example 41g — Two-level multinomial logistic regression (multilevel)

Description Remarks and examples References Also see

Description

We demonstrate two-level multinomial logistic regression with random effects by using the following data:

```
. use http://www.stata-press.com/data/r13/gsem_lineup
(Fictional suspect identification data)

. describe

Contains data from http://www.stata-press.com/data/r13/gsem_lineup.dta
  obs:         6,535                          Fictional suspect
                                                identification data
 vars:             6                          29 Mar 2013 10:35
 size:       156,840                          (_dta has notes)
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
suspect         float   %9.0g                 suspect id
suswhite        float   %9.0g                 suspect is white
violent         float   %9.0g                 violent crime
location        float   %14.0g     loc        lineup location
witmale         float   %9.0g                 witness is male
chosen          float   %9.0g      choice     indvidual identified in linup by
                                                witness
-------------------------------------------------------------------------------
Sorted by:  suspect

. notes

_dta:
  1.  Fictional data inspired by Wright, D.B and Sparks, A.T., 1994, "Using
      multilevel multinomial regression to analyse line-up data", _Multilevel
      Modeling Newsletter_, Vol. 6, No. 1
  2.  Data contain repeated values of variable suspect.  Each suspect is viewed
      by multiple witnesses and each witness (1) declines to identify a
      suspect, (2) chooses a foil, or (3) chooses the suspect.

. tabulate location

         lineup |
       location |      Freq.     Percent        Cum.
----------------+-----------------------------------
 police_station |      2,228       34.09       34.09
        suite_1 |      1,845       28.23       62.33
        suite_2 |      2,462       37.67      100.00
----------------+-----------------------------------
          Total |      6,535      100.00
```

```
. tabulate chosen
```

indvidual identified in linup by witness	Freq.	Percent	Cum.
none	2,811	43.01	43.01
foil	1,369	20.95	63.96
suspect	2,355	36.04	100.00
Total	6,535	100.00	

In what follows, we re-create results similar to those of Wright and Sparks (1994), but we use fictional data. These data resemble the real data used by the authors in proportion of observations having each level of the outcome variable `chosen`, and the data produce results similar to those presented by the authors.

See *Structural models 6: Multinomial logistic regression* and *Multilevel mixed-effects models* in [SEM] **intro 5** for background.

For additional discussion of fitting multilevel multinomial logistic regression models, see Skrondal and Rabe-Hesketh (2003).

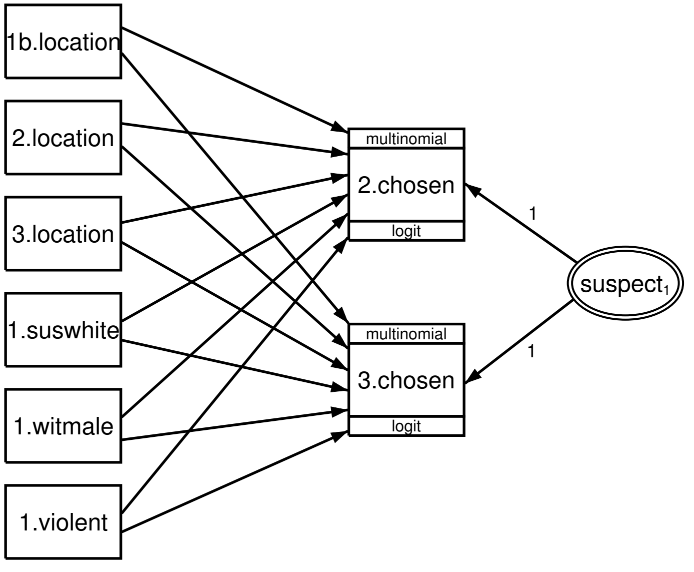
Remarks and examples

stata.com

Remarks are presented under the following headings:

Two-level multinomial logistic model with shared random effects

We wish to fit the following model:



This model concerns who is chosen in a police lineup. The response variables are `1.chosen`, `2.chosen`, and `3.chosen`, meaning `chosen` = 1 (code for not chosen), `chosen` = 2 (code for foil chosen), and `chosen` = 3 (code for suspect chosen). A foil is a stand-in who could not possibly be guilty of the crime.

We say the response variables are `1.chosen`, `2.chosen`, and `3.chosen`, but `1.chosen` does not even appear in the diagram. By its omission, we are specifying that `chosen` = 1 be treated as the base mlogit category. There are other ways we could have drawn this; see [SEM] **example 37g**.

In these data, each suspect was viewed by multiple witnesses. In the model, we include a random effect at the suspect level, and we constrain the effect to be equal for chosen values 2 and 3 (selecting the foil or the suspect).

