868   .8699284                      4.232001    7.693034
 var(e.paragraph)|   4.150919   .7114099                      2.966608    5.808024
  var(e.sentence)|   4.243196   1.346801                      2.277821    7.904358
     var(e.wordc)|   14.61309   2.026313                      11.13553    19.17667
     var(Spatial)|    24.4007   7.450847                      13.41173    44.39356
      var(Verbal)|   7.163992   1.340404                      4.964697    10.33754
-----------------+----------------------------------------------------------------
cov(Spatial,Verbal)|   7.303571   1.806168     4.04   0.000     3.763546     10.8436
----------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(16)  =     27.86, Prob > chi2 = 0.0328
```

Testing for Group Differences

- We could use a likelihood-ratio test to see whether the parameters vary across schools
- An alternative is to use `estat ginvariant` instead
  - ◇ Wald tests are used to test whether *unconstrained* coefficients are significantly different
  - ◇ Score tests are used to test whether *relaxing constraints* would improve model fit
  - ◇ In both cases the null hypothesis is that the constraint does not harm model fit
- Let's give it a try

```
. estat ginvariant

Tests for group invariance of parameters

----------------------------------------------------------------------------------
                 |          Wald Test                       Score Test
                 |      chi2          df    p>chi2       chi2            df    p>chi2
-----------------+----------------------------------------------------------------
Measurement      |
  visual         |
         Spatial |         .           .         .          .             0         .
```

```
                _cons |      0.007          1    0.9326           .            .          .
-----------------+------------------------------------------------------------------
cubes            |
          Spatial |      0.527          1    0.4678           .            .          .
            _cons |      2.579          1    0.1083           .            .          .
-----------------+------------------------------------------------------------------
paper            |
          Spatial |      1.031          1    0.3100           .            .          .
            _cons |      0.193          1    0.6602           .            .          .
-----------------+------------------------------------------------------------------
paragraph        |
           Verbal |          .          .         .           .            0          .
            _cons |     14.280          1    0.0002           .            .          .
-----------------+------------------------------------------------------------------
sentence         |
           Verbal |      0.013          1    0.9100           .            .          .
            _cons |     25.440          1    0.0000           .            .          .
-----------------+------------------------------------------------------------------
wordc            |
           Verbal |      0.014          1    0.9046           .            .          .
            _cons |     42.765          1    0.0000           .            .          .
-------------------+------------------------------------------------------------------
      mean(Spatial)|          .          .         .           .            0          .
       mean(Verbal)|          .          .         .           .            0          .
-------------------+------------------------------------------------------------------
     var(e.visual)|      0.727          1    0.3939           .            .          .
      var(e.cubes)|      2.335          1    0.1265           .            .          .
      var(e.paper)|      0.522          1    0.4700           .            .          .
  var(e.paragraph)|      0.192          1    0.6612           .            .          .
   var(e.sentence)|      1.904          1    0.1676           .            .          .
      var(e.wordc)|      0.609          1    0.4351           .            .          .
      var(Spatial)|      0.895          1    0.3442           .            .          .
       var(Verbal)|      0.267          1    0.6055           .            .          .
-------------------+------------------------------------------------------------------
cov(Spatial,Verbal)|      0.223          1    0.6371           .            .          .
-------------------------------------------------------------------------------------
```

- We fail to reject the null hypothesis that the measurement coefficients are the same across groups

---

Constraining Coefficients

- We could refit the model constraining measurement coefficients to equality across gourps using the option `ginvariant(mcoef)`
- Then we could use `estat ginvariant` to test for equality of the intercepts across groups
- A typical ordering of tests is
  - Measurement coefficients
  - Measurement intercepts
  - Measurement error variances

---

3 Generalized SEM

3.1 Introduction

Generalized Structure Equation Models

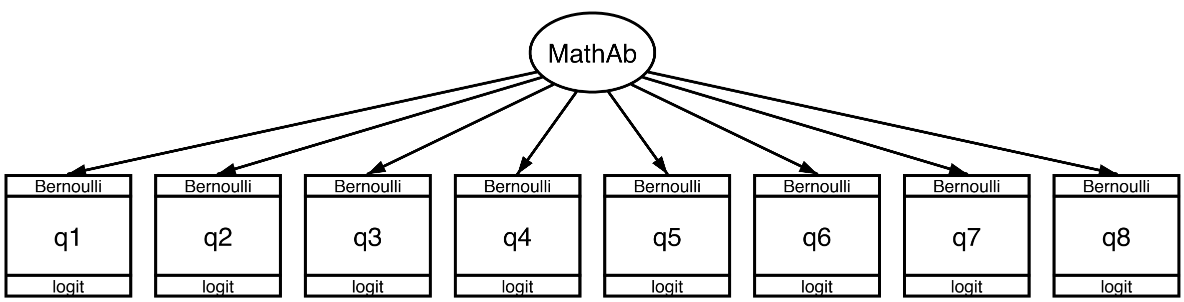
- gsem allows us to extend the types of models we can fit
- Models for binary, ordered, nominal, count, survival time, interval, and censored responses.
- Multilevel models, including models with random intercepts and slopes, for nested or crossed data
- Latent variables can be included at any level of the model

3.2 Models for Generalized Responses

CFA with Binary Indicators

- Many of the models we fit above can be extended to include generalized response variables
- In this example we will fit a confirmatory factor analysis model using binary indicators
- The dataset contains fictional data on students' math scores and attitudes towards math
- The binary indicators are 8 questions from a math test
- The latent variable is math ability

Path Model for a Generalized CFA



Fitting a Generalized CFA

- Let's begin by opening the dataset and looking at the items q1–q8

