mlogit — Multinomial (polytomous) logistic regression

Description

mlogit fits a multinomial logit (MNL) model for a categorical dependent variable with outcomes that have no natural ordering. The actual values taken by the dependent variable are irrelevant. The MNL model is also known as the polytomous logistic regression model. Some people refer to conditional logistic regression as multinomial logistic regression. If you are one of them, see [R] clogit.

Quick start

Multinomial logit model of y on x1, x2, and categorical variable a
mlogit y x1 x2 i.a

As above, but use y = 1 as the base outcome even if 1 is not the most frequent
mlogit y x1 x2 i.a, baseoutcome(1)

Report results as relative-risk ratios
mlogit y x1 x2 i.a, rrr

Constrain coefficient of x1 to be equal for second and third outcomes
constraint 1 [#2=#3]:x1
mlogit y x1 x2 i.a, constraints(1)

Menu

Statistics > Categorical outcomes > Multinomial logistic regression
Syntax

```
mlogit depvar [indepvars] [if] [in] [weight] [ , options]
```

<table>
<thead>
<tr>
<th>options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td></td>
</tr>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><code>baseoutcome(#)</code></td>
<td>value of depvar that will be the base outcome</td>
</tr>
<tr>
<td><code>constraints(clist)</code></td>
<td>apply specified linear constraints; clist has the form <code>#[-#] [, #[-#]] ...</code></td>
</tr>
<tr>
<td><code>collinear</code></td>
<td>keep collinear variables</td>
</tr>
<tr>
<td>SE/Robust</td>
<td></td>
</tr>
<tr>
<td><code>vce(vcetype)</code></td>
<td>vcetype may be oim, robust, cluster clustvar, bootstrap, or jackknife</td>
</tr>
<tr>
<td>Reporting</td>
<td></td>
</tr>
<tr>
<td><code>level(#)</code></td>
<td>set confidence level; default is level(95)</td>
</tr>
<tr>
<td><code>rrr</code></td>
<td>report relative-risk ratios</td>
</tr>
<tr>
<td><code>nocnsreport</code></td>
<td>do not display constraints</td>
</tr>
<tr>
<td><code>display_options</code></td>
<td>control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
<tr>
<td>Maximization</td>
<td></td>
</tr>
<tr>
<td><code>maximize_options</code></td>
<td>control the maximization process; seldom used</td>
</tr>
<tr>
<td><code>coeflegend</code></td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>

`indepvars` may contain factor variables; see [U] 11.4.3 Factor variables.
`indepvars` may contain time-series operators; see [U] 11.4.4 Time-series varlists.
`bayes`, bootstrap, by, fmm, fp, jackknife, mfp, mi estimate, rolling, statsby, and svy are allowed; see [U] 11.1.10 Prefix commands. For more details, see [BAYES] bayes: mlogit and [FMM] fmm: mlogit.
`vce(bootstrap)` and `vce(jackknife)` are not allowed with the mi estimate prefix; see [MI] mi estimate.
Weights are not allowed with the bootstrap prefix; see [R] bootstrap.
`vce()` and weights are not allowed with the svy prefix; see [SVY] svy.
fweights, iweights, and pweights are allowed; see [U] 11.1.6 weight.
`coeflegend` does not appear in the dialog box.
See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

Options

- `noconstant`; see [R] estimation options.
- `baseoutcome(#)`: specifies the value of depvar to be treated as the base outcome. The default is to choose the most frequent outcome.
- `constraints(clist), collinear`; see [R] estimation options.
vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim), that are robust to some kinds of misspecification (robust), that allow for intragroup correlation (cluster clustvar), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [R] vce_option.

If specifying vce(bootstrap) or vce(jackknife), you must also specify baseoutcome().

Reporting

level(#) reports the estimated coefficients transformed to relative-risk ratios, that is, \( e^b \) rather than \( b \); see Description of the model below for an explanation of this concept. Standard errors and confidence intervals are similarly transformed. This option affects how results are displayed, not how they are estimated. rrr may be specified at estimation or when replaying previously estimated results.

display_options: noci, nopvalues, noomitted, vsquish, noemptycells, baselevels, allbaselevels, nolabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

Maximization

maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs); see [R] maximize. These options are seldom used.

