### oprobit — Ordered probit regression

| Description | Quick start          |
|-------------|----------------------|
| Options     | Remarks and examples |
| References  | Also see             |

Menu Stored results N

Syntax Methods and formulas

# Description

oprobit fits ordered probit models of ordinal variable *depvar* on the independent variables *indepvars*. The actual values taken on by the dependent variable are irrelevant, except that larger values are assumed to correspond to "higher" outcomes.

# **Quick start**

Ordinal probit model of y on x1 and categorical variables a and b oprobit y x1 i.a i.b

Model of y on x1 and a one-period lagged value of x1 using tsset data oprobit y x1 L.x1

Same as above, but calculate results for each level of catvar and save statistics to myfile.dta statsby, by(catvar) saving(myfile): oprobit y x1 L.x1

## Menu

 $Statistics > Ordinal \ outcomes > Ordered \ probit \ regression$ 

# Syntax

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

| options   | Description  |
|---|--|
| Model   |  |
| <pre>offset(varname) constraints(constraints)</pre> | include <i>varname</i> in model with coefficient constrained to 1 apply specified linear constraints   |
| SE/Robust   |  |
| vce( <i>vcetype</i> )                               | vcetype may be oim, opg, <u>r</u> obust, <u>cl</u> uster clustvar, <u>boot</u> strap,  |
|   | or <u>jackknife</u>  |
| Reporting   |  |
| <u>l</u> evel(#)                                    | set confidence level; default is level(95)   |
| <u>nocnsr</u> eport                                 | do not display constraints   |
| display_options                                     | control columns and column formats, row spacing, line width,<br>display of omitted variables and base and empty cells, and<br>factor-variable labeling |
| Maximization  |  |
| maximize_options                                    | control the maximization process; seldom used  |
| <u>col</u> linear                                   | keep collinear variables   |
| <u>coefl</u> egend                                  | display legend instead of statistics   |

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

depvar and indepvars may contain time-series operators; see [U] 11.4.4 Time-series varlists.

bayes, bayesboot, bootstrap, by, collect, fmm, fp, jackknife, mfp, mi estimate, nestreg, rolling, statsby, stepwise, and svy are allowed; see [U] **11.1.10 Prefix commands**. For more details, see [BAYES] **bayes: oprobit** and [FMM] **fmm: oprobit**.

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.

collinear and coeflegend do not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

# Options

[ Model ]

offset(varname), constraints(constraints); 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. Reporting

level(#); see [R] Estimation options.

nocnsreport; see [R] Estimation options.

display\_options: noci, nopvalues, noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, 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 options are available with oprobit but is not shown in the dialog box:

collinear, coeflegend; see [R] Estimation options.

### **Remarks and examples**

An ordered probit model is used to estimate relationships between an ordinal dependent variable and a set of independent variables. An *ordinal* variable is a variable that is categorical and ordered, for instance, "poor", "good", and "excellent", which might indicate a person's current health status or the repair record of a car. If there are only two outcomes, see [R] **logistic**, [R] **logit**, and [R] **probit**. This entry is concerned only with more than two outcomes. If the outcomes cannot be ordered (for example, residency in the north, east, south, or west), see [R] **logistic** for a list of related estimation commands.

In ordered probit, an underlying score is estimated as a linear function of the independent variables and a set of cutpoints. The probability of observing outcome *i* corresponds to the probability that the estimated linear function, plus random error, is within the range of the cutpoints estimated for the outcome:

$$\Pr(\mathsf{outcome}_i = i) = \Pr(\kappa_{i-1} < \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_i \le \kappa_i)$$

 $u_j$  is assumed to be normally distributed. In either case, we estimate the coefficients  $\beta_1, \beta_2, \ldots, \beta_k$  together with the cutpoints  $\kappa_1, \kappa_2, \ldots, \kappa_{I-1}$ , where *I* is the number of possible outcomes.  $\kappa_0$  is taken as  $-\infty$ , and  $\kappa_I$  is taken as  $+\infty$ . All of this is a direct generalization of the ordinary two-outcome probit model.

