Example 30g — Two-level measurement model (multilevel, generalized response)

Description Remarks and examples References Also see

Description

We demonstrate a multilevel measurement model with the same data used in [SEM] **Example 29g**:

```
. use https://www.stata-press.com/data/r18/gsem_cfa
(Fictional math abilities data)
. summarize
```

Variable	Obs	Mean	Std. dev.	Min	Max
school	500	10.5	5.772056	1	20
id	500	50681.71	29081.41	71	100000
q1	500	.506	.5004647	0	1
q2	500	.394	.4891242	0	1
q3	500	.534	.4993423	0	1
q4	500	.424	.4946852	0	1
q5	500	.49	.5004006	0	1
q6	500	.434	.4961212	0	1
q7	500	.52	.5001002	0	1
q8	500	.494	.5004647	0	1
att1	500	2.946	1.607561	1	5
att2	500	2.948	1.561465	1	5
att3	500	2.84	1.640666	1	5
att4	500	2.91	1.566783	1	5
att5	500	3.086	1.581013	1	5
test1	500	75.548	5.948653	55	93
test2	500	80.556	4.976786	65	94
test3	500	75.572	6.677874	50	94
test4	500	74.078	8.845587	43	96

```
. notes
_dta:
  1.  Fictional data on math ability and attitudes of 500 students from 20
      schools.
  2.  Variables q1-q8 are incorrect/correct (0/1) on individual math questions.
  3.  Variables att1-att5 are items from a Likert scale measuring each
      student's attitude toward math.
  4.  Variables test1-test4 are test scores from tests of four different
      aspects of mathematical abilities.  Range of scores: 0-100.
```

These data record results from a fictional instrument measuring mathematical ability. Variables q1 through q8 are the items from the instrument.

For discussions of multilevel measurement models, including extensions beyond the example we present here, see Mehta and Neale (2005) and Skrondal and Rabe-Hesketh (2004).

See *Single-factor measurement models* and *Multilevel mixed-effects models* in [SEM] **Intro 5** for background.

Remarks and examples

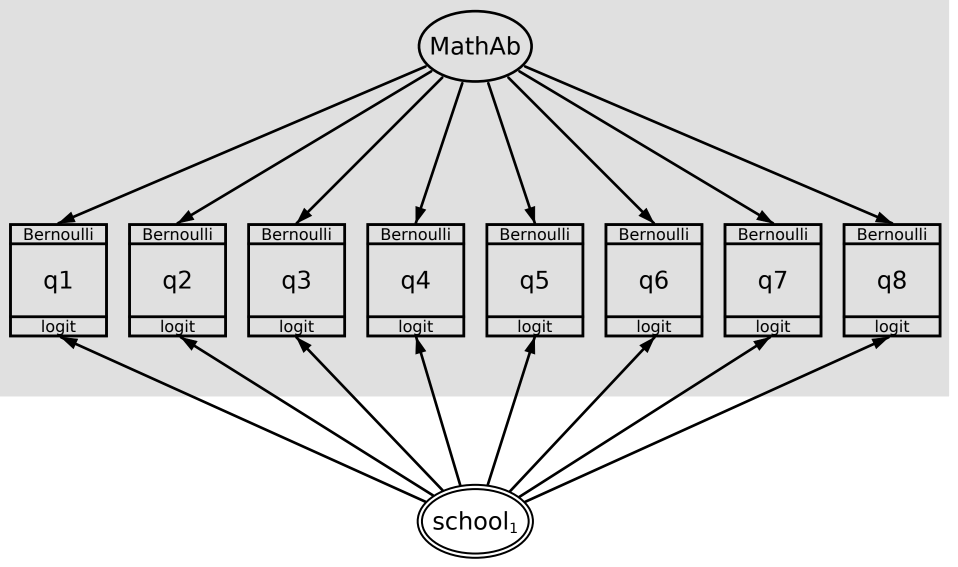
stata.com

Remarks are presented under the following headings:

Fitting the two-level model

We extend the measurement model fit in [SEM] **Example 29g** to better account for our (fictional) data. In the data, students are nested within school, but we have ignored that so far. In this example, we include a latent variable at the school level to account for possible school-by-school effects.