The following option is available with mlogit but is not shown in the dialog box:

coeflegend; see [R] estimation options.

Remarks and examples

Remarks are presented under the following headings:

Description of the model
Fitting unconstrained models
Fitting constrained models

mlogit fits maximum likelihood models with discrete dependent (left-hand-side) variables when the dependent variable takes on more than two outcomes and the outcomes have no natural ordering. If the dependent variable takes on only two outcomes, estimates are identical to those produced by logistic or logit; see [R] logistic or [R] logit. If the outcomes are ordered, see [R] ologit. See [R] logistic for a list of related estimation commands.

Description of the model

Consider the outcomes 1, 2, 3, \ldots, \textit{m} recorded in \textit{y}, and the explanatory variables \textit{X}. Assume that there are \textit{m} = 3 outcomes: “buy an American car”, “buy a Japanese car”, and “buy a European car”. The values of \textit{y} are then said to be “unordered”. Even though the outcomes are coded 1, 2, and 3, the numerical values are arbitrary because 1 < 2 < 3 does not imply that outcome 1 (buy American) is less than outcome 2 (buy Japanese) is less than outcome 3 (buy European). This unordered categorical property of \textit{y} distinguishes the use of \texttt{mlogit} from \texttt{regress} (which is appropriate for a continuous dependent variable), from \texttt{ologit} (which is appropriate for ordered categorical data), and from \texttt{logit} (which is appropriate for two outcomes, which can be thought of as ordered).

In the multinomial logit model, you estimate a set of coefficients, \(\beta(1)\), \(\beta(2)\), and \(\beta(3)\), corresponding to each outcome:

\[
\begin{align*}
\Pr(y = 1) &= \frac{e^{X\beta(1)}}{e^{X\beta(1)} + e^{X\beta(2)} + e^{X\beta(3)}} \\
\Pr(y = 2) &= \frac{e^{X\beta(2)}}{e^{X\beta(1)} + e^{X\beta(2)} + e^{X\beta(3)}} \\
\Pr(y = 3) &= \frac{e^{X\beta(3)}}{e^{X\beta(1)} + e^{X\beta(2)} + e^{X\beta(3)}}
\end{align*}
\]

The model, however, is unidentified in the sense that there is more than one solution to \(\beta(1)\), \(\beta(2)\), and \(\beta(3)\) that leads to the same probabilities for \(y = 1\), \(y = 2\), and \(y = 3\). To identify the model, you arbitrarily set one of \(\beta(1)\), \(\beta(2)\), or \(\beta(3)\) to 0—it does not matter which. That is, if you arbitrarily set \(\beta(1) = 0\), the remaining coefficients \(\beta(2)\) and \(\beta(3)\) will measure the change relative to the \(y = 1\) group. If you instead set \(\beta(2) = 0\), the remaining coefficients \(\beta(1)\) and \(\beta(3)\) will measure the change relative to the \(y = 2\) group. The coefficients will differ because they have different interpretations, but the predicted probabilities for \(y = 1\), \(y = 2\), and \(y = 3\) will still be the same. Thus either parameterization will be a solution to the same underlying model.

Setting \(\beta(1) = 0\), the equations become

\[
\begin{align*}
\Pr(y = 1) &= \frac{1}{1 + e^{X\beta(2)} + e^{X\beta(3)}} \\
\Pr(y = 2) &= \frac{e^{X\beta(2)}}{1 + e^{X\beta(2)} + e^{X\beta(3)}} \\
\Pr(y = 3) &= \frac{e^{X\beta(3)}}{1 + e^{X\beta(2)} + e^{X\beta(3)}}
\end{align*}
\]

The relative probability of \(y = 2\) to the base outcome is

\[
\frac{\Pr(y = 2)}{\Pr(y = 1)} = e^{X\beta(2)}
\]