#### Example 1

In example 2 of [R] **ologit**, we use a variation of the automobile dataset (see [U] **1.2.2 Example datasets**) to analyze the 1977 repair records of 66 foreign and domestic cars. We use ordered logit to explore the relationship of rep77 in terms of foreign (origin of manufacture), length (a proxy for size), and mpg. Here we fit the same model using ordered probit rather than ordered logit:

| . use https://<br>(Automobile mo   | /www.stata-pre<br>odels)  | ss.com/data/   | r19/ful]                   | Lauto                   |  |  |
|--|---|--|----------------------------|-------------------------|--|--|
| . oprobit rep7   | 77 foreign len  | gth mpg  |                            |                         |  |  |
| Iteration 0:<br>Iteration 1:<br>Iteration 2:<br>Iteration 3:<br>Iteration 4: | Log likelihoo<br>Log likelihoo<br>Log likelihoo<br>Log likelihoo<br>Log likelihoo | d = -89.8950<br>d = -78.1063<br>d = -78.0200<br>d = -78.0200<br>d = -78.0200<br>d = -78.0200 | 98<br>16<br>86<br>25<br>25 |                         |  |  |
| Ordered probit   | regression<br>1 = -78.020025  |  |                            |                         | Number of ob<br>LR chi2(3)<br>Prob > chi2<br>Pseudo R2 | s = 66<br>= 23.75<br>= 0.0000<br>= 0.1321    |
| rep77  | Coefficient   | Std. err.  | z                          | P> z                    | [95% conf.   | interval]                                    |
| foreign<br>length<br>mpg   | 1.704861<br>.0468675<br>.1304559  | .4246796<br>.012648<br>.0378628  | 4.01<br>3.71<br>3.45       | 0.000<br>0.000<br>0.001 | .8725037<br>.022078<br>.0562463                        | 2.537217<br>.0716571<br>.2046656             |
| /cut1<br>/cut2<br>/cut3<br>/cut4   | 10.1589<br>11.21003<br>12.54561<br>13.98059                                       | 3.076754<br>3.107527<br>3.155233<br>3.218793   |                            |                         | 4.128577<br>5.119389<br>6.361467<br>7.671874           | 16.18923<br>17.30067<br>18.72975<br>20.28931 |

We find that foreign cars have better repair records, as do larger cars and cars with better mileage ratings.

# **Stored results**

oprobit stores the following in e():

| Scal | ars           |  |
|------|---------------|--|
|      | e(N)          | number of observations                       |
|      | e(N_cd)       | number of completely determined observations |
|      | e(k_cat)      | number of categories                         |
|      | e(k)          | number of parameters                         |
|      | e(k_aux)      | number of auxiliary parameters               |
|      | e(k_eq)       | number of equations in e(b)                  |
|      | e(k_eq_model) | number of equations in overall model test    |
|      | e(k_dv)       | number of dependent variables                |
|      | e(df_m)       | model degrees of freedom                     |
|      | e(r2_p)       | pseudo- $R^2$                                |
|      | e(11)         | log likelihood                               |
|      | e(11_0)       | log likelihood, constant-only model          |
|      | e(N_clust)    | number of clusters                           |
|      | e(chi2)       | $\chi^2$                                     |
|      | e(p)          | <i>p</i> -value for model test               |
|      | e(rank)       | rank of e(V)                                 |
|      | e(ic)         | number of iterations                         |
|      | e(rc)         | return code                                  |
|      | e(converged)  | 1 if converged, 0 otherwise                  |
| Mac  | eros          |  |
|      | e(cmd)        | oprobit                                      |
|      | e(cmdline)    | command as typed                             |
|      | e(depvar)     | name of dependent variable                   |
|      | e(wtype)      | weight type                                  |
|      |               |  |

|     | e(wexp)           | weight expression  |
|-----|-------------------|--|
|     | e(title)          | title in estimation output   |
|     | e(clustvar)       | name of cluster variable   |
|     | e(offset)         | linear offset variable   |
|     | e(chi2type)       | Wald or LR; type of model $\chi^2$ test                                  |
|     | e(vce)            | vcetype specified in vce()   |
|     | e(vcetype)        | title used to label Std. err.  |
|     | 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(marginsdefault) | default predict() specification for margins                              |
|     | e(asbalanced)     | factor variables fvset as asbalanced                                     |
|     | e(asobserved)     | factor variables fvset as asobserved                                     |
| Ma  | trices            |  |
|     | e(b)              | coefficient vector   |
|     | e(Cns)            | constraints matrix   |
|     | e(ilog)           | iteration log (up to 20 iterations)                                      |
|     | e(gradient)       | gradient vector  |
|     | e(cat)            | category values  |
|     | e(V)              | variance-covariance matrix of the estimators                             |
|     | e(V_modelbased)   | model-based variance   |
| Fun | ctions            |  |
|     | e(sample)         | marks estimation sample  |
|     |                   |  |

