logistic — Logistic regression, reporting odds ratios

Description

`logistic` fits a logistic regression model of *depvar* on *indepvars*, where *depvar* is a 0/1 variable (or, more precisely, a 0/non-0 variable). Without arguments, `logistic` redisplays the last `logistic` estimates. `logistic` displays estimates as odds ratios; to view coefficients, type `logit` after running `logistic`. To obtain odds ratios for any covariate pattern relative to another, see [R] **lincom**.

Quick start

Report odds ratios from logistic regression of `y` on `x1` and `x2`

```
logistic y x1 x2
```

Add indicators for values of categorical variable `a`

```
logistic y x1 x2 i.a
```

As above, and apply frequency weights defined by `wvar`

```
logistic y x1 x2 i.a [fweight=wvar]
```

Show base level of `a`

```
logistic y x1 x2 i.a, baselevels
```

Menu

Statistics > Binary outcomes > Logistic regression

Syntax

`logistic` *depvar indepvars* [*if*] [*in*] [*weight*] [, *options*]

options	Description
Model	
`noconstant`	suppress constant term
`offset(`*varname*`)`	include *varname* in model with coefficient constrained to 1
`asis`	retain perfect predictor variables
`constraints(`*constraints*`)`	apply specified linear constraints
`collinear`	keep collinear variables
SE/Robust	
`vce(`*vcetype*`)`	*vcetype* may be `oim`, `robust`, `cluster` *clustvar*, `bootstrap`, or `jackknife`
Reporting	
`level(`*#*`)`	set confidence level; default is `level(95)`
`coef`	report estimated coefficients
`nocnsreport`	do not display constraints
display_options	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
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**.

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

`bayes`, `bootstrap`, `by`, `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: logistic**.

`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**.

`fweight`s, `iweight`s, and `pweight`s 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

Model

`noconstant`, `offset(`*varname*`)`, `constraints(`*constraints*`)`, `collinear`; see [R] **estimation options**.

`asis` forces retention of perfect predictor variables and their associated perfectly predicted observations and may produce instabilities in maximization; see [R] **probit**.

SE/Robust

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***.

Reporting

level(*#*); see [R] **estimation options**.

coef causes logistic to report the estimated coefficients rather than the odds ratios (exponentiated coefficients). coef may be specified when the model is fit or may be used later to redisplay results. coef affects only how results are displayed and not how they are estimated.

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 option is available with logistic but is not shown in the dialog box:

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

Remarks and examples

stata.com

Remarks are presented under the following headings:

logistic and logit
Robust estimate of variance
Video examples

logistic and logit

logistic provides an alternative and preferred way to fit maximum-likelihood logit models, the other choice being logit ([R] **logit**).

First, let's dispose of some confusing terminology. We use the words logit and logistic to mean the same thing: maximum likelihood estimation. To some, one or the other of these words connotes transforming the dependent variable and using weighted least squares to fit the model, but that is not how we use either word here. Thus the logit and logistic commands produce the same results.

The logistic command is generally preferred to the logit command because logistic presents the estimates in terms of odds ratios rather than coefficients. To some people, this may seem disadvantageous, but you can type logit without arguments after logistic to see the underlying coefficients. You should be cautious when interpreting the odds ratio of the constant term. Usually, this odds ratio represents the baseline odds of the model when all predictor variables are set to zero. However, you must verify that a zero value for all predictor variables in the model actually makes sense before continuing with this interpretation.

Nevertheless, [R] **logit** is still worth reading because `logistic` shares the same features as `logit`, including omitting variables due to collinearity or one-way causation.

For an introduction to logistic regression, see Lemeshow and Hosmer (2005), Pagano and Gauvreau (2000, 470–487), or Pampel (2000); for a complete but nonmathematical treatment, see Kleinbaum and Klein (2010); and for a thorough discussion, see Hosmer, Lemeshow, and Sturdivant (2013). See Gould (2000) for a discussion of the interpretation of logistic regression. See Dupont (2009) or Hilbe (2009) for a discussion of logistic regression with examples using Stata. For a discussion using Stata with an emphasis on model specification, see Vittinghoff et al. (2012).

