# Title

> **mprobit** — Multinomial probit regression

Syntax            Menu            Description            Options
Remarks and examples    Stored results    Methods and formulas    References
Also see

# Syntax

mprobit *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, *options*]

| *options* | Description |
|---|---|
| **Model** | |
| <u>nocon</u>stant | suppress constant terms |
| <u>base</u>outcome(*#* \| *lbl*) | outcome used to normalize location |
| probitparam | use the probit variance parameterization |
| <u>constraints</u>(*constraints*) | apply specified linear constraints |
| collinear | keep collinear variables |
| **SE/Robust** | |
| vce(*vcetype*) | *vcetype* may be oim, <u>r</u>obust, <u>cl</u>uster *clustvar*, opg, bootstrap, or jackknife |
| **Reporting** | |
| <u>l</u>evel(*#*) | set confidence level; default is level(95) |
| nocnsreport | do not display constraints |
| *display_options* | control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling |
| **Integration** | |
| intpoints(*#*) | number of quadrature points |
| **Maximization** | |
| *maximize_options* | control the maximization process; seldom used |
| coeflegend | display legend instead of statistics |

*indepvars* may contain factor variables; see [U] **11.4.3 Factor variables**.

bootstrap, by, fp, jackknife, mi estimate, rolling, statsby, and svy are allowed; see [U] **11.1.10 Prefix commands**.

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.

## Menu

## Description

mprobit fits multinomial probit (MNP) models via maximum likelihood. *depvar* contains the outcome for each observation, and *indepvars* are the associated covariates. The error terms are assumed to be independent, standard normal, random variables. See [R] **asmprobit** for the case where the latent-variable errors are correlated or heteroskedastic and you have alternative-specific variables.

## Options

###### Model

noconstant suppresses the $J - 1$ constant terms.

baseoutcome(# | *lbl*) specifies the outcome used to normalize the location of the latent variable. The base outcome may be specified as a number or a label. The default is to use the most frequent outcome. The coefficients associated with the base outcome are zero.

probitparam specifies to use the probit variance parameterization by fixing the variance of the differenced latent errors between the scale and the base alternatives to be one. The default is to make the variance of the base and scale latent errors one, thereby making the variance of the difference to be two.

constraints(*constraints*), collinear; see [R] **estimation options**.

###### SE/Robust

vce(*vcetype*) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim, opg), 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(#); see [R] **estimation options**.

nocnsreport; see [R] **estimation options**.

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

###### Integration

intpoints(#) specifies the number of Gaussian quadrature points to use in approximating the likelihood. The default is 15.

Maximization

*maximize_options*: <u>di</u>fficult, <u>tech</u>nique(*algorithm_spec*), <u>iter</u>ate(*#*), [<u>no</u>]<u>log</u>, <u>trace</u>,
<u>grad</u>ient, <u>showstep</u>, <u>hess</u>ian, <u>showtol</u>erance, <u>tol</u>erance(*#*), <u>ltol</u>erance(*#*),
<u>nrtol</u>erance(*#*), <u>nonrtol</u>erance, and <u>from</u>(*init_specs*); see [R] **maximize**. These options are
seldom used.

Setting the optimization type to technique(bhhh) resets the default *vcetype* to vce(opg).

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

<u>coefl</u>egend; see [R] **estimation options**.

# Remarks and examples

The MNP model is used with discrete dependent variables that take on more than two outcomes
that do not have a natural ordering. The stochastic error terms for this implementation of the model
are assumed to have independent, standard normal distributions. To use mprobit, you must have one
observation for each decision maker in the sample. See [R] **asmprobit** for another implementation of
the MNP model that permits correlated and heteroskedastic errors and is suitable when you have data
for each alternative that a decision maker faced.

The MNP model is frequently motivated using a latent-variable framework. The latent variable for
the $j$th alternative, $j = 1, \ldots, J$, is

$$\eta_{ij} = \mathbf{z}_i \boldsymbol{\alpha}_j + \xi_{ij}$$

where the $1 \times q$ row vector $\mathbf{z}_i$ contains the observed independent variables for the $i$th decision maker.
Associated with $\mathbf{z}_i$ are the $J$ vectors of regression coefficients $\boldsymbol{\alpha}_j$. The $\xi_{i,1}, \ldots, \xi_{i,J}$ are distributed
independently and identically standard normal. The decision maker chooses the alternative $k$ such
that $\eta_{ik} \geq \eta_{im}$ for $m \neq k$.