```
. use math
. codebook q1-q8

(Fictional math abilities data)

-------------------------------------------------------------------------------------------
q1                                                                               q1 correct
-------------------------------------------------------------------------------------------

                  type:  numeric (byte)
                 label:  result

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/500
```

```
          tabulation:  Freq.   Numeric  Label
                         247         0  Incorrect
                         253         1  Correct

------------------------------------------------------------------------------------------
q2                                                                              q2 correct
------------------------------------------------------------------------------------------

                type:  numeric (byte)
               label:  result

               range:  [0,1]                        units:  1
       unique values:  2                        missing .:  0/500

          tabulation:  Freq.   Numeric  Label
                         303         0  Incorrect
                         197         1  Correct

------------------------------------------------------------------------------------------
q3                                                                              q3 correct
------------------------------------------------------------------------------------------

                type:  numeric (byte)
               label:  result

               range:  [0,1]                        units:  1
       unique values:  2                        missing .:  0/500

          tabulation:  Freq.   Numeric  Label
                         233         0  Incorrect
                         267         1  Correct

------------------------------------------------------------------------------------------
q4                                                                              q4 correct
------------------------------------------------------------------------------------------

                type:  numeric (byte)
               label:  result

               range:  [0,1]                        units:  1
       unique values:  2                        missing .:  0/500

          tabulation:  Freq.   Numeric  Label
                         288         0  Incorrect
                         212         1  Correct

------------------------------------------------------------------------------------------
q5                                                                              q5 correct
------------------------------------------------------------------------------------------

                type:  numeric (byte)
               label:  result

               range:  [0,1]                        units:  1
       unique values:  2                        missing .:  0/500

          tabulation:  Freq.   Numeric  Label
                         255         0  Incorrect
                         245         1  Correct
```

```
----------------------------------------------------------------------------------------------
q6                                                                                  q6 correct
----------------------------------------------------------------------------------------------

                  type:  numeric (byte)
                 label:  result

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/500

            tabulation:  Freq.   Numeric  Label
                           283         0  Incorrect
                           217         1  Correct

----------------------------------------------------------------------------------------------
q7                                                                                  q7 correct
----------------------------------------------------------------------------------------------

                  type:  numeric (byte)
                 label:  result

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/500

            tabulation:  Freq.   Numeric  Label
                           240         0  Incorrect
                           260         1  Correct