Stata has a variety of commands for performing estimation when the dependent variable is dichotomous or polytomous. See Long and Freese (2014) for a book devoted to fitting these models with Stata. Below is a list of some estimation commands that may be of interest. See `help estimation commands` for a complete list of all of Stata's estimation commands.

asclogit	[R] **asclogit**	Alternative-specific conditional logit (McFadden's choice) model
asmixlogit	[R] **asmixlogit**	Alternative-specific mixed logit regression
asmprobit	[R] **asmprobit**	Alternative-specific multinomial probit regression
asroprobit	[R] **asroprobit**	Alternative-specific rank-ordered probit regression
bayes: *cmd*	[BAYES] **bayesian estimation**	Bayesian estimation is available for many of these commands
binreg	[R] **binreg**	Generalized linear models for the binomial family
biprobit	[R] **biprobit**	Bivariate probit regression
clogit	[R] **clogit**	Conditional (fixed-effects) logistic regression
cloglog	[R] **cloglog**	Complementary log-log regression
exlogistic	[R] **exlogistic**	Exact logistic regression
fmm: *cmd*	[FMM] **fmm estimation**	Finite mixture modeling is available for some of these commands
glm	[R] **glm**	Generalized linear models
heckoprobit	[R] **heckoprobit**	Ordered probit model with sample selection
heckprobit	[R] **heckprobit**	Probit model with sample selection
hetprobit	[R] **hetprobit**	Heteroskedastic probit model
ivprobit	[R] **ivprobit**	Probit model with endogenous covariates
logit	[R] **logit**	Logistic regression, reporting coefficients
mecloglog	[ME] **mecloglog**	Multilevel mixed-effects complementary log-log regression
meglm	[ME] **meglm**	Multilevel mixed-effects generalized linear model
melogit	[ME] **melogit**	Multilevel mixed-effects logistic regression
meologit	[ME] **meologit**	Multilevel mixed-effects ordered logistic regression
meoprobit	[ME] **meoprobit**	Multilevel mixed-effects ordered probit regression
meprobit	[ME] **meprobit**	Multilevel mixed-effects probit regression
mlogit	[R] **mlogit**	Multinomial (polytomous) logistic regression
mprobit	[R] **mprobit**	Multinomial probit regression
nlogit	[R] **nlogit**	Nested logit regression (RUM-consistent and nonnormalized)
ologit	[R] **ologit**	Ordered logistic regression
oprobit	[R] **oprobit**	Ordered probit regression
probit	[R] **probit**	Probit regression
rologit	[R] **rologit**	Rank-ordered logistic regression
scobit	[R] **scobit**	Skewed logistic regression
slogit	[R] **slogit**	Stereotype logistic regression
svy: *cmd*	[SVY] **svy estimation**	Survey versions of many of these commands are available
xtcloglog	[XT] **xtcloglog**	Random-effects and population-averaged cloglog models
xtgee	[XT] **xtgee**	GEE population-averaged generalized linear models
xtlogit	[XT] **xtlogit**	Fixed-effects, random-effects, and population-averaged logit models
xtologit	[XT] **xtologit**	Random-effects ordered logistic models
xtoprobit	[XT] **xtoprobit**	Random-effects ordered probit models
xtprobit	[XT] **xtprobit**	Random-effects and population-averaged probit models
zioprobit	[R] **zioprobit**	Zero-inflated ordered probit regression

▷ Example 1

Consider the following dataset from a study of risk factors associated with low birthweight described in Hosmer, Lemeshow, and Sturdivant (2013, 24).

```
. use http://www.stata-press.com/data/r15/lbw
(Hosmer & Lemeshow data)
. describe
Contains data from http://www.stata-press.com/data/r15/lbw.dta
  obs:           189                          Hosmer & Lemeshow data
 vars:            11                          15 Jan 2016 05:01
 size:         2,646
─────────────────────────────────────────────────────────────────────────────
              storage   display    value
variable name   type    format     label      variable label
─────────────────────────────────────────────────────────────────────────────
id              int     %8.0g                 identification code
low             byte    %8.0g                 birthweight<2500g
age             byte    %8.0g                 age of mother
lwt             int     %8.0g                 weight at last menstrual period
race            byte    %8.0g      race       race
smoke           byte    %9.0g      smoke      smoked during pregnancy
ptl             byte    %8.0g                 premature labor history (count)
ht              byte    %8.0g                 has history of hypertension
ui              byte    %8.0g                 presence, uterine irritability
ftv             byte    %8.0g                 number of visits to physician
                                                during 1st trimester
bwt             int     %8.0g                 birthweight (grams)
─────────────────────────────────────────────────────────────────────────────
Sorted by:
```