Suppose that case $i$ chooses alternative $k$, and take the difference between latent variable $\eta_{ik}$ and
the $J - 1$ others:

$$
\begin{aligned}
v_{ijk} &= \eta_{ij} - \eta_{ik} \\
&= \mathbf{z}_i (\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_k) + \xi_{ij} - \xi_{ik} \\
&= \mathbf{z}_i \boldsymbol{\gamma}_{j'} + \epsilon_{ij'}
\end{aligned}
\tag{1}
$$

where $j' = j$ if $j < k$ and $j' = j - 1$ if $j > k$ so that $j' = 1, \ldots, J - 1$. $\mathrm{Var}(\epsilon_{ij'}) = \mathrm{Var}(\xi_{ij} - \xi_{ik}) = 2$
and $\mathrm{Cov}(\epsilon_{ij'}, \epsilon_{il'}) = 1$ for $j' \neq l'$. The probability that alternative $k$ is chosen is

$$
\begin{aligned}
\mathrm{Pr}(i \text{ chooses } k) &= \mathrm{Pr}(v_{i1k} \leq 0, \ldots, v_{i,J-1,k} \leq 0) \\
&= \mathrm{Pr}(\epsilon_{i1} \leq -\mathbf{z}_i \boldsymbol{\gamma}_1, \ldots, \epsilon_{i,J-1} \leq -\mathbf{z}_i \boldsymbol{\gamma}_{J-1})
\end{aligned}
$$

Hence, evaluating the likelihood function involves computing probabilities from the multivariate
normal distribution. That all the covariances are equal simplifies the problem somewhat; see *Methods
and formulas* for details.

In (1), not all $J$ of the $\boldsymbol{\alpha}_j$ are identifiable. To remove the indeterminacy, $\boldsymbol{\alpha}_l$ is set to the zero vector,
where $l$ is the base outcome as specified in the baseoutcome() option. That fixes the $l$th latent
variable to zero so that the remaining variables measure the attractiveness of the other alternatives
relative to the base.

▷ Example 1

As discussed in example 1 of [R] **mlogit**, 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). Patients may have either an indemnity (fee-for-service) plan or a prepaid plan such as an HMO, or the patient may be uninsured. Demographic variables include age, gender, race, and site. Indemnity insurance is the most popular alternative, so mprobit will choose it as the base outcome by default.

```
. use http://www.stata-press.com/data/r13/sysdsn1
(Health insurance data)

. mprobit insure age male nonwhite i.site

Iteration 0:   log likelihood = -535.89424
Iteration 1:   log likelihood = -534.56173
Iteration 2:   log likelihood = -534.52835
Iteration 3:   log likelihood = -534.52833

Multinomial probit regression                  Number of obs   =        615
                                               Wald chi2(10)   =      40.18
Log likelihood = -534.52833                    Prob > chi2     =     0.0000
```

| insure | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| Indemnity | (base outcome) | | | | | |
| **Prepaid** | | | | | | |
| age | -.0098536 | .0052688 | -1.87 | 0.061 | -.0201802 | .000473 |
| male | .4774678 | .1718316 | 2.78 | 0.005 | .1406841 | .8142515 |
| nonwhite | .8245003 | .1977582 | 4.17 | 0.000 | .4369013 | 1.212099 |
| **site** | | | | | | |
| 2 | .0973956 | .1794546 | 0.54 | 0.587 | -.2543289 | .4491201 |
| 3 | -.495892 | .1904984 | -2.60 | 0.009 | -.869262 | -.1225221 |
| _cons | .22315 | .2792424 | 0.80 | 0.424 | -.324155 | .7704549 |
| **Uninsure** | | | | | | |
| age | -.0050814 | .0075327 | -0.67 | 0.500 | -.0198452 | .0096823 |
| male | .3332637 | .2432986 | 1.37 | 0.171 | -.1435929 | .8101203 |
| nonwhite | .2485859 | .2767734 | 0.90 | 0.369 | -.29388 | .7910518 |
| **site** | | | | | | |
| 2 | -.6899485 | .2804497 | -2.46 | 0.014 | -1.23962 | -.1402771 |
| 3 | -.1788447 | .2479898 | -0.72 | 0.471 | -.6648957 | .3072063 |
| _cons | -.9855917 | .3891873 | -2.53 | 0.011 | -1.748385 | -.2227986 |