We can fit this model with command syntax by typing

```
. gsem (i.chosen <- i.location i.suswhite i.witmale i.violent M1[suspect]@1),
> mlogit
Fitting fixed-effects model:
Iteration 0:   log likelihood = -6914.9098
Iteration 1:   log likelihood = -6696.7136
Iteration 2:   log likelihood = -6694.0006
Iteration 3:   log likelihood = -6693.9974
Iteration 4:   log likelihood = -6693.9974
Refining starting values:
Grid node 0:   log likelihood = -6705.0919
Fitting full model:
Iteration 0:   log likelihood = -6705.0919  (not concave)
Iteration 1:   log likelihood = -6654.5724
Iteration 2:   log likelihood = -6653.5717
Iteration 3:   log likelihood = -6653.5671
Iteration 4:   log likelihood = -6653.5671
Generalized structural equation model           Number of obs    =       6535
Log likelihood = -6653.5671
 ( 1)  [2.chosen]M1[suspect] = 1
 ( 2)  [3.chosen]M1[suspect] = 1
```

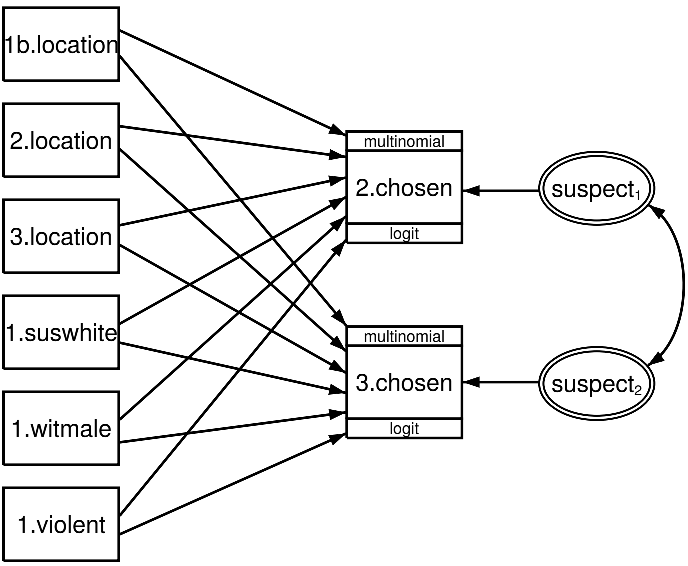
	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.chosen	(base outcome)					
2.chosen <-						
location						
suite_1	.3867066	.1027161	3.76	0.000	.1853868	.5880264
suite_2	.4915675	.0980312	5.01	0.000	.2994299	.6837051
1.suswhite	-.0275501	.0751664	-0.37	0.714	-.1748736	.1197734
1.witmale	-.0001844	.0680803	-0.00	0.998	-.1336193	.1332505
1.violent	.0356477	.0773658	0.46	0.645	-.1159864	.1872819
M1[suspect]	1	(constrained)				
_cons	-1.002334	.099323	-10.09	0.000	-1.197003	-.8076643
3.chosen <-						
location						
suite_1	-.2832042	.0936358	-3.02	0.002	-.4667271	-.0996814
suite_2	.1391796	.0863473	1.61	0.107	-.0300581	.3084172
1.suswhite	-.2397561	.0643075	-3.73	0.000	-.3657965	-.1137158
1.witmale	.1419285	.059316	2.39	0.017	.0256712	.2581857
1.violent	-1.376579	.0885126	-15.55	0.000	-1.55006	-1.203097
M1[suspect]	1	(constrained)				
_cons	.1781047	.0833393	2.14	0.033	.0147627	.3414468
var(M1[suspect])	.2538014	.0427302			.1824673	.3530228

Notes:

1. We show the interpretation of mlogit coefficients in [SEM] **example 37g**.
2. The estimated variance of the random effect is 0.2538, implying a standard deviation of 0.5038. Thus a 1-standard-deviation change in the random effect amounts to a $\exp(0.5038) = 1.655$ change in the relative-risk ratio. The effect is both practically significant and, from the output, statistically significant.
3. This is not the model fit by Wright and Sparks (1994). Those authors did not constrain the random effect to be the same for `chosen` equal to 2 and 3. They included separate but correlated random effects, and then took that even a step further.

Two-level multinomial logistic model with separate but correlated random effects

The model we wish to fit is



This is one of the models fit by Wright and Sparks (1994), although remember that we are using fictional data.