The model we wish to fit is



The double-ringed $\texttt{school}_1$ is new. That new component of the path diagram is saying, "I am a latent variable at the `school` level—meaning I am constant within school and vary across schools—and I correspond to a latent variable named `M1`"; see *Specifying generalized SEMs: Multilevel mixed effects (2 levels)* in [SEM] **Intro 2**. This new variable will account for the effect, if any, of the identity of the school.

To fit this model without this new, school-level component in it, we would type

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

To include the new school-level component, we add `M1[school]` to the exogenous variables:

```
. gsem (MathAb M1[school] -> q1-q8), logit
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 = -2649.0033
Fitting full model:
Iteration 0:  Log likelihood = -2649.0033  (not concave)
Iteration 1:  Log likelihood = -2645.0613  (not concave)
Iteration 2:  Log likelihood = -2641.9755  (not concave)
Iteration 3:  Log likelihood = -2634.3857
Iteration 4:  Log likelihood = -2631.1111
Iteration 5:  Log likelihood = -2630.7898
Iteration 6:  Log likelihood = -2630.2477
Iteration 7:  Log likelihood = -2630.2402
Iteration 8:  Log likelihood = -2630.2074
Iteration 9:  Log likelihood = -2630.2063
Iteration 10: Log likelihood = -2630.2063
Generalized structural equation model                   Number of obs = 500
Response: q1
Family:   Bernoulli
Link:     Logit
Response: q2
Family:   Bernoulli
Link:     Logit
Response: q3
Family:   Bernoulli
Link:     Logit
Response: q4
Family:   Bernoulli
Link:     Logit
Response: q5
Family:   Bernoulli
Link:     Logit
Response: q6
Family:   Bernoulli
Link:     Logit
Response: q7
Family:   Bernoulli
Link:     Logit
Response: q8
Family:   Bernoulli
Link:     Logit
Log likelihood = -2630.2063
 ( 1)  [q1]M1[school] = 1
 ( 2)  [q2]MathAb = 1
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
q1						
M1[school]	1	(constrained)				
MathAb	2.807515	.9468682	2.97	0.003	.9516878	4.663343
_cons	.0388021	.1608489	0.24	0.809	-.276456	.3540602
q2						
M1[school]	.6673925	.3058328	2.18	0.029	.0679712	1.266814
MathAb	1	(constrained)				
_cons	-.4631159	.1201227	-3.86	0.000	-.698552	-.2276798
q3						
M1[school]	.3555867	.3043548	1.17	0.243	-.2409377	.9521111
MathAb	1.455529	.5187786	2.81	0.005	.4387416	2.472316
_cons	.1537831	.1070288	1.44	0.151	-.0559894	.3635556
q4						
M1[school]	.7073241	.3419273	2.07	0.039	.037159	1.377489
MathAb	.8420897	.3528195	2.39	0.017	.1505762	1.533603
_cons	-.3252735	.1202088	-2.71	0.007	-.5608784	-.0896686
q5						
M1[school]	.7295553	.3330652	2.19	0.028	.0767595	1.382351
MathAb	2.399529	.8110973	2.96	0.003	.8098079	3.989251
_cons	-.0488674	.1378015	-0.35	0.723	-.3189533	.2212185
q6						
M1[school]	.484903	.2844447	1.70	0.088	-.0725983	1.042404
MathAb	1.840627	.5934017	3.10	0.002	.6775813	3.003673
_cons	-.3139302	.1186624	-2.65	0.008	-.5465042	-.0813563
q7						
M1[school]	.3677241	.2735779	1.34	0.179	-.1684787	.903927
MathAb	2.444023	.8016872	3.05	0.002	.8727449	4.015301
_cons	.1062164	.1220796	0.87	0.384	-.1330552	.3454881
q8						
M1[school]	.5851299	.3449508	1.70	0.090	-.0909612	1.261221
MathAb	1.606287	.5367614	2.99	0.003	.5542541	2.65832
_cons	-.0261962	.1189835	-0.22	0.826	-.2593995	.2070071
var(M1[school])	.2121216	.1510032			.052558	.8561121
var(MathAb)	.2461246	.1372513			.0825055	.7342217

Notes:

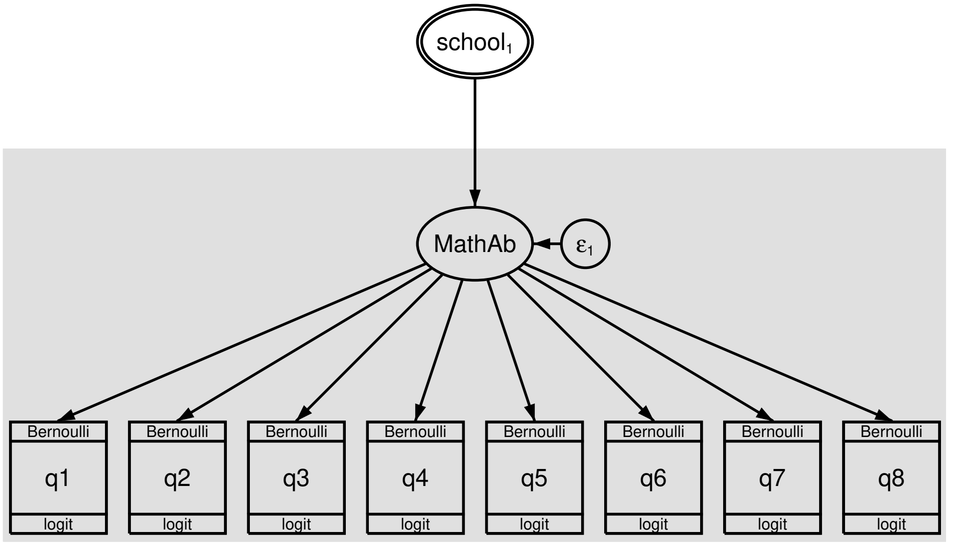
1. The variance of M1[school] is estimated to be 0.21.
2. So how important is M1[school]? The variance of MathAb is estimated to be 0.25, so math ability and school have roughly the same variance, and both of course have mean 0. The math ability coefficients, meanwhile, are larger—often much larger—than the school coefficients in every case, so math ability is certainly more important than school in explaining whether questions were answered correctly. At this point, we are merely exploring the magnitude of effect.
3. You could also include a school-level latent variable for each question. For instance, you could type