◁

The likelihood function for mprobit is derived under the assumption that all decision-making units face the same choice set, which is the union of all outcomes observed in the dataset. If that is not true for your model, then an alternative is to use the asmprobit command, which does not require this assumption. To do that, you will need to expand the dataset so that each decision maker has $k_i$ observations, where $k_i$ is the number of alternatives in the choice set faced by decision maker $i$. You will also need to create a binary variable to indicate the choice made by each decision maker. Moreover, you will need to use the correlation(independent) and stddev(homoskedastic) options with asmprobit unless you have alternative-specific variables.

## Stored results

mprobit stores the following in e():

Scalars
| | |
|---|---|
| e(N) | number of observations |
| e(k_out) | number of outcomes |
| e(k_points) | number of quadrature points |
| e(k) | number of parameters |
| e(k_eq) | number of equations in e(b) |
| e(k_eq_model) | number of equations in overall model test |
| e(k_indvars) | number of independent variables |
| e(k_dv) | number of dependent variables |
| e(df_m) | model degrees of freedom |
| e(ll) | log simulated-likelihood |
| e(N_clust) | number of clusters |
| e(chi2) | $\chi^2$ |
| e(p) | significance |
| e(i_base) | base outcome index |
| e(const) | 0 if noconstant is specified, 1 otherwise |
| e(probitparam) | 1 if probitparam is specified, 0 otherwise |
| e(rank) | rank of e(V) |
| e(ic) | number of iterations |
| e(rc) | return code |
| e(converged) | 1 if converged, 0 otherwise |

Macros
| | |
|---|---|
| e(cmd) | mprobit |
| e(cmdline) | command as typed |
| e(depvar) | name of dependent variable |
| e(indvars) | independent variables |
| e(wtype) | weight type |
| e(wexp) | weight expression |
| e(title) | title in estimation output |
| e(clustvar) | name of cluster variable |
| e(chi2type) | Wald, type of model $\chi^2$ test |
| e(vce) | *vcetype* specified in vce() |
| e(vcetype) | title used to label Std. Err. |
| e(outeqs) | outcome equations |
| e(out#) | outcome labels, #=1,...,e(k_out) |
| e(opt) | type of optimization |
| e(which) | max or min; whether optimizer is to perform maximization or minimization |
| e(ml_method) | type of ml method |
| e(user) | name of likelihood-evaluator program |
| e(technique) | maximization technique |
| e(properties) | b V |
| e(predict) | program used to implement predict |
| e(marginsnotok) | predictions disallowed by margins |
| e(asbalanced) | factor variables fvset as asbalanced |
| e(asobserved) | factor variables fvset as asobserved |

Matrices
    e(b)                          coefficient vector
    e(outcomes)             outcome values
    e(Cns)                     constraints matrix
    e(ilog)                  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

See Cameron and Trivedi (2005, chap. 15) for a discussion of multinomial models, including multinomial probit. Long and Freese (2014, chap. 8) discuss the multinomial logistic, multinomial probit, and stereotype logistic regression models, with examples using Stata.

As discussed in *Remarks and examples*, the latent variables for a $J$-alternative model are $\eta_{ij} = \mathbf{z}_i\boldsymbol{\alpha}_j + \xi_{ij}$, for $j = 1, \ldots, J$, $i = 1, \ldots, n$, and $\{\xi_{i,1}, \ldots, \xi_{i,J}\} \sim$ i.i.d. $N(0,1)$. The experimenter observes alternative $k$ for the $i$th observation if $\eta_{ik} > \eta_{il}$ for $l \neq k$. For $j \neq k$, let

$$
\begin{aligned}
v_{ij'} &= \eta_{ij} - \eta_{ik} \\
&= \mathbf{z}_i(\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_k) + \xi_{ij} - \xi_{ik} \\
&= \mathbf{z}_i\boldsymbol{\gamma}_{j'} + \epsilon_{ij'}
\end{aligned}
$$

where $j' = j$ if $j < k$ and $j' = j - 1$ if $j > k$ so that $j' = 1, \ldots, J-1$. $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{i,J-1}) \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$, where

$$
\boldsymbol{\Sigma} = \begin{pmatrix}
2 & 1 & 1 & \ldots & 1 \\
1 & 2 & 1 & \ldots & 1 \\
1 & 1 & 2 & \ldots & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 1 & 1 & \ldots & 2
\end{pmatrix}
$$