Let’s call this ratio the relative risk, and let’s further assume that \(X\) and \(\beta^{(2)}_k\) are vectors equal to \((x_1, x_2, \ldots, x_k)\) and \((\beta^{(2)}_1, \beta^{(2)}_2, \ldots, \beta^{(2)}_k)'\), respectively. The ratio of the relative risk for a one-unit change in \(x_i\) is then

\[
\frac{e^{\beta^{(2)}_1 x_1 + \cdots + \beta^{(2)}_i (x_i + 1) + \cdots + \beta^{(2)}_k x_k}}{e^{\beta^{(2)}_1 x_1 + \cdots + \beta^{(2)}_i x_i + \cdots + \beta^{(2)}_k x_k}} = e^{\beta^{(2)}_i}
\]
Thus the exponentiated value of a coefficient is the relative-risk ratio for a one-unit change in the corresponding variable (risk is measured as the risk of the outcome relative to the base outcome).

### Fitting unconstrained models

#### Example 1: A first example

We have data on the type of health insurance available to 616 psychologically depressed subjects in the United States (Tarlov et al. 1989; Wells et al. 1989). The insurance is categorized as either an indemnity plan (that is, regular fee-for-service insurance, which may have a deductible or coinsurance rate) or a prepaid plan (a fixed up-front payment allowing subsequent unlimited use as provided, for instance, by an HMO). The third possibility is that the subject has no insurance whatsoever. We wish to explore the demographic factors associated with each subject’s insurance choice. One of the demographic factors in our data is the race of the participant, coded as white or nonwhite:

```
use http://www.stata-press.com/data/r15/sysdsn1
(Health insurance data)
.tabulate insure nonwhite, chi2 col
```

<table>
<thead>
<tr>
<th></th>
<th>frequency</th>
<th>column percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>nonwhite</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Indemnity</td>
<td>251</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>50.71</td>
<td>35.54</td>
</tr>
<tr>
<td>Prepaid</td>
<td>208</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>42.02</td>
<td>57.02</td>
</tr>
<tr>
<td>Uninsured</td>
<td>36</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>7.27</td>
<td>7.44</td>
</tr>
<tr>
<td>Total</td>
<td>495</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Pearson chi2(2) = 9.5599 Pr = 0.008

Although `insure` appears to take on the values *Indemnity*, *Prepaid*, and *Uninsured*, it actually takes on the values 1, 2, and 3. The words appear because we have associated a value label with the numeric variable `insure`; see [U] 12.6.3 Value labels.
When we fit a multinomial logit model, we can tell `mlogit` which outcome to use as the base outcome, or we can let `mlogit` choose. To fit a model of `insure` on `nonwhite`, letting `mlogit` choose the base outcome, we type

```
. mlogit insure nonwhite
```

```
Iteration 0: log likelihood = -556.59502
Iteration 1: log likelihood = -551.78935
Iteration 2: log likelihood = -551.78348
Iteration 3: log likelihood = -551.78348
```

Multinomial logistic regression

```
Number of obs = 616
LR chi2(2) = 9.62
Prob > chi2 = 0.0081
Log likelihood = -551.78348
Pseudo R2 = 0.0086
```

```
insure | Coef. Std. Err.   z    P>|z|   [95% Conf. Interval]
-------|------------------|-------|-------|------------------|--------|
        | (base outcome)   |       |       |                  |         |
Indemnity|                   |       |       |                  |         |
Prepaid  |                   |       |       |                  |         |
 nonwhite|      0.6608212    | 0.2157321  | 3.06  | 0.002   | 0.2379942    | 1.083648|
     _cons|      -0.1879149  | 0.0937644  | -2.00 | 0.045   | -0.3716896   | -0.0041401|
Uninsured|                   |       |       |                  |         |
 nonwhite|      0.3779586    | 0.407589  | 0.93  | 0.354   | -0.4209011   | 1.176818|
     _cons|     -1.941934     | 0.1782185 | -10.90| 0.000   | -2.291236    | -1.592632|
```