```
. gsem (MathAb M1[school] N1[school] -> q1) ///
       (MathAb M1[school] N2[school] -> q2) ///
       (MathAb M1[school] N3[school] -> q3) ///
       (MathAb M1[school] N4[school] -> q4) ///
       (MathAb M1[school] N5[school] -> q5) ///
       (MathAb M1[school] N6[school] -> q6) ///
       (MathAb M1[school] N7[school] -> q7) ///
       (MathAb M1[school] N8[school] -> q8), logit
```

   You will sometimes see such effects included in multilevel measurement models in theoretical discussions of models. Be aware that estimation of models with many latent variables is problematic, requiring both time and luck.

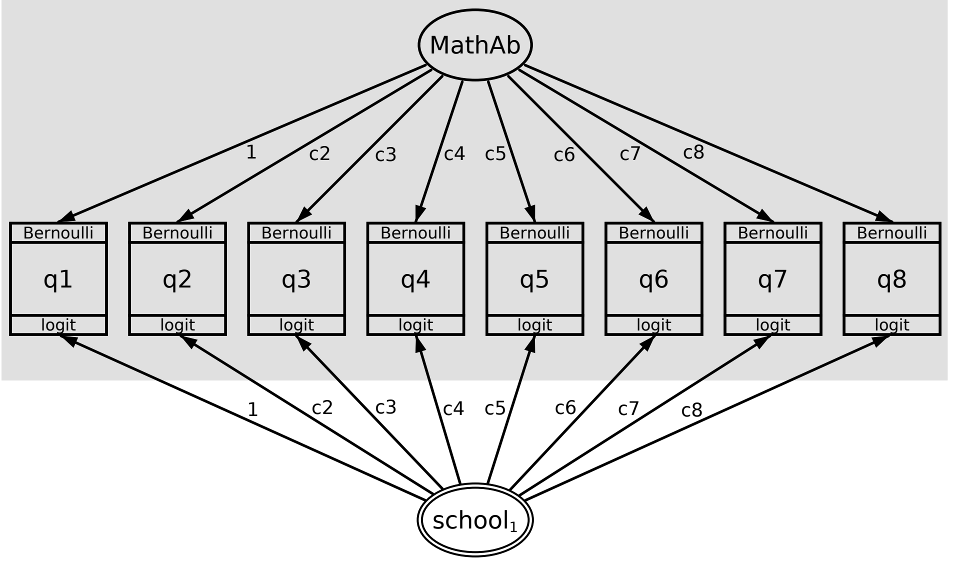
Fitting the variance-components model

In a variance-components model, school would affect math ability which would affect correctness of answers to questions. The model might be drawn like this:



The above is a great way to draw the model. Sadly, `gsem` cannot understand it. The problem from `gsem`'s perspective is that one latent variable is affecting another and the two latent variables are at different levels.

So we have to draw the model differently:



The models may look different, but constraining the coefficients along the paths from math ability and from school to each question is identical in effect to the model above.

The result of fitting the model is

