Description

To demonstrate a one-parameter logistic IRT (Rasch) model, we use the following data:

```
. use http://www.stata-press.com/data/r14/gsem_cfa
(Fictional math abilities data)
. summarize
```

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

. 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 Rasch models, IRT models, and their extensions, see Embretson and Reise (2000), van der Linden and Hambleton (1997), Skrondal and Rabe-Hesketh (2004), Andrich (1988), Bond and Fox (2007), and Fischer and Molenaar (1995). The standard one-parameter logistic model can be fit using the _irt_ 1pl command; see [**IRT**] _irt_ 1pl. This example demonstrates how to fit this model. With _gsem_, we can build on this model to fit many of the extensions to basic IRT models discussed in these books.

See _Item response theory (IRT) models_ in [SEM] _intro_ 5 for background.
Remarks and examples

Remarks are presented under the following headings:

- 1-PL IRT model with unconstrained variance
- 1-PL IRT model with variance constrained to 1
- Obtaining item characteristic curves
- Fitting the model with the Builder

1-PL IRT model with unconstrained variance

Mechanically speaking, one-parameter logistic (1-PL) IRT models are similar to the probit measurement model we demonstrated in [SEM] example 27g. The differences are that we will use logit rather than probit and that we will place various constraints on the logit model to obtain results that will allow us to judge the difficulty of the individual questions.

The model we wish to fit is

![Diagram of the 1-PL IRT model]

In the 1-PL model, we place constraints that all coefficients, the factor loadings, are equal to 1. The negative of the intercept for each question will then represent the difficulty of the question:
. gsem (MathAb -> (q1-q8)@1), 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 = -2653.2353

Fitting full model:
Iteration 0: log likelihood = -2653.2353
Iteration 1: log likelihood = -2651.2171
Iteration 2: log likelihood = -2650.9117
Iteration 3: log likelihood = -2650.9116

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 = -2650.9116
( 1) [q1]MathAb = 1
( 2) [q2]MathAb = 1
( 3) [q3]MathAb = 1
( 4) [q4]MathAb = 1
( 5) [q5]MathAb = 1
( 6) [q6]MathAb = 1
( 7) [q7]MathAb = 1
( 8) [q8]MathAb = 1
| q1 <- MathAb | Coef.       | Std. Err.   | z   | P>|z|  | [95% Conf. Interval] |
|------------|------------|-------------|-----|-------|---------------------|
|            | q1         | .0293252    | .1047674 | 0.28  | 0.780    | -.1760152 .2346656 |
| q2 <- MathAb | cons      | -.5025012   | .1068768 | -4.70 | 0.000    | -.7119759 -.2930264 |
| q3 <- MathAb | cons      | .1607425    | .104967  | 1.53  | 0.126    | -.044989 .3664739  |
| q4 <- MathAb | cons      | -.3574951   | .105835  | -3.38 | 0.001    | -.564928 -.1500623 |
| q5 <- MathAb | cons      | -.0456599   | .1047812 | -0.44 | 0.663    | -.2510274 .1597075 |
| q6 <- MathAb | cons      | -.3097521   | .1055691 | -2.93 | 0.003    | -.5166637 -.1028404 |
| q7 <- MathAb | cons      | .09497      | .1048315 | 0.91  | 0.365    | -.1104959 .300436  |
| q8 <- MathAb | cons      | -.0269104   | .1047691 | -0.26 | 0.797    | -.232254 .1784332  |

Notes:

1. We had to use `gsem` and not `sem` to fit this model because the response variables were 0/1 and not continuous and because we wanted to use logit and not a continuous model.

2. To place the constraints that all coefficients are equal to 1, in the diagram we placed 1s along the path from the underlying latent factor MathAb to each of the questions. In the command language, we added @1 to our command:

   ```
   gsem (MathAb -> (q1-q8)@1), logit
   ```

   Had we omitted the @1, we would have obtained coefficients about how well each question measured math ability.

   There are several ways we could have asked that the model above be fit. They include the following:

   ```
   gsem (MathAb -> q1 q2 q3 q4 q5 q6 q7 q8), logit
   gsem (MathAb -> (q1 q2 q3 q4 q5 q6 q7 q8)@1), logit
   ```

   Similarly, for the shorthand logit, we could have typed `family(bernoulli) link(logit)`.

3. The negative of the reported intercept is proportional to the difficulty of the item. The most difficult is q2, and the least difficult is q3.
The goal of the 1-PL model is in fact to constrain the loadings to be equal. In the previous model, that was achieved by constraining them to be 1 and letting the variance of the latent variable float. An alternative with perhaps easier-to-interpret results would constrain the variance of the latent variable to be 1—giving it a standard-normal interpretation—and constrain the loadings to be merely equal:

```
.gsem (MathAb -> (q1-q8)b), logit var(MathAb@1) 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 = -2656.1973
Iteration 1:  log likelihood = -2650.9139
Iteration 2:  log likelihood = -2650.9116
Iteration 3:  log likelihood = -2650.9116
Generalized structural equation model  Number of obs = 500
Log likelihood = -2650.9116
( 1)  [q1]MathAb - [q8]MathAb = 0
( 2)  [q2]MathAb - [q8]MathAb = 0
( 3)  [q3]MathAb - [q8]MathAb = 0
( 4)  [q4]MathAb - [q8]MathAb = 0
( 5)  [q5]MathAb - [q8]MathAb = 0
( 6)  [q6]MathAb - [q8]MathAb = 0
( 7)  [q7]MathAb - [q8]MathAb = 0
( 8)  [var(MathAb)]_cons = 1
```

