Example 28g — One-parameter logistic IRT (Rasch) model

Description

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

```
. use https://www.stata-press.com/data/r16/gsem_cfa
  (Fictional math abilities data)
. summarize
                  Variable |     Obs  Mean    Std. Dev.  Min   Max
-----------------+--------------------------------------------
            school |       500   10.5      5.772056    1  100
            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       .4993423     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 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 (2015), 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

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:
Example 28g — One-parameter logistic IRT (Rasch) model

```
. 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
```
### Example 28g — One-parameter logistic IRT (Rasch) model

|   | Coef.  | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|---|--------|-----------|-------|-------|----------------------|
| q1 |        |           |       |       |                      |
| MathAb | 1 (constrained) | .0293252 | .1047674 | 0.28  | 0.780 | -.1760152 | .2346656 |
| _cons |           |           |       |       |                      |
| q2 |        |           |       |       |                      |
| MathAb | 1 (constrained) | -.5025012 | .1068768 | -4.70 | 0.000 | -.7119759 | -.2930264 |
| _cons |           |           |       |       |                      |
| q3 |        |           |       |       |                      |
| MathAb | 1 (constrained) | .1607425 | .104967 | 1.53  | 0.126 | -.044989 | .3664739 |
| _cons |           |           |       |       |                      |
| q4 |        |           |       |       |                      |
| MathAb | 1 (constrained) | -.3574951 | .105835 | -3.38 | 0.001 | -.564928 | -.1500623 |
| _cons |           |           |       |       |                      |
| q5 |        |           |       |       |                      |
| MathAb | 1 (constrained) | -.0456599 | .1047812 | -0.44 | 0.663 | -.2510274 | .1597075 |
| _cons |           |           |       |       |                      |
| q6 |        |           |       |       |                      |
| MathAb | 1 (constrained) | -.3097521 | .1055691 | -2.93 | 0.003 | -.5166637 | -.1028404 |
| _cons |           |           |       |       |                      |
| q7 |        |           |       |       |                      |
| MathAb | 1 (constrained) | .09497 | .1048315 | 0.91  | 0.365 | -.1104959 | .300436 |
| _cons |           |           |       |       |                      |
| q8 |        |           |       |       |                      |
| MathAb | 1 (constrained) | -.0269104 | .1047691 | -0.26 | 0.797 | -.232254 | .1784332 |
| _cons |           |           |       |       |                      |
| var(MathAb) | | .7929701 | .1025406 |  | .6154407 | 1.02171 |

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


1-PL IRT model with variance constrained to 1

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:

<table>
<thead>
<tr>
<th>Iteration</th>
<th>log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-2750.3114</td>
</tr>
<tr>
<td>1</td>
<td>-2749.3709</td>
</tr>
<tr>
<td>2</td>
<td>-2749.3708</td>
</tr>
</tbody>
</table>

Refining starting values:

<table>
<thead>
<tr>
<th>Grid node</th>
<th>log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-2645.8536</td>
</tr>
</tbody>
</table>

Fitting full model:

<table>
<thead>
<tr>
<th>Iteration</th>
<th>log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-2656.1973</td>
</tr>
<tr>
<td>1</td>
<td>-2650.9139</td>
</tr>
<tr>
<td>2</td>
<td>-2650.9116</td>
</tr>
<tr>
<td>3</td>
<td>-2650.9116</td>
</tr>
</tbody>
</table>

Generalized structural equation model

<table>
<thead>
<tr>
<th>Number of obs = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log likelihood = -2650.9116</td>
</tr>
</tbody>
</table>

( 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) \(\var(MathAb) = 1\)

| 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 | -.0449899 .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  |
Example 28g — One-parameter logistic IRT (Rasch) model

<table>
<thead>
<tr>
<th></th>
<th>MathAb</th>
<th>_cons</th>
<th>var(MathAb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>q8</td>
<td>0.8904887</td>
<td>0.0575755</td>
<td>0.7776429</td>
</tr>
<tr>
<td></td>
<td>-0.0269103</td>
<td>0.1047691</td>
<td>-0.232254</td>
</tr>
</tbody>
</table>

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)
```
A less busy graph might show merely the most difficult and least difficult questions:

```stata
.twoway line pr2 pr3 ability, sort xlabel(-1.5(.5)1.5) legend(col(1))
```

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
   
   ```stata
   use https://www.stata-press.com/data/r16/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 \( \square \) button.

4. Create the measurement component for *MathAb*.
   
   Select the Add measurement component tool, \( \square \), 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, 
   b. Click on the path from MathAb to q1. In the Contextual Toolbar, type 1 in the β 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, 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, .
   c. Click on the path from MathAb to q1. In the Contextual Toolbar, type b in the β 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 , click on the oval for MathAb. In the Contextual Toolbar, type 1 in the σ² box and press Enter.

8. 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_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] **Intro 5** — Tour of models

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

[SEM] **predict after gsem** — Generalized linear predictions, etc.

[IRT] **irt 1pl** — One-parameter logistic model