```
. gsem (MathAb M1[school] ->
>       q1@1  q2@c2  q3@c3  q4@c4  q5@c5  q6@c6  q7@c7 q8@c8), logit
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 = -2642.8248
Fitting full model:
Iteration 0:  Log likelihood = -2651.7239  (not concave)
Iteration 1:  Log likelihood = -2644.4937
Iteration 2:  Log likelihood =   -2634.92
Iteration 3:  Log likelihood = -2633.9336
Iteration 4:  Log likelihood = -2633.5924
Iteration 5:  Log likelihood = -2633.5922
Generalized structural equation model                    Number of obs = 500
```

(output omitted)

```
Log likelihood = -2633.5922
 ( 1)  [q1]M1[school] = 1
 ( 2)  [q1]MathAb = 1
 ( 3)  [q2]M1[school] - [q2]MathAb = 0
 ( 4)  [q3]M1[school] - [q3]MathAb = 0
 ( 5)  [q4]M1[school] - [q4]MathAb = 0
 ( 6)  [q5]M1[school] - [q5]MathAb = 0
 ( 7)  [q6]M1[school] - [q6]MathAb = 0
 ( 8)  [q7]M1[school] - [q7]MathAb = 0
 ( 9)  [q8]M1[school] - [q8]MathAb = 0
```

	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
q1						
M1[school]	1	(constrained)				
MathAb	1	(constrained)				
_cons	.0385522	.1556214	0.25	0.804	-.2664601	.3435646
q2						
M1[school]	.3876281	.1156823	3.35	0.001	.1608951	.6143612
MathAb	.3876281	.1156823	3.35	0.001	.1608951	.6143612
_cons	-.4633143	.1055062	-4.39	0.000	-.6701028	-.2565259
q3						
M1[school]	.4871164	.1295515	3.76	0.000	.2332001	.7410328
MathAb	.4871164	.1295515	3.76	0.000	.2332001	.7410328
_cons	.1533212	.1098068	1.40	0.163	-.0618962	.3685386
q4						
M1[school]	.3407151	.1058542	3.22	0.001	.1332446	.5481856
MathAb	.3407151	.1058542	3.22	0.001	.1332446	.5481856
_cons	-.3246936	.1011841	-3.21	0.001	-.5230108	-.1263763

q5						
M1[school]	.8327426	.1950955	4.27	0.000	.4503624	1.215123
MathAb	.8327426	.1950955	4.27	0.000	.4503624	1.215123
_cons	-.0490579	.1391324	-0.35	0.724	-.3217524	.2236365
q6						
M1[school]	.6267415	.1572247	3.99	0.000	.3185868	.9348962
MathAb	.6267415	.1572247	3.99	0.000	.3185868	.9348962
_cons	-.3135398	.1220389	-2.57	0.010	-.5527317	-.074348
q7						
M1[school]	.7660343	.187918	4.08	0.000	.3977219	1.134347
MathAb	.7660343	.187918	4.08	0.000	.3977219	1.134347
_cons	.1039102	.1330652	0.78	0.435	-.1568927	.3647131
q8						
M1[school]	.5600833	.1416542	3.95	0.000	.2824462	.8377203
MathAb	.5600833	.1416542	3.95	0.000	.2824462	.8377203
_cons	-.0264193	.1150408	-0.23	0.818	-.2518951	.1990565
var(M1[school])	.1719347	.1150138			.0463406	.6379187
var(MathAb)	2.062489	.6900045			1.070589	3.973385

1. Note that for each question, the coefficient on `MathAb` is identical to the coefficient on `M1[school]`.
2. We estimate separate variances for `M1[school]` and `MathAb`. They are 0.17 and 2.06. Now that the coefficients are the same on school and ability, we can directly compare these variances. We see that math ability has a much larger affect than does school.

Fitting the model with the Builder

Use the diagram in *Fitting the two-level model* above for reference.

1. Open the dataset.

   In the Command window, type

   ```
   . use https://www.stata-press.com/data/r18/gsem_cfa
   ```

2. Open a new Builder diagram.

   Select menu item **Statistics > SEM (structural equation modeling) > Model building and estimation**.

3. Put the Builder in `gsem` mode by clicking on the GSEM button.
4. Create the measurement component for `MathAb`.

   Select the Add measurement component tool, and then click in the diagram about one-fourth of the way down from the top and slightly left of the center.

   In the resulting dialog box,

