Example 35g — Ordered probit and ordered logit

Description   Remarks and examples   Reference   Also see

Description

Below we demonstrate ordered probit and ordered logit in a measurement-model context. We are not going to illustrate every family/link combination. Ordered probit and logit, however, are unique in that a single equation is able to predict a set of ordered outcomes. The unordered alternative, `mlogit`, requires $k-1$ equations to fit k (unordered) outcomes.

To demonstrate ordered probit and ordered logit, we use the following data:

```
. use https://www.stata-press.com/data/r19/gsem_issp93
(Selection for ISSP 1993)

. describe

Contains data from https://www.stata-press.com/data/r19/gsem_issp93.dta
 Observations:           871                  Selection for ISSP 1993
    Variables:             8                  21 Mar 2024 16:03
                                              (_dta has notes)
--------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
--------------------------------------------------------------------------------
id              int     %9.0g                 Respondent identifier
y1              byte    %26.0g     agree5     Too much science, not enough
                                                feelings & faith
y2              byte    %26.0g     agree5     Science does more harm than good
y3              byte    %26.0g     agree5     Any change makes nature worse
y4              byte    %26.0g     agree5     Science will solve environmental
                                                problems
sex             byte    %9.0g      sex        Sex
age             byte    %9.0g      age        Age (6 categories)
edu             byte    %20.0g     edu        Education (6 categories)
--------------------------------------------------------------------------------
Sorted by:

. notes

_dta:
  1.   Source: Data from pages 42-43 of Greenacre, M. J., and J. Blasius. 2006.
       Multiple Correspondence Analysis and Related Methods. Boca Raton, FL:
       Chapman & Hall. Data are part of the International Social Survey Program
       (ISSP) 1993.
  2.   Full text of y1: We believe too often in science, and not enough in
       feelings and faith.
  3.   Full text of y2: Overall, modern science does more harm than good.
  4.   Full text of y3: Any change humans cause in nature, no matter how
       scientific, is likely to make things worse.
  5.   Full text of y4: Modern science will solve our environmental problems
       with little change to our way of life.
```

See *Structural models 5: Ordinal models* in [SEM] **Intro 5** for background.

Remarks and examples

Remarks are presented under the following headings:

Ordered probit
Ordered logit
Fitting the model with the Builder

Ordered probit

For the measurement model, we focus on variables `y1` through `y4`. Each variable contains 1–5, with 1 meaning strong disagreement and 5 meaning strong agreement with a statement about science.

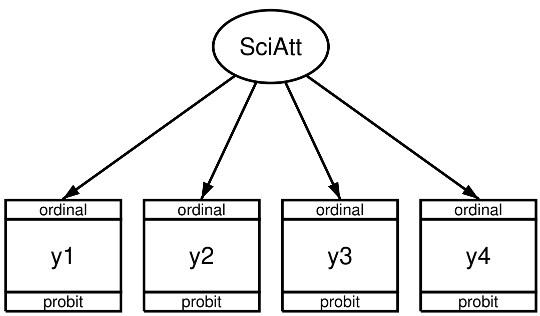
Ordered probit produces predictions about the probabilities that a respondent gives response 1, response 2, ..., response k. It does this by dividing up the domain of an $N(0, 1)$ distribution into k categories defined by $k - 1$ cutpoints, c_1, c_2, ..., c_{k-1}. Individual respondents are assumed to have a score $s = X\beta + \epsilon$, where $\epsilon \sim N(0, 1)$, and then that score is used along with the cutpoints to produce probabilities for each respondent producing response 1, 2, ..., k.

$$\text{Pr}(\text{response is } i \,|\, X) = \text{Pr}(c_{i-1} < X\beta + \epsilon \leq c_i)$$

where $c_0 = -\infty$; $c_k = +\infty$; and c_1, c_2, ..., c_{k-1} and β are parameters of the model to be fit. This ordered probit model has long been known in Stata circles as oprobit.

We have a set of four questions designed to determine the respondent's attitude toward science, each question with $k = 5$ possible answers ranging on a Likert scale from 1 to 5. With ordered probit in hand, we have a way to take a continuous variable, say, a latent variable we will call `SciAtt`, and produce predicted categorical responses.