We can fit this model with command syntax by typing

```
. gsem (2.chosen <- i.location i.suswhite i.witmale i.violent M1[suspect])  ///
>      (3.chosen <- i.location i.suswhite i.witmale i.violent M2[suspect]), ///
>      mlogit
```

We did not even mention the assumed covariance between the random effects because latent exogenous variables are assumed to be correlated in the command language. Even so, we can specify the `cov()` option if we wish, and we might do so for emphasis or because we are unsure whether the parameter would be included.

```
. gsem (2.chosen <- i.location i.suswhite i.witmale i.violent M1[suspect])
>      (3.chosen <- i.location i.suswhite i.witmale i.violent M2[suspect]),
>      cov(M1[suspect]*M2[suspect]) mlogit
Fitting fixed-effects model:
Iteration 0:   log likelihood = -6914.9098
```

```
Iteration 1:   log likelihood = -6696.7136
Iteration 2:   log likelihood = -6694.0006
Iteration 3:   log likelihood = -6693.9974
Iteration 4:   log likelihood = -6693.9974
Refining starting values:
Grid node 0:   log likelihood = -6793.4228
Fitting full model:
Iteration 0:   log likelihood = -6793.4228  (not concave)
Iteration 1:   log likelihood = -6717.7507  (not concave)
Iteration 2:   log likelihood = -6684.6592
Iteration 3:   log likelihood = -6660.1404
Iteration 4:   log likelihood = -6652.1368
Iteration 5:   log likelihood = -6651.7841
Iteration 6:   log likelihood = -6651.7819
Iteration 7:   log likelihood = -6651.7819
Generalized structural equation model          Number of obs    =       6535
Log likelihood = -6651.7819
 ( 1)  [2.chosen]M1[suspect] = 1
 ( 2)  [3.chosen]M2[suspect] = 1
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.chosen	(base outcome)					
2.chosen <-						
location						
suite_1	.3881676	.1004754	3.86	0.000	.1912394	.5850958
suite_2	.48938	.0960311	5.10	0.000	.3011625	.6775974
1.suswhite	-.0260152	.0749378	-0.35	0.728	-.1728906	.1208602
1.witmale	-.0007652	.0679187	-0.01	0.991	-.1338833	.132353
1.violent	.0369381	.0771594	0.48	0.632	-.1142915	.1881677
M1[suspect]	1	(constrained)				
_cons	-1.000382	.0992546	-10.08	0.000	-1.194918	-.8058469
3.chosen <-						
location						
suite_1	-.2904225	.0968578	-3.00	0.003	-.4802604	-.1005847
suite_2	.1364246	.089282	1.53	0.127	-.0385649	.3114142
1.suswhite	-.2437654	.0647275	-3.77	0.000	-.370629	-.1169018
1.witmale	.139826	.0596884	2.34	0.019	.0228389	.256813
1.violent	-1.388013	.0891863	-15.56	0.000	-1.562815	-1.213212
M2[suspect]	1	(constrained)				
_cons	.1750622	.0851614	2.06	0.040	.008149	.3419754
var(M1[suspect])	.2168248	.0549321			.131965	.3562533
var(M2[suspect])	.2978104	.0527634			.2104416	.421452
cov(M2[suspect], M1[suspect])	.2329749	.0438721	5.31	0.000	.1469872	.3189627

Notes:

1. The estimated variances of the two random effects are 0.2168 and 0.2978, which as explained in the second note of above example, are both practically and statistically significant.

2. The covariance is estimated to be 0.2300. Therefore, $0.2300/\sqrt{0.2168 \times 0.2978} = 0.9052$ is the estimated correlation.

3. Wright and Sparks (1994) were interested in whether the location of the lineup mattered. They found that it did, and that foils were more likely to be chosen at lineups outside of the police station (at the two "specialist" suites). They speculated the cause might be that the police at the station strongly warn witnesses against misidentification, or possibly because the specialist suites had better foils.

Fitting the model with the Builder

Use the first diagram in *Two-level multinomial logistic model with shared random effects* above for reference.

1. Open the dataset.

   In the Command window, type

   ```
   . use http://www.stata-press.com/data/r13/gsem_lineup
   ```

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 independent variables.

   Select the Add Observed Variables Set tool, , and then click near the bottom of the diagram about one-third of the way in from the left.

   In the resulting dialog box,

   a. select the *Select variables* radio button (it may already be selected);

   b. include the levels of the factor variable `location` by clicking on the ... button next to the *Variables* control. In the resulting dialog box, select the *Factor variable* radio button, select `Main effect` in the *Specification* control, and select `location` in the *Variables* control for *Variable 1*. Click on **Add to varlist**, and then click on **OK**;

   c. type `1.suswhite 1.witmale 1.violent` in the *Variables* control after `i.location` (typing `1.`*varname* rather than using the ... button to create them as `i.`*varname* factor variables prevents rectangles corresponding to the base categories for these binary variables from being created);

   d. select `Vertical` in the *Orientation* control;

   e. click on **OK**.

   If you wish, move the set of variables by clicking on any variable and dragging it.

5. Create the rectangles for the possible outcomes of the multinomial endogenous variable.

   Select the Add Observed Variables Set tool, , and then click in the diagram about one-third of the way in from the right and one-fourth of the way up from the bottom.