Denote the deterministic part of the model as $\lambda_{ij'} = \mathbf{z}_i\boldsymbol{\gamma}_{j'}$; the probability that subject $i$ chooses outcome $k$ is

$$
\begin{aligned}
\Pr(y_i = k) &= \Pr(v_{i1} \leq 0, \ldots, v_{i,J-1} \leq 0) \\
&= \Pr(\epsilon_{i1} \leq -\lambda_{i1}, \ldots, \epsilon_{i,J-1} \leq -\lambda_{i,J-1}) \\
&= \frac{1}{(2\pi)^{(J-1)/2} |\boldsymbol{\Sigma}|^{1/2}} \int_{-\infty}^{-\lambda_{i1}} \cdots \int_{-\infty}^{-\lambda_{i,J-1}} \exp\left(-\tfrac{1}{2}\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}\right) d\mathbf{z}
\end{aligned}
$$

Because of the exchangeable correlation structure of $\boldsymbol{\Sigma}$ ($\rho_{ij} = 1/2$ for all $i \neq j$), we can use Dunnett's (1989) result to reduce the multidimensional integral to one dimension:

$$
\Pr(y_i = k) = \frac{1}{\sqrt{\pi}} \int_0^\infty \left\{ \prod_{j=1}^{J-1} \Phi\left(-z\sqrt{2} - \lambda_{ij}\right) + \prod_{j=1}^{J-1} \Phi\left(z\sqrt{2} - \lambda_{ij}\right) \right\} e^{-z^2} dz
$$

Gaussian quadrature is used to approximate this integral, resulting in the $K$-point quadrature formula

$$\Pr(y_i = k) \approx \frac{1}{2} \sum_{k=1}^{K} w_k \left\{ \prod_{j=1}^{J-1} \Phi\left(-\sqrt{2x_k} - \lambda_{ij}\right) + \prod_{j=1}^{J-1} \Phi\left(\sqrt{2x_k} - \lambda_{ij}\right) \right\}$$

where $w_k$ and $x_k$ are the weights and roots of the Laguerre polynomial of order $K$. In mprobit, $K$ is specified by the intpoints() option.

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] _robust, particularly *Maximum likelihood estimators* and *Methods and formulas*.

mprobit also supports estimation with survey data. For details on VCEs with survey data, see [SVY] **variance estimation**.

## References

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Dunnett, C. W. 1989. Algorithm AS 251: Multivariate normal probability integrals with product correlation structure. *Journal of the Royal Statistical Society, Series C* 38: 564–579.

Haan, P., and A. Uhlendorff. 2006. Estimation of multinomial logit models with unobserved heterogeneity using maximum simulated likelihood. *Stata Journal* 6: 229–245.

Hole, A. R. 2007. Fitting mixed logit models by using maximum simulated likelihood. *Stata Journal* 7: 388–401.

Long, J. S., and J. Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*. 3rd ed. College Station, TX: Stata Press.

Tarlov, A. R., J. E. Ware, Jr., S. Greenfield, E. C. Nelson, E. Perrin, and M. Zubkoff. 1989. The medical outcomes study. An application of methods for monitoring the results of medical care. *Journal of the American Medical Association* 262: 925–930.

Wells, K. B., R. D. Hays, M. A. Burnam, W. H. Rogers, S. Greenfield, and J. E. Ware, Jr. 1989. Detection of depressive disorder for patients receiving prepaid or fee-for-service care. Results from the Medical Outcomes Survey. *Journal of the American Medical Association* 262: 3298–3302.

## Also see

[R] **mprobit postestimation** — Postestimation tools for mprobit

[R] **asmprobit** — Alternative-specific multinomial probit regression

[R] **clogit** — Conditional (fixed-effects) logistic regression

[R] **mlogit** — Multinomial (polytomous) logistic regression

[R] **nlogit** — Nested logit regression

[R] **ologit** — Ordered logistic regression

[R] **oprobit** — Ordered probit regression

[MI] **estimation** — Estimation commands for use with mi estimate

[SVY] **svy estimation** — Estimation commands for survey data

[U] **20 Estimation and postestimation commands**