`mlogit` chose the indemnity outcome as the base outcome and presented coefficients for the outcomes prepaid and uninsured. According to the model, the probability of prepaid for whites (\(\text{nonwhite} = 0\)) is

\[
Pr(\text{insure} = \text{Prepaid}) = \frac{e^{-0.188}}{1 + e^{-0.188} + e^{-1.942}} = 0.420
\]

Similarly, for nonwhites, the probability of prepaid is

\[
Pr(\text{insure} = \text{Prepaid}) = \frac{e^{-0.188 + 0.661}}{1 + e^{-0.188 + 0.661} + e^{-1.942 + 0.378}} = 0.570
\]

These results agree with the column percentages presented by `tabulate` because the `mlogit` model is fully saturated. That is, there are enough terms in the model to fully explain the column percentage in each cell. The model chi-squared and the `tabulate` chi-squared are in almost perfect agreement; both test that the column percentages of `insure` are the same for both values of `nonwhite`.

> **Example 2: Specifying the base outcome**

By specifying the `baseoutcome()` option, we can control which outcome of the dependent variable is treated as the base. Left to its own, `mlogit` chose to make outcome 1, indemnity, the base outcome. To make outcome 2, prepaid, the base, we would type
. mlogit insure nonwhite, base(2)

## Multinomial Logistic Regression

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-556.59502</td>
</tr>
<tr>
<td>1</td>
<td>-551.78935</td>
</tr>
<tr>
<td>2</td>
<td>-551.78348</td>
</tr>
<tr>
<td>3</td>
<td>-551.78348</td>
</tr>
</tbody>
</table>

### Multinomial Logistic Regression

- **Number of obs**: 616
- **LR chi2(2)**: 9.62
- **Prob > chi2**: 0.0081
- **Log likelihood**: -551.78348
- **Pseudo R2**: 0.0086

| insure   | Coef.  | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|----------|--------|-----------|------|------|----------------------|
| Indemnity |        |           |      |      |                      |
| nonwhite | -.6608212 | .2157321  | -3.06| 0.002| -1.083648 -.2379942  |
| _cons    | .1879149 | .0937644  | 2.00 | 0.045| .0041401 .3716896    |
| Prepaid   | (base outcome) | | | | |
| Uninsure  |        |           |      |      |                      |
| nonwhite | -.2828627 | .3977302  | -0.71| 0.477| -1.0624 .4966742    |
| _cons    | -1.754019 | .1805145  | -9.72| 0.000| -2.107821 -1.400217  |

The `baseoutcome()` option requires that we specify the numeric value of the outcome, so we could not type `base(Prepaid)`. Although the coefficients now appear to be different, the summary statistics reported at the top are identical. With this parameterization, the probability of prepaid insurance for whites is:

\[
\Pr(\text{insure} = \text{Prepaid}) = \frac{1}{1 + e^{1.188} + e^{-1.754}} = 0.420
\]

This is the same answer we obtained previously.

### Example 3: Displaying Relative-Risk Ratios

By specifying `rrr`, which we can do at estimation time or when we redisplay results, we see the model in terms of relative-risk ratios:

. mlogit, rrr

## Multinomial Logistic Regression

<table>
<thead>
<tr>
<th>Number of obs</th>
<th>616</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR chi2(2)</td>
<td>9.62</td>
</tr>
<tr>
<td>Prob &gt; chi2</td>
<td>0.0081</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-551.78348</td>
</tr>
<tr>
<td>Pseudo R2</td>
<td>0.0086</td>
</tr>
</tbody>
</table>

| insure   | RRR  | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|----------|------|-----------|------|------|----------------------|
| Indemnity |      |           |      |      |                      |
| nonwhite | .516427 | .1114099  | -3.06| 0.002| .3383588 .7882073    |
| _cons    | 1.206731 | .1131483  | 2.00 | 0.045| 1.004149 1.450183   |
| Prepaid   | (base outcome) | | | | | |
| Uninsure  |      |           |      |      |                      |
| nonwhite | .7536233 | .2997387  | -0.71| 0.477| .3456255 1.643247   |
| _cons    | .1730769 | .0312429  | -9.72| 0.000| .1215024 .2465434   |

Note: `_cons` estimates baseline relative risk for each outcome.
Looked at this way, the relative risk of choosing an indemnity over a prepaid plan is 0.516 for nonwhites relative to whites.

