# Title

stata.com

**logit postestimation —** Postestimation tools for logit


Description | Syntax for predict | Menu for predict | Options for predict
Remarks and examples | Methods and formulas | References | Also see


## Description

The following postestimation commands are of special interest after `logit`:

| Command | Description |
|---|---|
| `estat classification` | report various summary statistics, including the classification table |
| `estat gof` | Pearson or Hosmer–Lemeshow goodness-of-fit test |
| `lroc` | compute area under ROC curve and graph the curve |
| `lsens` | graph sensitivity and specificity versus probability cutoff |

These commands are not appropriate after the `svy` prefix.

The following standard postestimation commands are also available:

| Command | Description |
|---|---|
| `contrast` | contrasts and ANOVA-style joint tests of estimates |
| `estat ic` | Akaike's and Schwarz's Bayesian information criteria (AIC and BIC) |
| `estat summarize` | summary statistics for the estimation sample |
| `estat vce` | variance–covariance matrix of the estimators (VCE) |
| `estat (svy)` | postestimation statistics for survey data |
| `estimates` | cataloging estimation results |
| `forecast`[1] | dynamic forecasts and simulations |
| `lincom` | point estimates, standard errors, testing, and inference for linear combinations of coefficients |
| `linktest` | link test for model specification |
| `lrtest`[2] | likelihood-ratio test |
| `margins` | marginal means, predictive margins, marginal effects, and average marginal effects |
| `marginsplot` | graph the results from margins (profile plots, interaction plots, etc.) |
| `nlcom` | point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients |
| `predict` | predictions, residuals, influence statistics, and other diagnostic measures |
| `predictnl` | point estimates, standard errors, testing, and inference for generalized predictions |
| `pwcompare` | pairwise comparisons of estimates |
| `suest` | seemingly unrelated estimation |
| `test` | Wald tests of simple and composite linear hypotheses |
| `testnl` | Wald tests of nonlinear hypotheses |

[1] `forecast` is not appropriate with `mi` or `svy` estimation results.

[2] `lrtest` is not appropriate with `svy` estimation results.

# Syntax for predict

```
predict [type] newvar [if] [in] [, statistic nooffset rules asif]
```

| *statistic* | Description |
|---|---|
| Main | |
| pr | probability of a positive outcome; the default |
| xb | linear prediction |
| stdp | standard error of the prediction |
| *dbeta | Pregibon (1981) $\Delta\widehat{\beta}$ influence statistic |
| *deviance | deviance residual |
| *dx2 | Hosmer, Lemeshow, and Sturdivant (2013) $\Delta\chi^2$ influence statistic |
| *ddeviance | Hosmer, Lemeshow, and Sturdivant (2013) $\Delta D$ influence statistic |
| *hat | Pregibon (1981) leverage |
| *number | sequential number of the covariate pattern |
| *residuals | Pearson residuals; adjusted for number sharing covariate pattern |
| *rstandard | standardized Pearson residuals; adjusted for number sharing covariate pattern |
| score | first derivative of the log likelihood with respect to $\mathbf{x}_j\boldsymbol{\beta}$ |

Unstarred statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample. Starred statistics are calculated only for the estimation sample, even when `if e(sample)` is not specified.

`pr`, `xb`, `stdp`, and `score` are the only options allowed with `svy` estimation results.

# Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

# Options for predict

## Main

`pr`, the default, calculates the probability of a positive outcome.

`xb` calculates the linear prediction.

`stdp` calculates the standard error of the linear prediction.

`dbeta` calculates the Pregibon (1981) $\Delta\widehat{\beta}$ influence statistic, a standardized measure of the difference in the coefficient vector that is due to deletion of the observation along with all others that share the same covariate pattern. In Hosmer, Lemeshow, and Sturdivant (2013, 154–155) jargon, this statistic is $M$-asymptotic; that is, it is adjusted for the number of observations that share the same covariate pattern.

`deviance` calculates the deviance residual.

`dx2` calculates the Hosmer, Lemeshow, and Sturdivant (2013, 191) $\Delta\chi^2$ influence statistic, reflecting the decrease in the Pearson $\chi^2$ that is due to deletion of the observation and all others that share the same covariate pattern.

`ddeviance` calculates the Hosmer, Lemeshow, and Sturdivant (2013, 191) $\Delta D$ influence statistic, which is the change in the deviance residual that is due to deletion of the observation and all others that share the same covariate pattern.

`hat` calculates the Pregibon (1981) leverage or the diagonal elements of the hat matrix adjusted for the number of observations that share the same covariate pattern.