We want to investigate the causes of low birthweight. Here `race` is a categorical variable indicating whether a person is white (`race` = 1), black (`race` = 2), or some other race (`race` = 3). We want indicator (dummy) variables for `race` included in the regression, so we will use factor variables.

```
. logistic low age lwt i.race smoke ptl ht ui
Logistic regression                             Number of obs     =        189
                                                LR chi2(8)        =      33.22
                                                Prob > chi2       =     0.0001
Log likelihood =   -100.724                     Pseudo R2         =     0.1416

─────────────┬────────────────────────────────────────────────────────────────
         low │ Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
─────────────┼────────────────────────────────────────────────────────────────
         age │   .9732636   .0354759    -0.74   0.457     .9061578    1.045339
         lwt │   .9849634   .0068217    -2.19   0.029     .9716834    .9984249
             │
        race │
      black  │   3.534767   1.860737     2.40   0.016     1.259736    9.918406
      other  │   2.368079   1.039949     1.96   0.050     1.001356    5.600207
             │
       smoke │   2.517698    1.00916     2.30   0.021     1.147676    5.523162
         ptl │   1.719161   .5952579     1.56   0.118     .8721455    3.388787
          ht │   6.249602   4.322408     2.65   0.008     1.611152    24.24199
          ui │     2.1351   .9808153     1.65   0.099     .8677528      5.2534
       _cons │   1.586014   1.910496     0.38   0.702     .1496092     16.8134
─────────────┴────────────────────────────────────────────────────────────────
Note: _cons estimates baseline odds.
```

The odds ratios are for a one-unit change in the variable. If we wanted the odds ratio for `age` to be in terms of 4-year intervals, we would type

```
. generate age4 = age/4
. logistic low age4 lwt i.race smoke ptl ht ui
(output omitted)
```

After `logistic`, we can type `logit` to see the model in terms of coefficients and standard errors:

```
. logit
Logistic regression                             Number of obs     =        189
                                                LR chi2(8)        =      33.22
                                                Prob > chi2       =     0.0001
Log likelihood =   -100.724                     Pseudo R2         =     0.1416

------------------------------------------------------------------------------
         low |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age4 |  -.1084012   .1458017    -0.74   0.457    -.3941673    .1773649
         lwt |  -.0151508   .0069259    -2.19   0.029    -.0287253   -.0015763
             |
        race |
      black  |   1.262647   .5264101     2.40   0.016     .2309024    2.294392
      other  |   .8620792   .4391532     1.96   0.050     .0013548    1.722804
             |
       smoke |   .9233448   .4008266     2.30   0.021      .137739    1.708951
         ptl |   .5418366    .346249     1.56   0.118     -.136799    1.220472
          ht |   1.832518   .6916292     2.65   0.008     .4769494    3.188086
          ui |   .7585135   .4593768     1.65   0.099    -.1418484    1.658875
       _cons |   .4612239    1.20459     0.38   0.702    -1.899729    2.822176
------------------------------------------------------------------------------
```

If we wanted to see the logistic output again, we would type `logistic` without arguments.