To illustrate, from the output and discussions of examples 1 and 2 we find that

$$\frac{\Pr(\text{insure} = \text{Indemnity} \mid \text{white})}{\Pr(\text{insure} = \text{Prepaid} \mid \text{white})} = \frac{1}{1 + e^{-1.88} + e^{-1.942}} = 0.507$$

and thus the relative risk of choosing indemnity over prepaid (for whites) is

$$\frac{\Pr(\text{insure} = \text{Indemnity} \mid \text{white})}{\Pr(\text{insure} = \text{Prepaid} \mid \text{white})} = \frac{0.507}{0.420} = 1.207$$

For nonwhites,

$$\frac{\Pr(\text{insure} = \text{Indemnity} \mid \text{not white})}{\Pr(\text{insure} = \text{Prepaid} \mid \text{not white})} = \frac{1}{1 + e^{-1.88+.661} + e^{-1.942+.378}} = 0.355$$

and thus the relative risk of choosing indemnity over prepaid (for nonwhites) is

$$\frac{\Pr(\text{insure} = \text{Indemnity} \mid \text{not white})}{\Pr(\text{insure} = \text{Prepaid} \mid \text{not white})} = \frac{0.355}{0.570} = 0.623$$

The ratio of these two relative risks, hence the name “relative-risk ratio”, is $0.623/1.207 = 0.516$, as given in the output under the heading “RRR”.

Technical note

In models where only two categories are considered, the `mlogit` model reduces to standard `logit`. Consequently the exponentiated regression coefficients, labeled as RRR within `mlogit`, are equal to the odds ratios as given when the `or` option is specified under `logit`; see `[R] logit`.

As such, always referring to `mlogit`’s exponentiated coefficients as odds ratios may be tempting. However, the discussion in example 3 demonstrates that doing so would be incorrect. In general `mlogit` models, the exponentiated coefficients are ratios of relative risks, not ratios of odds.

Example 4: Model with continuous and multiple categorical variables

One of the advantages of `mlogit` over `tabulate` is that we can include continuous variables and multiple categorical variables in the model. In examining the data on insurance choice, we decide that we want to control for age, gender, and site of study (the study was conducted in three sites):
. mlogit insure age male nonwhite i.site

Iteration 0:  log likelihood = -555.85446
Iteration 1:  log likelihood = -534.67443
Iteration 2:  log likelihood = -534.36284
Iteration 3:  log likelihood = -534.36165
Iteration 4:  log likelihood = -534.36165

Multinomial logistic regression

Number of obs   = 615
LR chi2(10)     = 42.99
Prob > chi2     = 0.0000
Log likelihood  = -534.36165
Pseudo R2       =  0.0387

insure         Coef.  Std. Err.    z    P>|z|     [95% Conf. Interval]

Indemnity      (base outcome)

Prepaid
    age    -.011745   .0061946  -1.90  0.058    -.0238862   .0003962
    male    .5616934   .2027465   2.77  0.006    .1643175   .9590693
    nonwhite    .9747768   .2363213   4.12  0.000    .5115955   1.437958

    site
       2    .1130359   .2101903   0.54  0.591   -.2989296   .5250013
       3   -.5879879   .2279351  -2.58  0.010   -1.034733  -.1412433
    _cons    .2697127   .3284422   0.82  0.412   -.3740222   .9134476

Uninsure
    age   -.0077961   .0114418  -0.68  0.496   -.0302217   .0146294
    male    .4518496   .3674867   1.23  0.219   -.2684111   1.17211
    nonwhite    .2170589   .4256361   0.51  0.610   -.6171725   1.05129

    site
       2   -1.211563   .4705127  -2.57  0.010   -.2133751  -.2893747
       3  -.2078123   .3662926  -0.57  0.570   -.9257327   .510108
    _cons   -1.286943   .5923219  -2.17  0.030   -.2447872  -.1260134

These results suggest that the inclination of nonwhites to choose prepaid care is even stronger than it was without controlling. We also see that subjects in site 2 are less likely to be uninsured.