   a. change the *Latent variable name* to `MathAb`;

   b. select `q1`, `q2`, `q3`, `q4`, `q5`, `q6`, `q7`, and `q8` by using the *Measurement variables* control;

c. check *Make measurements generalized*;

d. select `Bernoulli, Logit` in the *Family/Link* control;

e. select `Down` in the *Measurement direction* control;

f. click on **OK**.

If you wish, move the component by clicking on any variable and dragging it.

5. Create the school-level latent variable.

a. Select the Add multilevel latent variable tool, , and click about one-fourth of the way up from the bottom and slightly left of the center.

b. In the Contextual Toolbar, click on the button.

c. Select the nesting level and nesting variable by selecting `2` from the *Nesting depth* control and selecting `school` > *Observations* in the next control.

d. Specify `M1` as the *Base name*.

e. Click on **OK**.

6. Create the factor-loading paths for the multilevel latent variable.

a. Select the Add path tool, .

b. Click in the top-left quadrant of the double oval for `school`$_1$ (it will highlight when you hover over it), and drag a path to the bottom of the `q1` rectangle (it will highlight when you can release to connect the path).

c. Continuing with the tool, draw paths from `school`$_1$ to each of the remaining rectangles.

7. Clean up paths.

If you do not like where a path has been connected to its variables, use the Select tool, , to click on the path, and then simply click on where it connects to a rectangle or oval and drag the endpoint.

8. Estimate.

Click on the **Estimate** button, , in the Standard Toolbar, and then click on **OK** in the resulting *GSEM estimation options* dialog box.

9. To fit the model in *Fitting the variance-components model*, add constraints to the diagram created above.

a. From the SEM Builder menu, select **Estimation > Clear estimates** to clear results from the previous model.

b. Choose the Select tool, .

c. Click on the path from `MathAb` to `q1`. In the Contextual Toolbar, type `1` in the box and press *Enter*.

d. Click on the path from `school`$_1$ to `q1`. In the Contextual Toolbar, type `1` in the box and press *Enter*.

c. Click on the path from `MathAb` to `q2`. In the Contextual Toolbar, type `c2` in the box and press *Enter*.

d. Click on the path from `school`$_1$ to `q2`. In the Contextual Toolbar, type `c2` in the box and press *Enter*.

e. Repeat this process to add the c3 constraint on both paths to q3, the c4 constraint on both paths to q4, ..., and the c8 constraint on both paths to q8.

10. Estimate again.

Click on the **Estimate** button, , in the Standard Toolbar, and then click on **OK** in the resulting *GSEM estimation options* dialog box.

You can open a completed diagram in the Builder for the first model by typing

```
. webgetsem gsem_mlcfa1
```

You can open a completed diagram in the Builder for the second model by typing

```
. webgetsem gsem_mlcfa2
```

References

Mehta, P. D., and M. C. Neale. 2005. People are variables too: Multilevel structural equations modeling. *Psychological Methods* 10: 259–284. https://doi.org/10.1037/1082-989X.10.3.259.

Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman and Hall/CRC.

Also see

[SEM] **Example 27g** — Single-factor measurement model (generalized response)

[SEM] **Example 29g** — Two-parameter logistic IRT model

[SEM] **Intro 5** — Tour of models

[SEM] **gsem** — Generalized structural equation model estimation command