◁

▷ Example 2

We can specify the confidence interval for the odds ratios with the `level()` option, and we can do this either at estimation time or when replaying the model. For instance, to see our first model in example 1 with narrower, 90% confidence intervals, we might type

```
. logistic, level(90)
Logistic regression                             Number of obs     =        189
                                                LR chi2(8)        =      33.22
                                                Prob > chi2       =     0.0001
Log likelihood =   -100.724                     Pseudo R2         =     0.1416

------------------------------------------------------------------------------
         low | Odds Ratio   Std. Err.      z    P>|z|     [90% Conf. Interval]
-------------+----------------------------------------------------------------
        age4 |   .8972675   .1308231    -0.74   0.457     .7059409    1.140448
         lwt |   .9849634   .0068217    -2.19   0.029     .9738063    .9962483
             |
        race |
      black  |   3.534767   1.860737     2.40   0.016     1.487028    8.402379
      other  |   2.368079   1.039949     1.96   0.050     1.149971    4.876471
             |
       smoke |   2.517698    1.00916     2.30   0.021     1.302185    4.867819
         ptl |   1.719161   .5952579     1.56   0.118     .9726876    3.038505
          ht |   6.249602   4.322408     2.65   0.008     2.003487    19.49478
          ui |     2.1351   .9808153     1.65   0.099      1.00291    4.545424
       _cons |   1.586014   1.910496     0.38   0.702     .2186791    11.50288
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

◁

Robust estimate of variance

If you specify `vce(robust)`, Stata reports the robust estimate of variance described in [U] 20.22 Obtaining robust variance estimates. Here is the model previously fit with the robust estimate of variance:

```
. logistic low age lwt i.race smoke ptl ht ui, vce(robust)
Logistic regression                             Number of obs     =        189
                                                Wald chi2(8)      =      29.02
                                                Prob > chi2       =     0.0003
Log pseudolikelihood =   -100.724               Pseudo R2         =     0.1416
```

low	Odds Ratio	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.9732636	.0329376	-0.80	0.423	.9108015	1.040009
lwt	.9849634	.0070209	-2.13	0.034	.9712984	.9988206
race						
black	3.534767	1.793616	2.49	0.013	1.307504	9.556051
other	2.368079	1.026563	1.99	0.047	1.012512	5.538501
smoke	2.517698	.9736417	2.39	0.017	1.179852	5.372537
ptl	1.719161	.7072902	1.32	0.188	.7675715	3.850476
ht	6.249602	4.102026	2.79	0.005	1.726445	22.6231
ui	2.1351	1.042775	1.55	0.120	.8197749	5.560858
_cons	1.586014	1.939482	0.38	0.706	.144345	17.42658

```
Note: _cons estimates baseline odds.
```

Also you can specify `vce(cluster` *clustvar*`)` and then, within cluster, relax the assumption of independence. To illustrate this, we have made some fictional additions to the low-birthweight data.

Say that these data are not a random sample of mothers but instead are a random sample of mothers from a random sample of hospitals. In fact, that may be true—we do not know the history of these data.

Hospitals specialize, and it would not be too incorrect to say that some hospitals specialize in more difficult cases. We are going to show two extremes. In one, all hospitals are alike, but we are going to estimate under the possibility that they might differ. In the other, hospitals are strikingly different. In both cases, we assume that patients are drawn from 20 hospitals.

In both examples, we will fit the same model, and we will type the same command to fit it. Below are the same data we have been using but with a new variable, `hospid`, that identifies from which of the 20 hospitals each patient was drawn (and which we have made up):

```
. use http://www.stata-press.com/data/r15/hospid1, clear
. logistic low age lwt i.race smoke ptl ht ui, vce(cluster hospid)
Logistic regression                             Number of obs     =        189
                                                Wald chi2(8)      =      49.67
                                                Prob > chi2       =     0.0000
Log pseudolikelihood =   -100.724               Pseudo R2         =     0.1416
                               (Std. Err. adjusted for 20 clusters in hospid)
```

low	Odds Ratio	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.9732636	.0397476	-0.66	0.507	.898396	1.05437
lwt	.9849634	.0057101	-2.61	0.009	.9738352	.9962187
race						
black	3.534767	2.013285	2.22	0.027	1.157563	10.79386
other	2.368079	.8451325	2.42	0.016	1.176562	4.766257
smoke	2.517698	.8284259	2.81	0.005	1.321062	4.79826
ptl	1.719161	.6676221	1.40	0.163	.8030814	3.680219
ht	6.249602	4.066275	2.82	0.005	1.74591	22.37086
ui	2.1351	1.093144	1.48	0.138	.7827337	5.824014
_cons	1.586014	1.661913	0.44	0.660	.2034094	12.36639

```
Note: _cons estimates baseline odds.
```

The standard errors are similar to the standard errors we have previously obtained, whether we used the robust or conventional estimators. In this example, we invented the hospital IDs randomly.

Here are the results of the estimation with the same data but with a different set of hospital IDs:

```
. use http://www.stata-press.com/data/r15/hospid2
. logistic low age lwt i.race smoke ptl ht ui, vce(cluster hospid)
Logistic regression                             Number of obs     =        189
                                                Wald chi2(8)      =       7.19
                                                Prob > chi2       =     0.5167
Log pseudolikelihood =   -100.724               Pseudo R2         =     0.1416
                               (Std. Err. adjusted for 20 clusters in hospid)
```

low	Odds Ratio	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.9732636	.0293064	-0.90	0.368	.9174862	1.032432
lwt	.9849634	.0106123	-1.41	0.160	.9643817	1.005984
race						
black	3.534767	3.120338	1.43	0.153	.6265521	19.9418
other	2.368079	1.297738	1.57	0.116	.8089594	6.932114
smoke	2.517698	1.570287	1.48	0.139	.7414969	8.548655
ptl	1.719161	.6799153	1.37	0.171	.7919045	3.732161
ht	6.249602	7.165454	1.60	0.110	.660558	59.12808
ui	2.1351	1.411977	1.15	0.251	.5841231	7.804266
_cons	1.586014	1.946253	0.38	0.707	.1431423	17.573

```
Note: _cons estimates baseline odds.
```

Note the strikingly larger standard errors. What happened? In these data, women most likely to have low-birthweight babies are sent to certain hospitals, and the decision on likeliness is based not just on age, smoking history, etc., but on other things that doctors can see but that are not recorded in our data. Thus merely because a woman is at one of the centers identifies her to be more likely to have a low-birthweight baby.

Video examples

Logistic regression, part 1: Binary predictors

Logistic regression, part 2: Continuous predictors

Logistic regression, part 3: Factor variables

Stored results

`logistic` stores the following in `e()`:

Scalars

`e(N)`	number of observations
`e(N_cds)`	number of completely determined successes
`e(N_cdf)`	number of completely determined failures
`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_dv)`	number of dependent variables
`e(df_m)`	model degrees of freedom
`e(r2_p)`	pseudo-R-squared
`e(ll)`	log likelihood
`e(ll_0)`	log likelihood, constant-only model
`e(N_clust)`	number of clusters
`e(chi2)`	χ^2
`e(p)`	significance of model test
`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)`	`logistic`
`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 χ^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(estat_cmd)`	program used to implement `estat`
`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(Cns)`	constraints matrix
`e(ilog)`	iteration log (up to 20 iterations)
`e(gradient)`	gradient vector
`e(mns)`	vector of means of the independent variables
`e(rules)`	information about perfect predictors
`e(V)`	variance–covariance matrix of the estimators
`e(V_modelbased)`	model-based variance

Functions

`e(sample)`	marks estimation sample

Methods and formulas

Define $\mathbf{x}_j$ as the (row) vector of independent variables, augmented by 1, and $\mathbf{b}$ as the corresponding estimated parameter (column) vector. The logistic regression model is fit by `logit`; see [R] **logit** for details of estimation.

The odds ratio corresponding to the ith coefficient is $\psi_i = \exp(b_i)$. The standard error of the odds ratio is $s_i^{\psi} = \psi_i s_i$, where s_i is the standard error of b_i estimated by `logit`.

Define $I_j = \mathbf{x}_j\mathbf{b}$ as the predicted index of the jth observation. The predicted probability of a positive outcome is

$$p_j = \frac{\exp(I_j)}{1 + \exp(I_j)}$$

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*.

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

References

Archer, K. J., and S. A. Lemeshow. 2006. Goodness-of-fit test for a logistic regression model fitted using survey sample data. *Stata Journal* 6: 97–105.

Brady, A. R. 1998. sbe21: Adjusted population attributable fractions from logistic regression. *Stata Technical Bulletin* 42: 8–12. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, pp. 137–143. College Station, TX: Stata Press.

Buis, M. L. 2010a. Direct and indirect effects in a logit model. *Stata Journal* 10: 11–29.

——. 2010b. Stata tip 87: Interpretation of interactions in nonlinear models. *Stata Journal* 10: 305–308.

Cleves, M. A., and A. Tosetto. 2000. sg139: Logistic regression when binary outcome is measured with uncertainty. *Stata Technical Bulletin* 55: 20–23. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 152–156. College Station, TX: Stata Press.