Fitting constrained models

mlogit can fit models with subsets of coefficients constrained to be zero, with subsets of coefficients constrained to be equal both within and across equations, and with subsets of coefficients arbitrarily constrained to equal linear combinations of other estimated coefficients.

Before fitting a constrained model, you define the constraints with the constraint command; see [R] constraint. Once the constraints are defined, you estimate using mlogit, specifying the constraint() option. Typing constraint(4) would use the constraint you previously saved as 4. Typing constraint(1,4,6) would use the previously stored constraints 1, 4, and 6. Typing constraint(1-4,6) would use the previously stored constraints 1, 2, 3, 4, and 6.

Sometimes you will not be able to specify the constraints without knowing the omitted outcome. In such cases, assume that the omitted outcome is whatever outcome is convenient for you, and include the baseoutcome() option when you specify the mlogit command.
Example 5: Specifying constraints to test hypotheses

We can use constraints to test hypotheses, among other things. In our insurance-choice model, let's test the hypothesis that there is no distinction between having indemnity insurance and being uninsured. Indemnity-style insurance was the omitted outcome, so we type

```
   . test [Uninsure]
   ( 1) [Uninsure]age = 0
   ( 2) [Uninsure]male = 0
   ( 3) [Uninsure]nonwhite = 0
   ( 4) [Uninsure]1b.site = 0
   ( 5) [Uninsure]2.site = 0
   ( 6) [Uninsure]3.site = 0
   Constraint 4 dropped
   chi2(  5) =  9.31
   Prob > chi2 =  0.0973
```

If indemnity had not been the omitted outcome, we would have typed `test [Uninsure=Indemnity].`

The results produced by `test` are an approximation based on the estimated covariance matrix of the coefficients. Because the probability of being uninsured is low, the log likelihood may be nonlinear for the uninsured. Conventional statistical wisdom is not to trust the asymptotic answer under these circumstances but to perform a likelihood-ratio test instead.

To use Stata’s `lrtest` (likelihood-ratio test) command, we must fit both the unconstrained and constrained models. The unconstrained model is the one we have previously fit. Following the instruction in [R] `lrtest`, we first store the unconstrained model results:

```
   . estimates store unconstrained
```

To fit the constrained model, we must refit our model with all the coefficients except the constant set to 0 in the `Uninsure` equation. We define the constraint and then refit:

```
```
. constraint 1 [Uninsure]
. mlogit insure age male nonwhite i.site, constraints(1)
Iteration 0:  log likelihood =  -555.85446
Iteration 1:  log likelihood =  -539.80523
Iteration 2:  log likelihood =  -539.75644
Iteration 3:  log likelihood =  -539.75643

Multinomial logistic regression  Number of obs =  615
Log likelihood =  -539.75643  Wald chi2(5) =  29.70

( 1)  [Uninsure]o.age = 0
( 2)  [Uninsure]o.male = 0
( 3)  [Uninsure]o.nonwhite = 0
( 4)  [Uninsure]2o.site = 0
( 5)  [Uninsure]3o.site = 0

          Coef.  Std. Err.     z  P>|z|      [95% Conf. Interval]
indemnity (base outcome)
    prepaid
        age  -.0107025  .0060039  -1.78  0.075   -.0224699   .0010649
        male   .4963616  .1939683   2.56  0.010   .1161907   .8765324
        nonwhite   .9421369  .2252094   4.18  0.000   .5007346   1.383539
        site
            2  .2530912  .2029465   1.25  0.212  -.1446767   .6508591
            3 -.5521773  .2187237  -2.52  0.012  -.9808678  -.1234869
    _cons   .1792752  .3171372   0.57  0.572  -.4423023   .8008527

uninsure
        age   0 (omitted)
        male   0 (omitted)
        nonwhite   0 (omitted)
        site
            2  0 (omitted)
            3  0 (omitted)
    _cons  -1.87351  .1601099  -11.70  0.000  -2.18732  -1.5597

We can now perform the likelihood-ratio test:
. lrtest unconstrained

Likelihood-ratio test     LR chi2(5) =   10.79
(Assumption: . nested in unconstrained)  Prob > chi2 =  0.0557

The likelihood-ratio chi-squared is 10.79 with 5 degrees of freedom—just slightly greater than the magic \( p = 0.05 \) level—so we should not call this difference significant.