`number` numbers the covariate patterns—observations with the same covariate pattern have the same `number`. Observations not used in estimation have `number` set to missing. The first covariate pattern is numbered 1, the second 2, and so on.

`residuals` calculates the Pearson residual as given by Hosmer, Lemeshow, and Sturdivant (2013, 155) and adjusted for the number of observations that share the same covariate pattern.

`rstandard` calculates the standardized Pearson residual as given by Hosmer, Lemeshow, and Sturdivant (2013, 191) and adjusted for the number of observations that share the same covariate pattern.

`score` calculates the equation-level score, $\partial \ln L/\partial(\mathbf{x}_j\boldsymbol{\beta})$.

Options

`nooffset` is relevant only if you specified `offset(`*varname*`)` for `logit`. It modifies the calculations made by `predict` so that they ignore the offset variable; the linear prediction is treated as $\mathbf{x}_j\mathbf{b}$ rather than as $\mathbf{x}_j\mathbf{b} + \text{offset}_j$.

`rules` requests that Stata use any rules that were used to identify the model when making the prediction. By default, Stata calculates missing for excluded observations.

`asif` requests that Stata ignore the rules and exclusion criteria and calculate predictions for all observations possible by using the estimated parameter from the model.

## Remarks and examples

stata.com

Once you have fit a logit model, you can obtain the predicted probabilities by using the `predict` command for both the estimation sample and other samples; see **[U] 20 Estimation and postestimation commands** and [R] **predict**. Here we will make only a few more comments.

`predict` without arguments calculates the predicted probability of a positive outcome, that is, $\Pr(y_j = 1) = F(\mathbf{x}_j\mathbf{b})$. With the `xb` option, `predict` calculates the linear combination $\mathbf{x}_j\mathbf{b}$, where $\mathbf{x}_j$ are the independent variables in the $j$th observation and $\mathbf{b}$ is the estimated parameter vector. This is sometimes known as the index function because the cumulative distribution function indexed at this value is the probability of a positive outcome.

In both cases, Stata remembers any rules used to identify the model and calculates missing for excluded observations, unless `rules` or `asif` is specified. For information about the other statistics available after `predict`, see [R] **logistic postestimation**.

### ▷ Example 1: Predicted probabilities

In example 2 of [R] **logit**, we fit the logit model `logit foreign b3.repair`. To obtain predicted probabilities, type

```
. use http://www.stata-press.com/data/r13/repair
(1978 Automobile Data)
. logit foreign b3.repair
note: 1.repair != 0 predicts failure perfectly
      1.repair dropped and 10 obs not used
  (output omitted)
. predict p
(option pr assumed; Pr(foreign))
(10 missing values generated)
```

```
. summarize foreign p
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     foreign |         58    .2068966    .4086186          0          1
           p |         48         .25    .1956984         .1         .5
```

Stata remembers any rules used to identify the model and sets predictions to missing for any excluded observations. `logit` dropped the variable `1.repair` from our model and excluded 10 observations. Thus when we typed `predict p`, those same 10 observations were again excluded, and their predictions were set to missing.

`predict`'s `rules` option uses the rules in the prediction. During estimation, we were told "1.repair != 0 predicts failure perfectly", so the rule is that when `1.repair` is not zero, we should predict 0 probability of success or a positive outcome:

```
. predict p2, rules
(option pr assumed; Pr(foreign))
. summarize foreign p p2
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     foreign |         58    .2068966    .4086186          0          1
           p |         48         .25    .1956984         .1         .5
          p2 |         58    .2068966    .2016268          0         .5
```

`predict`'s `asif` option ignores the rules and exclusion criteria and calculates predictions for all observations possible by using the estimated parameters from the model:

```
. predict p3, asif
(option pr assumed; Pr(foreign))
. summarize foreign p p2 p3
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     foreign |         58    .2068966    .4086186          0          1
           p |         48         .25    .1956984         .1         .5
          p2 |         58    .2068966    .2016268          0         .5
          p3 |         58    .2931035    .2016268         .1         .5
```

Which is right? What `predict` does by default is the most conservative approach. If many observations had been excluded because of a simple rule, we could be reasonably certain that the `rules` prediction is correct. The `asif` prediction is correct only if the exclusion is a fluke, and we would be willing to exclude the variable from the analysis anyway. Then, however, we would refit the model to include the excluded observations.