----------------------------------------------------------------------------------------------
q8                                                                                  q8 correct
----------------------------------------------------------------------------------------------

                  type:  numeric (byte)
                 label:  result

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/500

            tabulation:  Freq.   Numeric  Label
                           253         0  Incorrect
                           247         1  Correct
```

- There are only a few changes to the command
  - We will use the `gsem` command to fit this model
  - Specify the `logit` option to fit a logit model to our binary response variables `q1-q8`
  - We'll use the `nodvheader`, otherwise `gsem` will list the family and link function for each dependent variable

```
. gsem (MathAb -> q1-q8, logit), nodvheader

Fitting fixed-effects model:

Iteration 0:   log likelihood = -2750.3114
Iteration 1:   log likelihood = -2749.3709
Iteration 2:   log likelihood = -2749.3708

Refining starting values:

Grid node 0:   log likelihood = -2645.8536
```

```
Fitting full model:

Iteration 0:   log likelihood = -2645.8536
Iteration 1:   log likelihood =  -2638.477
Iteration 2:   log likelihood = -2637.6526
Iteration 3:   log likelihood = -2637.3803
Iteration 4:   log likelihood =  -2637.376
Iteration 5:   log likelihood = -2637.3759

Generalized structural equation model           Number of obs     =        500
Log likelihood = -2637.3759

 ( 1)  [q1]MathAb = 1
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
q1           |
      MathAb |          1  (constrained)
       _cons |   .0373365   .1252279     0.30   0.766    -.2081058    .2827787
-------------+----------------------------------------------------------------
q2           |
      MathAb |    .381626    .116809     3.27   0.001     .1526845    .6105674
       _cons |  -.4613391   .0989722    -4.66   0.000     -.655321   -.2673571
-------------+----------------------------------------------------------------
q3           |
      MathAb |   .4993762    .134314     3.72   0.000     .2361255    .7626269
       _cons |   .1533362   .1006072     1.52   0.127    -.0438503    .3505228
-------------+----------------------------------------------------------------
q4           |
      MathAb |   .3299698   .1063034     3.10   0.002     .1216189    .5383207
       _cons |  -.3230667   .0957983    -3.37   0.001     -.510828   -.1353054
-------------+----------------------------------------------------------------
q5           |
      MathAb |   .8401762   .1995336     4.21   0.000     .4490975    1.231255
       _cons |  -.0494684   .1163093    -0.43   0.671    -.2774304    .1784937
-------------+----------------------------------------------------------------
q6           |
      MathAb |   .6453722   .1639865     3.94   0.000     .3239646    .9667798
       _cons |   -.314723   .1083049    -2.91   0.004    -.5269968   -.1024493
-------------+----------------------------------------------------------------
q7           |
      MathAb |   .8163613   .2045448     3.99   0.000     .4154609    1.217262
       _cons |   .1053404   .1152979     0.91   0.361    -.1206393    .3313201
-------------+----------------------------------------------------------------
q8           |
      MathAb |   .5769516   .1473524     3.92   0.000     .2881463     .865757
       _cons |   -.026705   .1034396    -0.26   0.796    -.2294429    .1760328
-------------+----------------------------------------------------------------
  var(MathAb)|   2.151059   .7298407                      1.106229    4.182728
------------------------------------------------------------------------------
```

- If we include certain constraints on this model, it can be interpreted as an item response theory (IRT) model
  - See `help irt` for information on Stata's `irt` commands

A Generalized Structural Equation Model

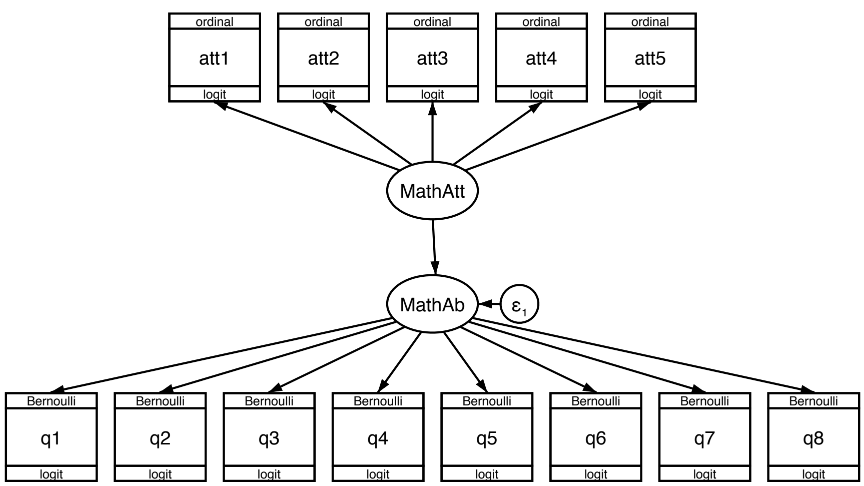
- The dataset also includes information on student's attitudes towards math, we may want to see if these predict math ability
- Let's look more closely at these items