In addition to the above, the following is stored in r():

Matrices

r(table) matrix containing the coefficients with their standard errors, test statistics, p-values, and confidence intervals

Note that results stored in r() are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

# Methods and formulas

See Methods and formulas of [R] ologit.

## References

- Aitchison, J., and S. D. Silvey. 1957. The generalization of probit analysis to the case of multiple responses. *Biometrika* 44: 131–140. https://doi.org/10.2307/2333245.
- Bauldry, S., J. Xu, and A. S. Fullerton. 2018. gencrm: A new command for generalized continuation-ratio models. *Stata Journal* 18: 924–936.
- Cameron, A. C., and P. K. Trivedi. 2005. Microeconometrics: Methods and Applications. New York: Cambridge University Press.
- Canette, I. 2013. Fitting ordered probit models with endogenous covariates with Stata's gsem command. The Stata Blog: Not Elsewhere Classified. https://blog.stata.com/2013/11/07/fitting-ordered-probit-models-with-endogenouscovariates-with-statas-gsem-command/.
- Chiburis, R., and M. Lokshin. 2007. Maximum likelihood and two-step estimation of an ordered-probit selection model. *Stata Journal* 7: 167–182.

- De Luca, G., and V. Perotti. 2011. Estimation of ordered response models with sample selection. *Stata Journal* 11: 213–239.
- Drukker, D. M. 2016. An ordered-probit inverse probability weighted (IPW) estimator. *The Stata Blog: Not Elsewhere Classified*. https://blog.stata.com/2016/09/13/an-ordered-probit-inverse-probability-weighted-ipw-estimator/.
- Huismans, J., J. W. Nijenhuis, and A. Sirchenko. 2022. A mixture of ordered probit models with endogenous switching between two latent classes. Stata Journal 22: 557–596.
- Long, J. S. 1997. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage.
- Long, J. S., and J. Freese. 2014. Regression Models for Categorical Dependent Variables Using Stata. 3rd ed. College Station, TX: Stata Press.
- Miranda, A., and S. Rabe-Hesketh. 2006. Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. *Stata Journal* 6: 285–308.
- Smith, E. K., M. G. Lacy, and A. Mayer. 2019. Performance simulations for categorical mediation: Analyzing khb estimates of mediation in ordinal regression models. *Stata Journal* 19: 913–930.
- Stewart, M. B. 2004. Semi-nonparametric estimation of extended ordered probit models. Stata Journal 4: 27-39.
- Williams, R. 2010. Fitting heterogeneous choice models with oglm. Stata Journal 10: 540-567.
- Xu, J., and J. S. Long. 2005. Confidence intervals for predicted outcomes in regression models for categorical outcomes. Stata Journal 5: 537–559.

## Also see

- [R] oprobit postestimation Postestimation tools for oprobit
- [R] heckoprobit Ordered probit model with sample selection
- [R] **hetoprobit** Heteroskedastic ordered probit regression
- [R] logistic Logistic regression, reporting odds ratios
- [R] **mlogit** Multinomial (polytomous) logistic regression
- [R] mprobit Multinomial probit regression
- [R] ologit Ordered logistic regression
- [R] **probit** Probit regression
- [R] **zioprobit** Zero-inflated ordered probit regression
- [BAYES] bayes: oprobit Bayesian ordered probit regression
- [CM] **cmroprobit** Rank-ordered probit choice model
- [ERM] **eoprobit** Extended ordered probit regression
- [FMM] **fmm: oprobit** Finite mixtures of ordered probit regression models
- [ME] meoprobit Multilevel mixed-effects ordered probit regression
- [MI] Estimation Estimation commands for use with mi estimate
- [SVY] svy estimation Estimation commands for survey data
- [XT] **xtoprobit** Random-effects ordered probit model

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

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.



For suggested citations, see the FAQ on citing Stata documentation.