de Irala-Estévez, J., and M. A. Martínez. 2000. sg125: Automatic estimation of interaction effects and their confidence intervals. *Stata Technical Bulletin* 53: 29–31. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 270–273. College Station, TX: Stata Press.

Dupont, W. D. 2009. *Statistical Modeling for Biomedical Researchers: A Simple Introduction to the Analysis of Complex Data*. 2nd ed. Cambridge: Cambridge University Press.

Freese, J. 2002. Least likely observations in regression models for categorical outcomes. *Stata Journal* 2: 296–300.

Garrett, J. M. 1997. sbe14: Odds ratios and confidence intervals for logistic regression models with effect modification. *Stata Technical Bulletin* 36: 15–22. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 104–114. College Station, TX: Stata Press.

Gould, W. W. 2000. sg124: Interpreting logistic regression in all its forms. *Stata Technical Bulletin* 53: 19–29. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 257–270. College Station, TX: Stata Press.

Hilbe, J. M. 1997. sg63: Logistic regression: Standardized coefficients and partial correlations. *Stata Technical Bulletin* 35: 21–22. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 162–163. College Station, TX: Stata Press.

——. 2009. *Logistic Regression Models*. Boca Raton, FL: Chapman & Hill/CRC.

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

Kleinbaum, D. G., and M. Klein. 2010. *Logistic Regression: A Self-Learning Text*. 3rd ed. New York: Springer.

Lalanne, C., and M. Mesbah. 2016. *Biostatistics and Computer-based Analysis of Health Data Using Stata*. London: ISTE Press.

Lemeshow, S. A., and J.-R. L. Gall. 1994. Modeling the severity of illness of ICU patients: A systems update. *Journal of the American Medical Association* 272: 1049–1055.

Lemeshow, S. A., and D. W. Hosmer, Jr. 2005. Logistic regression. In Vol. 2 of *Encyclopedia of Biostatistics*, ed. P. Armitage and T. Colton, 2870–2880. Chichester, UK: Wiley.

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

Mehmetoglu, M., and T. G. Jakobsen. 2017. *Applied Statistics Using Stata: A Guide for the Social Sciences*. Thousand Oaks, CA: Sage.

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.

Mitchell, M. N., and X. Chen. 2005. Visualizing main effects and interactions for binary logit models. *Stata Journal* 5: 64–82.

Pagano, M., and K. Gauvreau. 2000. *Principles of Biostatistics*. 2nd ed. Belmont, CA: Duxbury.

Pampel, F. C. 2000. *Logistic Regression: A Primer*. Thousand Oaks, CA: Sage.

Paul, C. 1998. sg92: Logistic regression for data including multiple imputations. *Stata Technical Bulletin* 45: 28–30. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 180–183. College Station, TX: Stata Press.

Pearce, M. S. 2000. sg148: Profile likelihood confidence intervals for explanatory variables in logistic regression. *Stata Technical Bulletin* 56: 45–47. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 211–214. College Station, TX: Stata Press.

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

Reilly, M., and A. Salim. 2000. sg156: Mean score method for missing covariate data in logistic regression models. *Stata Technical Bulletin* 58: 25–27. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 256–258. College Station, TX: Stata Press.

Schonlau, M. 2005. Boosted regression (boosting): An introductory tutorial and a Stata plugin. *Stata Journal* 5: 330–354.

Vittinghoff, E., D. V. Glidden, S. C. Shiboski, and C. E. McCulloch. 2012. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. 2nd ed. New York: Springer.

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] **logistic postestimation** — Postestimation tools for logistic

[R] **brier** — Brier score decomposition

[R] **cloglog** — Complementary log-log regression

[R] **exlogistic** — Exact logistic regression

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

[R] **roc** — Receiver operating characteristic (ROC) analysis

[BAYES] **bayes: logistic** — Bayesian logistic regression, reporting odds ratios

[FMM] **fmm: logit** — Finite mixtures of logistic regression models

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

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

[XT] **xtlogit** — Fixed-effects, random-effects, and population-averaged logit models

[U] 20 Estimation and postestimation commands