```
. codebook att*

------------------------------------------------------------------------------------------
att1                            Skills taught in math class will help me get a better job.
------------------------------------------------------------------------------------------

                  type:  numeric (float)
                 label:  agree

                 range:  [1,5]                        units:  1
         unique values:  5                        missing .:  0/500

            tabulation:  Freq.   Numeric  Label
                           150         1  Strongly disagree
                            78         2  Disagree
                            52         3  Neither agree nor disagree
                            89         4  Agree
                           131         5  Strongly agree

------------------------------------------------------------------------------------------
att2                                                     Math is important in everyday life
------------------------------------------------------------------------------------------

                  type:  numeric (float)
                 label:  agree

                 range:  [1,5]                        units:  1
         unique values:  5                        missing .:  0/500

            tabulation:  Freq.   Numeric  Label
                           134         1  Strongly disagree
                            93         2  Disagree
                            65         3  Neither agree nor disagree
                            81         4  Agree
                           127         5  Strongly agree

------------------------------------------------------------------------------------------
att3                                                Working math problems makes me anxious.
------------------------------------------------------------------------------------------

                  type:  numeric (float)
                 label:  agree

                 range:  [1,5]                        units:  1
         unique values:  5                        missing .:  0/500

            tabulation:  Freq.   Numeric  Label
                           171         1  Strongly disagree
                            75         2  Disagree
                            47         3  Neither agree nor disagree
                            77         4  Agree
                           130         5  Strongly agree

------------------------------------------------------------------------------------------
att4                                                 Math has always been my worst subject.
------------------------------------------------------------------------------------------
```

```
                  type:  numeric (float)
                 label:  agree

                 range:  [1,5]                        units:  1
         unique values:  5                        missing .:  0/500

            tabulation:  Freq.   Numeric  Label
                           145         1  Strongly disagree
                            83         2  Disagree
                            63         3  Neither agree nor disagree
                            90         4  Agree
                           119         5  Strongly agree

--------------------------------------------------------------------------------------------
att5                                          I am able to learn new math concepts easily.
--------------------------------------------------------------------------------------------

                  type:  numeric (float)
                 label:  agree

                 range:  [1,5]                        units:  1
         unique values:  5                        missing .:  0/500

            tabulation:  Freq.   Numeric  Label
                           121         1  Strongly disagree
                            91         2  Disagree
                            62         3  Neither agree nor disagree
                            76         4  Agree
                           150         5  Strongly agree
```

- The math attitude items appear to be Likert-type items

Path Diagram for the GSEM



Fitting a GSEM

- We will use the ologit option to model the responses att1-att5 using an ordered logistic model

```
. gsem (MathAb -> q1-q8, logit) ///
       (MathAtt -> att1-att5, ologit) ///
       (MathAtt -> MathAb), nodvheader

Fitting fixed-effects model:

Iteration 0:   log likelihood = -6629.7253
Iteration 1:   log likelihood = -6628.7848
Iteration 2:   log likelihood = -6628.7848