|       | Coef.   | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-------|---------|-----------|-------|------|---------------------|
| q1 <- | MathAb_ | .8904887  | .0575755 | 15.47 | 0.000  | .7776429 1.003335 |
|       | _cons   | .0293253  | .1047674 | 0.28  | 0.780  | -.1760151 .2346657 |
| q2 <- | MathAb_ | .8904887  | .0575755 | 15.47 | 0.000  | .7776429 1.003335 |
|       | _cons   | -.5025011 | .1068768 | -4.70 | 0.000  | -.7119758 -.2930264 |
| q3 <- | MathAb_ | .8904887  | .0575755 | 15.47 | 0.000  | .7776429 1.003335 |
|       | _cons   | .1607425  | .104967  | 1.53  | 0.126  | -.044989 .366474  |
| q4 <- | MathAb_ | .8904887  | .0575755 | 15.47 | 0.000  | .7776429 1.003335 |
|       | _cons   | -.3574951 | .105835  | -3.38 | 0.001  | -.5649279 -.1500622 |
| q5 <- | MathAb_ | .8904887  | .0575755 | 15.47 | 0.000  | .7776429 1.003335 |
|       | _cons   | -.0456599 | .1047812 | -0.44 | 0.663  | -.2510273 .1597076 |
| q6 <- | MathAb_ | .8904887  | .0575755 | 15.47 | 0.000  | .7776429 1.003335 |
|       | _cons   | -.309752  | .1055691 | -2.93 | 0.003  | -.5166637 -.1028403 |
| q7 <- | MathAb_ | .8904887  | .0575755 | 15.47 | 0.000  | .7776429 1.003335 |
|       | _cons   | .0949701  | .1048315 | 0.91  | 0.365  | -.1104959 .300436  |
```plaintext
q8 <- 
MathAb  .8904887  .0575755  15.47  0.000  .7776429  1.003335
_cons  -.0269103  .1047691  -0.26  0.797  -.232254  .1784333

var(MathAb)  1 (constrained)

Notes:
1. The log-likelihood values of both models is $-2650.9116$. The models are equivalent.
2. Intercepts are unchanged.

**Obtaining item characteristic curves**

Item characteristic curves graph the conditional probability of a particular response given the latent trait. In our case, this simply amounts to graphing the probability of a correct answer against math ability. After estimation, we can obtain the predicted probabilities of a correct answer by typing

```
.predict pr*, pr
(option conditional(ebmeans) assumed)
(using 7 quadrature points)
```

We can obtain the predicted value of the latent variable by typing

```
.predict ability, latent(MathAb)
(option ebmeans assumed)
(using 7 quadrature points)
```

and thus we can obtain the item characteristic curves for all eight questions by typing

```
.twoway line pr1 pr2 pr3 pr4 pr5 pr6 pr7 pr8 ability, sort xlabel(-1.5(.5)1.5)
```

![Graph of item characteristic curves](image.png)

**Predicted mean (q1 correct)**

**Predicted mean (q2 correct)**

**Predicted mean (q3 correct)**

**Predicted mean (q4 correct)**

**Predicted mean (q5 correct)**

**Predicted mean (q6 correct)**

**Predicted mean (q7 correct)**

**Predicted mean (q8 correct)**
A less busy graph might show merely the most difficult and least difficult questions:

```
.twoway line pr2 pr3 ability, sort xlabel(-1.5(.5)1.5)
```

The slopes of each curve are identical because we have constrained them to be identical. Thus we just see the shift between difficulties with the lower items having higher levels of difficulty.

**Fitting the model with the Builder**

Use the diagram in *1-PL IRT model with unconstrained variance* above for reference.

1. Open the dataset.
   In the Command window, type
   ```
   . use http://www.stata-press.com/data/r14/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 button.

4. Create the measurement component for MathAb.
   Select the Add Measurement Component tool, , and then click in the diagram about one-third 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. Constrain all path coefficients to 1.
   a. Choose the Select tool, ![Select tool].
   b. Click on the path from MathAb to q1. In the Contextual Toolbar, type 1 in the ![Beta] box and press Enter.
   c. Repeat this process to add the 1 constraint on the paths from MathAb to each of the other measurement variables.

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

7. To fit the model in 1-PL IRT model with variance constrained to 1, change the constraints in 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, ![Select tool].
   c. Click on the path from MathAb to q1. In the Contextual Toolbar, type b in the ![Beta] box and press Enter.
   d. Repeat this process to add the b constraint on the paths from MathAb to each of the other measurement variables.
   e. With ![Select tool], click on the oval for MathAb. In the Contextual Toolbar, type 1 in the ![Sigma squared] box and press Enter.

8. Estimate again.
   Click on the Estimate button, ![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_irt1
```
You can open a completed diagram in the Builder for the second model by typing
```
.webgetsem gsem_irt2
```

References


**Also see**

[SEM] example 27g — Single-factor measurement model (generalized response)
[SEM] example 29g — Two-parameter logistic IRT model
[SEM] gsem — Generalized structural equation model estimation command
[SEM] predict after gsem — Generalized linear predictions, etc.
[SEM] intro 5 — Tour of models
[IRT] irt 1pl — One-parameter logistic model