◁

## ▷ Example 2: Predictive margins

We can use the command `margins, contrast` after `logit` to make comparisons on the probability scale. Let's fit a model predicting low birthweight from characteristics of the mother:

```
. use http://www.stata-press.com/data/r13/lbw, clear
(Hosmer & Lemeshow data)
```

```
. logit low age i.race i.smoke ptl i.ht i.ui
Iteration 0:   log likelihood =   -117.336
Iteration 1:   log likelihood = -103.81846
Iteration 2:   log likelihood = -103.40486
Iteration 3:   log likelihood = -103.40384
Iteration 4:   log likelihood = -103.40384
Logistic regression                             Number of obs   =        189
                                                LR chi2(7)      =      27.86
                                                Prob > chi2     =     0.0002
Log likelihood = -103.40384                     Pseudo R2       =     0.1187
```

| low | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| age | -.0403293 | .0357127 | -1.13 | 0.259 | -.1103249 | .0296663 |
| race | | | | | | |
| black | 1.009436 | .5025122 | 2.01 | 0.045 | .0245302 | 1.994342 |
| other | 1.001908 | .4248342 | 2.36 | 0.018 | .1692485 | 1.834568 |
| smoke | | | | | | |
| smoker | .9631876 | .3904357 | 2.47 | 0.014 | .1979477 | 1.728427 |
| ptl | .6288678 | .3399067 | 1.85 | 0.064 | -.0373371 | 1.295073 |
| 1.ht | 1.358142 | .6289555 | 2.16 | 0.031 | .125412 | 2.590872 |
| 1.ui | .8001832 | .4572306 | 1.75 | 0.080 | -.0959724 | 1.696339 |
| _cons | -1.184127 | .9187461 | -1.29 | 0.197 | -2.984837 | .6165818 |

The coefficients are log odds-ratios: conditional on the other predictors, smoking during pregnancy is associated with an increase of 0.96 in the log odds-ratios of low birthweight. The model is linear in the log odds-scale, so the estimate of 0.96 has the same interpretation, whatever the values of the other predictors might be. We could convert 0.96 to an odds ratio by replaying the results with `logit, or`.

But what if we want to talk about the probability of low birthweight, and not the odds? Then we will need the command `margins, contrast`. We will use the `r.` contrast operator to compare each level of `smoke` with a reference level. (`smoke` has only two levels, so there will be only one comparison: a comparison of smokers with nonsmokers.)

```
. margins r.smoke, contrast
Contrasts of predictive margins
Model VCE    : OIM
Expression   : Pr(low), predict()
```

| | df | chi2 | P>chi2 |
|---|---|---|---|
| smoke | 1 | 6.32 | 0.0119 |

| | Contrast | Delta-method Std. Err. | [95% Conf. | Interval] |
|---|---|---|---|---|
| smoke (smoker vs nonsmoker) | .1832779 | .0728814 | .0404329 | .3261229 |

We see that maternal smoking is associated with an 18.3% increase in the probability of low birthweight. (We received a contrast in the probability scale because predicted probabilities are the default when `margins` is used after `logit`.)

The contrast of 18.3% is a difference of margins that are computed by averaging over the predictions for observations in the estimation sample. If the values of the other predictors were different, the contrast for `smoke` would be different, too. Let's estimate the contrast for 25-year-old mothers:

```
. margins r.smoke, contrast at(age=25)
Contrasts of predictive margins
Model VCE    : OIM
Expression   : Pr(low), predict()
at           : age             =          25
```

| | df | chi2 | P>chi2 |
|---|---|---|---|
| smoke | 1 | 6.19 | 0.0129 |

| | Contrast | Delta-method Std. Err. | [95% Conf. | Interval] |
|---|---|---|---|---|
| smoke (smoker vs nonsmoker) | .1808089 | .0726777 | .0383632 | .3232547 |

Specifying a maternal `age` of 25 changed the contrast to 18.1%. Our contrast of probabilities changed because the `logit` model is nonlinear in the probability scale. A contrast of log odds-ratios would not have changed.

◁

## Methods and formulas

See *Methods and formulas* of the individual postestimation commands for details.

## References


Hosmer, D. W., Jr., S. A. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: Wiley.

Newson, R. B. 2013. Attributable and unattributable risks and fractions and other scenario comparisons. *Stata Journal* 13: 672–698.

Powers, D. A., H. Yoshioka, and M.-S. Yun. 2011. mvdcmp: Multivariate decomposition for nonlinear response models. *Stata Journal* 11: 556–576.

Pregibon, D. 1981. Logistic regression diagnostics. *Annals of Statistics* 9: 705–724.


## Also see

[R] **logit** — Logistic regression, reporting coefficients

[R] **estat classification** — Classification statistics and table

[R] **estat gof** — Pearson or Hosmer–Lemeshow goodness-of-fit test

[R] **lroc** — Compute area under ROC curve and graph the curve

[R] **lsens** — Graph sensitivity and specificity versus probability cutoff

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