Refining starting values:

Grid node 0:   log likelihood = -6429.1636

Fitting full model:

Iteration 0:   log likelihood = -6429.1636
Iteration 1:   log likelihood = -6396.7471
Iteration 2:   log likelihood = -6394.6197
Iteration 3:   log likelihood = -6394.3949
Iteration 4:   log likelihood = -6394.3923
Iteration 5:   log likelihood = -6394.3923

Generalized structural equation model           Number of obs     =        500
Log likelihood = -6394.3923

 ( 1)  [q1]MathAb = 1
 ( 2)  [att1]MathAtt = 1
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
q1           |
      MathAb |          1  (constrained)
       _cons |    .044612   .1272967     0.35   0.726     -.204885     .294109
-------------+----------------------------------------------------------------
q2           |
      MathAb |   .3446066   .1050261     3.28   0.001     .1387593     .550454
       _cons |  -.4572215   .0979965    -4.67   0.000    -.6492911   -.2651519
-------------+----------------------------------------------------------------
q3           |
      MathAb |   .5445222   .1386992     3.93   0.000     .2726767    .8163677
       _cons |   .1591406   .1033116     1.54   0.123    -.0433465    .3616276
-------------+----------------------------------------------------------------
q4           |
      MathAb |   .2858862   .0948549     3.01   0.003      .099974    .4717984
       _cons |  -.3196648   .0947684    -3.37   0.001    -.5054075   -.1339222
-------------+----------------------------------------------------------------
q5           |
      MathAb |   .8174769   .1867022     4.38   0.000     .4515473    1.183406
       _cons |    -.04543    .116575    -0.39   0.697    -.2739129    .1830528
-------------+----------------------------------------------------------------
q6           |
      MathAb |   .6030423   .1471949     4.10   0.000     .3145457    .8915389
       _cons |  -.3099919   .1070853    -2.89   0.004    -.5198754   -.1001085
-------------+----------------------------------------------------------------
q7           |
      MathAb |   .7208369    .171309     4.21   0.000     .3850774    1.056597
```