In the resulting dialog box,

a. select the *Select variables* radio button (it may already be selected);

b. check *Make variables generalized responses*;

c. select `Multinomial, Logit` in the *Family/Link* control;

d. select `chosen` in the *Variable* control;

e. under *Levels*, remove `1b` to prevent the rectangle corresponding to the base category from being created;

f. select `Vertical` in the *Orientation* control;

g. select the *Distances* tab;

h. select `.5` (inch) from the from the *Distance between variables* control;

i. click on **OK**.

6. Create the paths from the independent variables to the rectangles for outcomes `chosen` = 2 and `chosen` = 3.

   a. Select the Add Path tool, →.

   b. Click in the right side of the `1b.location` rectangle (it will highlight when you hover over it), and drag a path to the left side of the `2.chosen` rectangle (it will highlight when you can release to connect the path).

   c. Continuing with the → tool, click in the right side of each independent variable and drag a path to both the `2.chosen` and `3.chosen` rectangles.

7. Create the suspect-level latent variable.

   a. Select the Add Multilevel Latent Variable tool, , and click near the right side of the diagram, vertically centered between `2.chosen` and `3.chosen`.

   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 `suspect` > *Observations* in the next control.

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

   e. Click on **OK**.

8. Create the paths from the multilevel latent variable to the rectangles for outcomes `chosen` = 2 and `chosen` = 3.

   a. Select the Add Path tool, →.

   b. Click in the upper-left quadrant of the `suspect`$_1$ double oval, and drag a path to the right side of the `2.chosen` rectangle.

   c. Continuing with the → tool, click in the lower-left quadrant of the `suspect`$_1$ double oval, and drag a path to the right side of the `3.chosen` rectangle.

9. Place constraints on path coefficients from the multilevel latent variable.

   Use the Select tool, , to select the path from the `suspect`$_1$ double oval to the `2.chosen` rectangle. Type `1` in the β box in the Contextual Toolbar and press *Enter*. Repeat this process to constrain the coefficient on the path from the `suspect`$_1$ double oval to the `3.chosen` rectangle to `1`.

10. Clean up the location of the paths.

    If you do not like where the paths have been connected to the rectangles, use the Select tool, , to click on the path, and then simply click on where it connects to a rectangle and drag the endpoint.

11. Estimate.

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

12. If you wish to fit the model described in *Two-level multinomial logistic model with separate but correlated random effects*, use the Select tool to select the path from the `suspect`$_1$ double oval to the `3.chosen` rectangle in the diagram created above. Select **Object > Delete** from the SEM Builder menu.

    Using the Select tool, select the `suspect`$_1$ double oval and move it up so that it is parallel with the rectangle for `2.chosen`.

13. Create the multilevel latent variable corresponding to the random effects of suspect in the `3.chosen` equation.

    a. Select the Add Multilevel Latent Variable tool, , and click near the right side of the diagram, next to the `3.chosen` rectangle.

    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 `suspect` > *Observations* in the next control.

    d. Specify `M2` as the *Base name*.

    e. Click on **OK**.

14. Draw a path from the newly added suspect-level latent variable to `3.chosen`.

    Select the Add Path tool, click in the left of the `suspect`$_2$ double oval, and drag a path to the right side of the `3.chosen` rectangle.

15. Create the covariance between the random effects.

    a. Select the Add Covariance tool, .

    b. Click in the bottom-right quadrant of the `suspect`$_1$ double oval, and drag a covariance to the top right of the `suspect`$_2$ double oval.

16. Clean up paths and covariance.

    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 and drag the endpoint. Similarly, you can change where the covariance connects to the latent variables by clicking on the covariance and dragging the endpoint. You can also change the bow of the covariance by clicking on the covariance and dragging the control point that extends from one end of the selected covariance.

17. 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 first model in the Builder by typing

```
. webgetsem gsem_mlmlogit1
```

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

```
. webgetsem gsem_mlmlogit2
```

References

Skrondal, A., and S. Rabe-Hesketh. 2003. Multilevel logistic regression for polytomous data and rankings. *Psychometrika* 68: 267–287.

Wright, D. B., and A. T. Sparks. 1994. Using multilevel multinomial regression to analyse line-up data. *Multilevel Modelling Newsletter* 6: 8–10.

Also see

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

[SEM] **example 38g** — Random-intercept and random-slope models (multilevel)

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

[SEM] **intro 5** — Tour of models