The measurement model we want to fit is



We fit the model in the command language by typing

```
. gsem (y1 y2 y3 y4 <- SciAtt), oprobit
Fitting fixed-effects model:
Iteration 0:  Log likelihood = -5227.8743
Iteration 1:  Log likelihood = -5227.8743
Refining starting values:
Grid node 0:  Log likelihood = -5230.8106
Fitting full model:
Iteration 0:  Log likelihood = -5230.8106  (not concave)
Iteration 1:  Log likelihood = -5132.1849  (not concave)
Iteration 2:  Log likelihood = -5069.5037
Iteration 3:  Log likelihood = -5040.4779
Iteration 4:  Log likelihood = -5040.2397
Iteration 5:  Log likelihood = -5039.8242
Iteration 6:  Log likelihood =  -5039.823
Iteration 7:  Log likelihood =  -5039.823
Generalized structural equation model                 Number of obs = 871
Response: y1
Family:   Ordinal
Link:     Probit
Response: y2
Family:   Ordinal
Link:     Probit
Response: y3
Family:   Ordinal
Link:     Probit
Response: y4
Family:   Ordinal
Link:     Probit
Log likelihood = -5039.823
 ( 1)  [y1]SciAtt = 1
```

	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
y1						
SciAtt	1	(constrained)				
y2						
SciAtt	1.424366	.2126574	6.70	0.000	1.007565	1.841167
y3						
SciAtt	1.283359	.1797557	7.14	0.000	.931044	1.635674
y4						
SciAtt	-.0322354	.0612282	-0.53	0.599	-.1522405	.0877697
/y1						
cut1	-1.343148	.0726927			-1.485623	-1.200673
cut2	.0084719	.0521512			-.0937426	.1106863
cut3	.7876538	.0595266			.6709837	.9043238
cut4	1.989985	.0999181			1.794149	2.18582

/y2						
cut1	-1.997245	.1311972			-2.254387	-1.740104
cut2	-.8240241	.0753839			-.9717738	-.6762743
cut3	.0547025	.0606036			-.0640784	.1734834
cut4	1.419923	.1001258			1.22368	1.616166
/y3						
cut1	-1.271915	.0847483			-1.438019	-1.105812
cut2	.1249493	.0579103			.0114472	.2384515
cut3	.9752553	.0745052			.8292277	1.121283
cut4	2.130661	.1257447			1.884206	2.377116
/y4						
cut1	-1.484063	.0646856			-1.610844	-1.357281
cut2	-.4259356	.0439145			-.5120065	-.3398647
cut3	.1688777	.0427052			.0851771	.2525782
cut4	.9413113	.0500906			.8431356	1.039487
var(SciAtt)	.5265523	.0979611			.3656637	.7582305

Notes:

1. The cutpoints $c_1, \ldots, c_4$ are labeled `cut1`, ..., `cut4` in the output. We have a separate cutpoint for each of the four questions `y1`, ..., `y4`. Look at the estimated cutpoints for `y1`, which are -1.343, 0.008, 0.788, and 1.99. The probabilities that a person with `SciAtt` $= 0$ (its mean) would give the various responses are

$$
\begin{aligned}
\Pr(\text{response } 1) &= \texttt{normal(-1.343)} = 0.090 \\
\Pr(\text{response } 2) &= \texttt{normal(0.008)} - \texttt{normal}(-1.343) = 0.414 \\
\Pr(\text{response } 3) &= \texttt{normal(0.788)} - \texttt{normal}(0.008) = 0.281 \\
\Pr(\text{response } 4) &= \texttt{normal(1.99)} - \texttt{normal}(0.788) = 0.192 \\
\Pr(\text{response } 5) &= 1 - \texttt{normal(1.99)} = 0.023
\end{aligned}
$$

2. The path coefficients (`y1 y2 y3 y4 <- SciAtt`) measure the effect of the latent variable we called science attitude on each of the responses.

3. The estimated path coefficients are 1, 1.42, 1.28, and -0.03 for the four questions.

4. If you read the questions—they are listed above—you will find that in all but the fourth question, agreement signifies a negative attitude toward science. Thus `SciAtt` measures a negative attitude toward science because the loadings on negative questions are positive and the loading on the single positive question is negative.

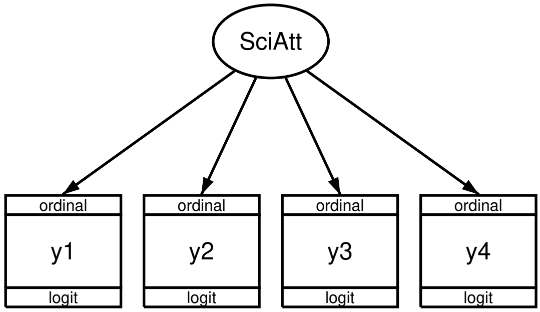
5. The direction of the meanings of latent variables is always a priori indeterminate and is set by the identifying restrictions we apply. We applied—or more correctly, `gsem` applied for us—the constraint that `y1 <- SciAtt` has path coefficient 1. Because statement 1 was a negative statement about science, that was sufficient to set the direction of `SciAtt` to be the opposite of what we hoped for.