```
       _cons |   .1047264   .1116494     0.94   0.348    -.1141024    .3235553
-------------+----------------------------------------------------------------
q8           |
      MathAb |   .5814736   .1426725     4.08   0.000     .3018406    .8611067
       _cons |  -.0250443   .1045135    -0.24   0.811    -.2298869    .1797984
-------------+----------------------------------------------------------------
att1         |
     MathAtt |          1  (constrained)
-------------+----------------------------------------------------------------
att2         |
     MathAtt |   .3788715   .0971234     3.90   0.000     .1885131    .5692299
-------------+----------------------------------------------------------------
att3         |
     MathAtt |  -1.592717   .3614956    -4.41   0.000    -2.301236   -.8841989
-------------+----------------------------------------------------------------
att4         |
     MathAtt |  -.8100108   .1530675    -5.29   0.000    -1.110017   -.510004
-------------+----------------------------------------------------------------
att5         |
     MathAtt |   .5225425   .1170166     4.47   0.000     .2931942    .7518907
-------------+----------------------------------------------------------------
MathAb       |
     MathAtt |    .581103     .14776     3.93   0.000     .2914987    .8707072
-------------+----------------------------------------------------------------
/att1        |
        cut1 |   -1.10254    .131228                     -1.359742   -.8453377
        cut2 |  -.2495339   .1160385                     -.4769652   -.0221025
        cut3 |   .2983261   .1164415                       .070105    .5265472
        cut4 |   1.333052   .1391919                      1.060241    1.605864
-------------+----------------------------------------------------------------
/att2        |
        cut1 |  -1.055791   .1062977                     -1.264131   -.8474513
        cut2 |  -.1941211   .0941435                      -.378639   -.0096032
        cut3 |   .3598488   .0952038                      .1732528    .5464448
        cut4 |   1.132624   .1082204                      .9205156    1.344732
-------------+----------------------------------------------------------------
/att3        |
        cut1 |  -1.053519   .1734001                     -1.393377   -.7136612
        cut2 |  -.0491074   .1442846                        -.3319    .2336853
        cut3 |   .5570672   .1538702                      .2554871    .8586472
        cut4 |   1.666859   .2135557                      1.248297     2.08542
-------------+----------------------------------------------------------------
/att4        |
        cut1 |   -1.07378   .1214071                     -1.311734   -.8358264
        cut2 |  -.2112462   .1076501                     -.4222365   -.0002559
        cut3 |    .406347   .1094847                       .191761     .620933
        cut4 |   1.398185   .1313327                      1.140778    1.655593
-------------+----------------------------------------------------------------
/att5        |
        cut1 |  -1.244051   .1148443                     -1.469142   -1.018961
        cut2 |   -.336135   .0986678                     -.5295203   -.1427498
        cut3 |   .2137776   .0978943                      .0219084    .4056468
        cut4 |   .9286849    .107172                      .7186316    1.138738
-------------+----------------------------------------------------------------
 var(e.MathAb)|   1.787117   .5974753                      .9280606    3.441357
  var(MathAtt)|   1.520854   .4077885                      .8991947    2.572298
------------------------------------------------------------------------------
```

3.3 Multilevel Models

Multilevel SEM

- Many of the types of structural equation models that we have discussed can be extended to multilevel models using `gsem`
- Because `gsem` can be used to include random effects and model generalized responses a large number of models can be fit
  - ◇ Including a multilevel multinomial logit model that cannot be fit elsewhere

---

Multilevel CFA

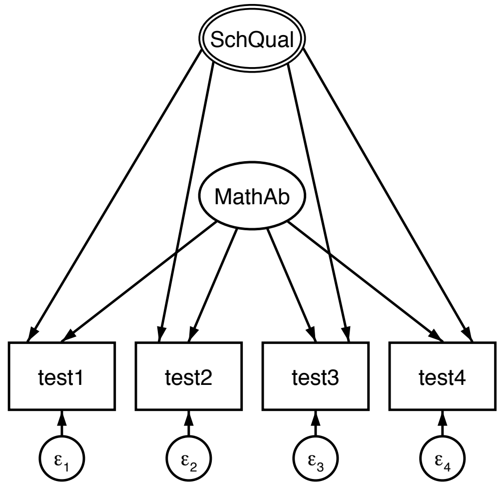
- We will look at an example of a multilevel CFA
- We will continue using the the same dataset
- The students are clustered within schools
- This time we will measure the latent variable `MathAb` using test scores, let's take a look

```
. codebook school test*

------------------------------------------------------------------------------------------
school                                                                           School id
------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [1,20]                       units:  1
         unique values:  20                       missing .:  0/500

                  mean:      10.5
              std. dev:   5.77206

           percentiles:        10%       25%       50%       75%       90%
                               2.5       5.5      10.5      15.5      18.5

------------------------------------------------------------------------------------------
test1                                                                   Score, math test 1
------------------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [55,93]                      units:  1
         unique values:  36                       missing .:  0/500

                  mean:    75.548
              std. dev:   5.94865

           percentiles:        10%       25%       50%       75%       90%
                                68        72        76        79        83

------------------------------------------------------------------------------------------
test2                                                                   Score, math test 2
------------------------------------------------------------------------------------------

                  type:  numeric (byte)
```

```
           range:  [65,94]                      units:  1
   unique values:  28                       missing .:  0/500

            mean:    80.556
        std. dev:   4.97679

     percentiles:        10%       25%       50%       75%       90%
                          74        77        80        84        88

------------------------------------------------------------------------------------------
test3                                                                   Score, math test 3
------------------------------------------------------------------------------------------

            type:  numeric (byte)

           range:  [50,94]                      units:  1
   unique values:  36                       missing .:  0/500

            mean:    75.572
        std. dev:   6.67787

     percentiles:        10%       25%       50%       75%       90%
                          67      71.5        76        80        84