Technical note

In certain circumstances, you should fit a multinomial logit model with conditional logit; see \([R] clogit\). With substantial data manipulation, clogit can handle the same class of models with some interesting additions. For example, if we had available the price and deductible of the most competitive insurance plan of each type, mlogit could not use this information, but clogit could.
mlogit stores the following in \texttt{e()}: 

Scalars
\begin{itemize}
\item \texttt{e(N)} \quad \text{number of observations}
\item \texttt{e(N_cd)} \quad \text{number of completely determined observations}
\item \texttt{e(k_out)} \quad \text{number of outcomes}
\item \texttt{e(k)} \quad \text{number of parameters}
\item \texttt{e(k_eq)} \quad \text{number of equations in \texttt{e(b)}}
\item \texttt{e(k_eq_model)} \quad \text{number of equations in overall model test}
\item \texttt{e(k_dv)} \quad \text{number of dependent variables}
\item \texttt{e(df_m)} \quad \text{model degrees of freedom}
\item \texttt{e(r2_p)} \quad \text{pseudo-R-squared}
\item \texttt{e(ll)} \quad \text{log likelihood}
\item \texttt{e(ll0)} \quad \text{log likelihood, constant-only model}
\item \texttt{e(N_clust)} \quad \text{number of clusters}
\item \texttt{e(chi2)} \quad \chi^2 \text{ model test}
\item \texttt{e(p)} \quad \text{p-value for model test}
\item \texttt{e(k_eq_base)} \quad \text{equation number of the base outcome}
\item \texttt{e(baseout)} \quad \text{the value of \texttt{depvar} to be treated as the base outcome}
\item \texttt{e(baseout)} \quad \text{index of the base outcome}
\item \texttt{e(rank)} \quad \text{rank of \texttt{e(V)}}
\item \texttt{e(ic)} \quad \text{number of iterations}
\item \texttt{e(rc)} \quad \text{return code}
\item \texttt{e(converged)} \quad 1 \text{ if converged, 0 otherwise}
\end{itemize}

Macros
\begin{itemize}
\item \texttt{e(cmd)} \quad \texttt{mlogit}
\item \texttt{e(cmdline)} \quad \text{command as typed}
\item \texttt{e(depvar)} \quad \text{name of dependent variable}
\item \texttt{e(wtype)} \quad \text{weight type}
\item \texttt{e(wexp)} \quad \text{weight expression}
\item \texttt{e(title)} \quad \text{title in estimation output}
\item \texttt{e(clustvar)} \quad \text{name of cluster variable}
\item \texttt{e(chi2type)} \quad \text{Wald or LR; type of model \chi^2 \text{ test}}
\item \texttt{e(vcetype)} \quad \text{vcetype specified in \texttt{vce()}}
\item \texttt{e(eqnames)} \quad \text{names of equations}
\item \texttt{e(baselab)} \quad \text{value label corresponding to base outcome}
\item \texttt{e(opt)} \quad \text{type of optimization}
\item \texttt{e(which)} \quad \text{max or min; whether optimizer is to perform maximization or minimization}
\item \texttt{e(ml_method)} \quad \text{type of \texttt{ml} method}
\item \texttt{e(user)} \quad \text{name of likelihood-evaluator program}
\item \texttt{e(technique)} \quad \text{maximization technique}
\item \texttt{e(properties)} \quad \text{b V}
\item \texttt{e(predict)} \quad \text{program used to implement \texttt{predict}}
\item \texttt{e(marginsnotok)} \quad \text{predictions disallowed by \texttt{margins}}
\item \texttt{e(marginsdefault)} \quad \text{default \texttt{predict()} specification for \texttt{margins}}
\item \texttt{e(asbalanced)} \quad \text{factor variables \texttt{fvset} as \texttt{asbalanced}}
\item \texttt{e(asobserved)} \quad \text{factor variables \texttt{fvset} as \texttt{asobserved}}
\end{itemize}
Matrices
- \( e(b) \): coefficient vector
- \( e(out) \): outcome values
- \( e(Cns) \): constraints matrix
- \( e(i/log) \): iteration log (up to 20 iterations)
- \( e(gradient) \): gradient vector
- \( e(V) \): variance–covariance matrix of the estimators
- \( e(V_{modelbased}) \): model-based variance