   The direction does not matter. You simply must remember to interpret the latent variable correctly when reading results based on it. In the models we fit, including more complicated models, the signs of the coefficients will work themselves out to adjust for the direction of the variable.

Ordered logit

The description of the ordered logit model is identical to that of the ordered probit model except that where we assumed a normal distribution in our explanation above, we now assume a logit distribution. The distributions are similar.

To fit an ordered logit (ologit) model, the link function shown in the boxes merely changes from probit to logit:



We can fit the model in the command language by typing

```
. gsem (y1 y2 y3 y4 <- SciAtt), ologit
Fitting fixed-effects model:

Iteration 0:  Log likelihood = -5227.8743
Iteration 1:  Log likelihood = -5227.8743

Refining starting values:

Grid node 0:  Log likelihood = -5127.9026

Fitting full model:

Iteration 0:  Log likelihood = -5127.9026  (not concave)
Iteration 1:  Log likelihood = -5065.4679
Iteration 2:  Log likelihood = -5035.9766
Iteration 3:  Log likelihood = -5035.0943
Iteration 4:  Log likelihood = -5035.0353
Iteration 5:  Log likelihood = -5035.0352

Generalized structural equation model                  Number of obs = 871

Response: y1
Family:   Ordinal
Link:     Logit

Response: y2
Family:   Ordinal
Link:     Logit

Response: y3
Family:   Ordinal
Link:     Logit

Response: y4
Family:   Ordinal
Link:     Logit

Log likelihood = -5035.0352

 ( 1)  [y1]SciAtt = 1
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
y1						
SciAtt	1	(constrained)				
y2						
SciAtt	1.394767	.2065479	6.75	0.000	.9899406	1.799593
y3						
SciAtt	1.29383	.1845113	7.01	0.000	.9321939	1.655465
y4						
SciAtt	-.0412446	.0619936	-0.67	0.506	-.1627498	.0802606
/y1						
cut1	-2.38274	.1394292			-2.656016	-2.109464
cut2	-.0088393	.0889718			-.1832207	.1655422
cut3	1.326292	.106275			1.117997	1.534587
cut4	3.522017	.1955535			3.138739	3.905295
/y2						
cut1	-3.51417	.2426595			-3.989774	-3.038566
cut2	-1.421711	.135695			-1.687669	-1.155754
cut3	.0963154	.1046839			-.1088612	.3014921
cut4	2.491459	.1840433			2.130741	2.852178
/y3						
cut1	-2.263557	.1618806			-2.580838	-1.946277
cut2	.2024798	.1012122			.0041075	.400852
cut3	1.695997	.1393606			1.422855	1.969138
cut4	3.828154	.2464566			3.345108	4.3112
/y4						
cut1	-2.606013	.1338801			-2.868413	-2.343613
cut2	-.6866159	.0718998			-.8275369	-.5456949
cut3	.268862	.0684577			.1346874	.4030366
cut4	1.561921	.0895438			1.386419	1.737424
var(SciAtt)	1.715641	.3207998			1.189226	2.475077

Note:

1. Results are nearly identical to those reported for ordered probit.

Fitting the model with the Builder

Use the diagram in *Ordered probit* above for reference.

1. Open the dataset.

   In the Command window, type

   ```
   . use https://www.stata-press.com/data/r19/gsem_issp93
   ```

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 `SciAtt`.

   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 `SciAtt`;
   b. select `y1`, `y2`, `y3`, and `y4` by using the *Measurement variables* control;
   c. check *Make measurements generalized*;
   d. select `Ordinal, Probit` 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. Estimate.

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

6. To fit the model in Ordered logit, change the type of generalized response for each of the measurement variables.

   a. Choose the Select tool, .
   b. Click on the `y1` rectangle. In the Contextual Toolbar, select `Ordinal, Logit` in the *Family/Link* control.
   c. Repeat this process to change the family and link to `Ordinal, Logit` for `y2`, `y3`, and `y4`.

7. 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 for the ordered probit model in the Builder by typing

```
. webgetsem gsem_oprobit
```

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

```
. webgetsem gsem_ologit
```

Reference

Greenacre, M. J. 2006. "From simple to multiple correspondence analysis". In *Multiple Correspondence Analysis and Related Methods*, edited by M. J. Greenacre and J. Blasius. Boca Raton, FL: Chapman and Hall.

Also see

[SEM] **Example 1** — Single-factor measurement model

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

[SEM] **Example 33g** — Logistic regression

[SEM] **Example 36g** — MIMIC model (generalized response)

[SEM] **Example 37g** — Multinomial logistic regression

[SEM] **Intro 5** — Tour of models

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