------------------------------------------------------------------------------------------
test4                                                                   Score, math test 4
------------------------------------------------------------------------------------------

            type:  numeric (byte)

           range:  [43,96]                      units:  1
   unique values:  48                       missing .:  0/500

            mean:    74.078
        std. dev:   8.84559

     percentiles:        10%       25%       50%       75%       90%
                          63        69        74        80        86
```

Path Diagram for a Multilevel CFA

- Random effects are denoted as ovals with double rings
- Graphically the model is



Fitting a Multilevel CFA

- Here the test items are predicted by `MathAb` and the random intercept denoted `SchQual`
- The square brackets around school indicate that `SchQual` is constant within school and varies across schools
- Run the model

```
. gsem (MathAb SchQual[school] -> test1 test2 test3 test4)

Fitting fixed-effects model:

Iteration 0:   log likelihood = -6569.2088
Iteration 1:   log likelihood = -6569.2088

Refining starting values:

Grid node 0:   log likelihood = -5394.8535

Fitting full model:

Iteration 0:   log likelihood = -5394.8535  (not concave)
Iteration 1:   log likelihood = -5391.8634
Iteration 2:   log likelihood =  -5386.954
Iteration 3:   log likelihood =  -5386.132
Iteration 4:   log likelihood =  -5386.112
Iteration 5:   log likelihood = -5386.1119

Generalized structural equation model           Number of obs     =        500

Response       : test1
Family         : Gaussian
Link           : identity

Response       : test2
Family         : Gaussian
Link           : identity

Response       : test3
Family         : Gaussian
```

```
Link           : identity

Response       : test4
Family         : Gaussian
Link           : identity

Log likelihood = -5386.1119

 ( 1)  [test1]SchQual[school] = 1
 ( 2)  [test2]MathAb = 1
-------------------------------------------------------------------------------------
                     |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+---------------------------------------------------------------
test1                |
     SchQual[school] |          1  (constrained)
                     |
              MathAb |   1.505647   .0781211    19.27   0.000     1.352532    1.658761
               _cons |   75.97363   .4239572   179.20   0.000     75.14269    76.80457
---------------------+---------------------------------------------------------------
test2                |
     SchQual[school] |   .2535074   .1413808     1.79   0.073     -.023594    .5306087
                     |
              MathAb |          1  (constrained)
               _cons |   80.83869   .2262333   357.32   0.000     80.39528     81.2821
---------------------+---------------------------------------------------------------
test3                |
     SchQual[school] |   .8253025   .0975692     8.46   0.000     .6340704    1.016535
                     |
              MathAb |    1.76592   .0900749    19.61   0.000     1.589377    1.942464
               _cons |   76.07121   .3911863   194.46   0.000      75.3045    76.83792
---------------------+---------------------------------------------------------------
test4                |
     SchQual[school] |   1.352783   .0963211    14.04   0.000     1.163997    1.541569
                     |
              MathAb |   2.394628   .1180165    20.29   0.000      2.16332    2.625936
               _cons |   74.75494   .5875971   127.22   0.000     73.60327    75.90661
---------------------+---------------------------------------------------------------
 var(SchQual[school])|   2.637378   1.196124                      1.084248    6.415285
          var(MathAb)|   12.00398   1.401103                      9.549339    15.08959
---------------------+---------------------------------------------------------------
         var(e.test1)|   3.980296   .2902873                      3.450137    4.591921
         var(e.test2)|   12.46348   .8074865                       10.9772      14.151
         var(e.test3)|   4.087937   .3289069                       3.49155    4.786192
         var(e.test4)|   1.465088   .3576202                      .9080087    2.363946
-------------------------------------------------------------------------------------
```

4 Conclusion

4.1 Conclusion

Conclusion

- We have learned a bit about structural equation models and generalized structural equation models
- We have seen how to use `sem` to fit linear SEM models
- We have also seen how `gsem` can be used to fit more general models
- We have also touched on the flexibility of `gsem`

Index