Functions
- \( e(sample) \): marks estimation sample

Methods and formulas

The multinomial logit model is described in Greene (2018, 829–833).

Suppose that there are \( k \) categorical outcomes and—without loss of generality—let the base outcome be 1. The probability that the response for the \( j \)th observation is equal to the \( i \)th outcome is

\[
p_{ij} = \Pr(y_j = i) = \begin{cases} 
1, & \text{if } i = 1 \\
1 + \sum_{m=2}^{k} \exp(x_j \beta_m), & \text{if } i > 1
\end{cases}
\]

where \( x_j \) is the row vector of observed values of the independent variables for the \( j \)th observation and \( \beta_m \) is the coefficient vector for outcome \( m \). The log pseudolikelihood is

\[
\ln L = \sum_j w_j \sum_{i=1}^{k} I_i(y_j) \ln p_{ik}
\]

where \( w_j \) is an optional weight and

\[
I_i(y_j) = \begin{cases} 
1, & \text{if } y_j = i \\
0, & \text{otherwise}
\end{cases}
\]

Newton–Raphson maximum likelihood is used; see [R] maximize.

For constrained equations, the set of constraints is orthogonalized, and a subset of maximizable parameters is selected. For example, a parameter that is constrained to zero is not a maximizable parameter. If two parameters are constrained to be equal to each other, only one is a maximizable parameter.

Let \( r \) be the vector of maximizable parameters. \( r \) is physically a subset of the solution parameters, \( b \). A matrix, \( T \), and a vector, \( m \), are defined as

\[
b = Tr + m
\]
so that

\[ \frac{\partial f}{\partial \mathbf{b}} = \frac{\partial f}{\partial \mathbf{r}} \mathbf{T}' \]

\[ \frac{\partial^2 f}{\partial \mathbf{b}^2} = \mathbf{T} \frac{\partial^2 f}{\partial \mathbf{r}^2} \mathbf{T}' \]

\( \mathbf{T} \) consists of a block form in which one part is a permutation of the identity matrix and the other part describes how to calculate the constrained parameters from the maximizable parameters.

This command supports the Huber/White/sandwich estimator of the variance and its clustered version using `vce(robust)` and `vce(cluster clustvar)`, respectively. See \([P] \_\text{robust}\), particularly \textit{Maximum likelihood estimators} and \textit{Methods and formulas}.

\texttt{mlogit} also supports estimation with survey data. For details on VCEs with survey data, see \([SVY] \text{ variance estimation}\).

---

**References**


Also see

[R] mlogit postestimation — Postestimation tools for mlogit
[R] clogit — Conditional (fixed-effects) logistic regression
[R] logistic — Logistic regression, reporting odds ratios
[R] logit — Logistic regression, reporting coefficients
[R] mprobit — Multinomial probit regression
[R] nlogit — Nested logit regression
[R] ologit — Ordered logistic regression
[R] rologit — Rank-ordered logistic regression
[R] slogit — Stereotype logistic regression
[BAYES] bayes: mlogit — Bayesian multinomial logistic regression
[FMM] fmm: mlogit — Finite mixtures of multinomial (polytomous) logistic regression models
[MI] estimation — Estimation commands for use with mi estimate
[SVY] svy estimation — Estimation commands for survey data
[U] 20 Estimation and postestimation commands