

truncreg — Truncated regression

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

Description

`truncreg` fits a regression model of *depvar* on *indepvars* from a sample drawn from a restricted part of the population. Under the normality assumption for the whole population, the error terms in the truncated regression model have a truncated normal distribution, which is a normal distribution that has been scaled upward so that the distribution integrates to one over the restricted range.

Quick start

Truncated regression of `y` on `x1` and `x2` truncated below 16

```
truncreg y x1 x2, ll(16)
```

Specify that `y` is truncated above 35

```
truncreg y x1 x2, ul(35)
```

With `y` truncated below 17 and above 35

```
truncreg y x1 x2, ll(17) ul(35)
```

Specify a lower truncation point that varies across observations using the variable `trunc`

```
truncreg y x1 x2, ll(trunc)
```

Same as above, but with bootstrap standard errors using 200 replications

```
truncreg y x1 x2, ll(trunc) vce(bootstrap, reps(200))
```

See last estimates with legend of coefficient names instead of statistics

```
truncreg, coeflegend
```

Menu

Statistics > Linear models and related > Truncated regression

Syntax

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

<i>options</i>	Description
Model	
<code>noconstant</code>	suppress constant term
<code>ll(<i>varname</i> #)</code>	left-truncation variable or limit
<code>ul(<i>varname</i> #)</code>	right-truncation variable or limit
<code>offset(<i>varname</i>)</code>	include <i>varname</i> in model with coefficient constrained to 1
<code>constraints(<i>constraints</i>)</code>	apply specified linear constraints
SE/Robust	
<code>vce(<i>vcetype</i>)</code>	<i>vcetype</i> may be <code>oim</code> , <code>robust</code> , <code>cluster</code> <i>clustvar</i> , <code>opg</code> , <code>bootstrap</code> , or <code>jackknife</code>
Reporting	
<code>level(#)</code>	set confidence level; default is <code>level(95)</code>
<code>lrmodel</code>	perform the likelihood-ratio model test instead of the default Wald test
<code>nocnsreport</code>	do not display constraints
<code>display_options</code>	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Maximization	
<code>maximize_options</code>	control the maximization process; seldom used
<code>collinear</code>	keep collinear variables
<code>coeflegend</code>	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`, `collect`, `fmm`, `fp`, `jackknife`, `mi estimate`, `rolling`, `statsby`, and `svy` are allowed; see [U] 11.1.10 **Prefix commands**. For more details, see [BAYES] `bayes: truncreg` and [FMM] `fmm: truncreg`.

`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**.

`aweight`s are not allowed with the `jackknife` prefix; see [R] **jackknife**.

`vce()`, `lrmodel`, and `weights` are not allowed with the `svy` prefix; see [SVY] **svy**.

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

`noconstant`; see [R] **Estimation options**.

`ll(varname | #)` and `ul(varname | #)` indicate the lower and upper limits for truncation, respectively. You may specify one or both. Observations with `depvar ≤ ll()` are left-truncated, observations with `depvar ≥ ul()` are right-truncated, and the remaining observations are not truncated. See [R] **tobit** for a more detailed description.

`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(#)`, `lrmodel`, `nocnsreport`; see [R] [Estimation options](#).

`display_options`: `nocl`, `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, but you may use the `ltol(#)` option to relax the convergence criterion; the default is $1e-6$ during specification searches.

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

The following options are available with `truncreg` but are not shown in the dialog box:

`collinear`, `coeflegend`; see [R] [Estimation options](#).

Remarks and examples

[stata.com](http://www.stata.com)

Truncated regression fits a model of a dependent variable on independent variables from a restricted part of a population. Truncation is essentially a characteristic of the distribution from which the sample data are drawn. If x has a normal distribution with mean μ and standard deviation σ , the density of the truncated normal distribution is

$$\begin{aligned} f(x \mid a < x < b) &= \frac{f(x)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \\ &= \frac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \end{aligned}$$

where ϕ and Φ are the density and distribution functions of the standard normal distribution.

Compared with the mean of the untruncated variable, the mean of the truncated variable is greater if the truncation is from below, and the mean of the truncated variable is smaller if the truncation is from above. Moreover, truncation reduces the variance compared with the variance in the untruncated distribution.

▷ Example 1

We will demonstrate `truncreg` with part of the Mroz dataset distributed with [Berndt \(1996\)](#). This dataset contains 753 observations on women's labor supply. Our subsample is of 250 observations, with 150 market laborers and 100 nonmarket laborers.

```
. use https://www.stata-press.com/data/r18/laborsub
. describe
Contains data from https://www.stata-press.com/data/r18/laborsub.dta
Observations:      250
Variables:         6                               25 Sep 2022 18:36
```

Variable name	Storage type	Display format	Value label	Variable label
lfp	byte	%9.0g		1 if woman worked in 1975
whrs	int	%9.0g		Wife's hours of work
kl6	byte	%9.0g		# of children younger than 6
k618	byte	%9.0g		# of children between 6 and 18
wa	byte	%9.0g		Wife's age
we	byte	%9.0g		Wife's educational attainment

Sorted by:

```
. summarize, sep(0)
```

Variable	Obs	Mean	Std. dev.	Min	Max
lfp	250	.6	.4908807	0	1
whrs	250	799.84	915.6035	0	4950
kl6	250	.236	.5112234	0	3
k618	250	1.364	1.370774	0	8
wa	250	42.92	8.426483	30	60
we	250	12.352	2.164912	5	17

We first perform ordinary least-squares estimation on the market laborers.

```
. regress whrs kl6 k618 wa we if whrs > 0
```

Source	SS	df	MS	Number of obs	=	150
Model	7326995.15	4	1831748.79	F(4, 145)	=	2.80
Residual	94793104.2	145	653745.546	Prob > F	=	0.0281
				R-squared	=	0.0717
				Adj R-squared	=	0.0461
Total	102120099	149	685369.794	Root MSE	=	808.55

whrs	Coefficient	Std. err.	t	P> t	[95% conf. interval]
kl6	-421.4822	167.9734	-2.51	0.013	-753.4748 -89.48953
k618	-104.4571	54.18616	-1.93	0.056	-211.5538 2.639668
wa	-4.784917	9.690502	-0.49	0.622	-23.9378 14.36797
we	9.353195	31.23793	0.30	0.765	-52.38731 71.0937
_cons	1629.817	615.1301	2.65	0.009	414.0371 2845.597

Now, we use `truncreg` to perform truncated regression with truncation from below zero.

```
. truncreg whrs kl6 k618 wa we, ll(0)
(100 obs truncated)
```

Fitting full model:

```
Iteration 0: Log likelihood = -1205.6992
Iteration 1: Log likelihood = -1200.9873
Iteration 2: Log likelihood = -1200.9159
```

Iteration 3: Log likelihood = -1200.9157

Iteration 4: Log likelihood = -1200.9157

Truncated regression

Limit: Lower = 0	Number of obs = 150
Upper = +inf	Wald chi2(4) = 10.05
Log likelihood = -1200.9157	Prob > chi2 = 0.0395

whrs	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
kl6	-803.0042	321.3614	-2.50	0.012	-1432.861	-173.1474
k618	-172.875	88.72898	-1.95	0.051	-346.7806	1.030578
wa	-8.821123	14.36848	-0.61	0.539	-36.98283	19.34059
we	16.52873	46.50375	0.36	0.722	-74.61695	107.6744
_cons	1586.26	912.355	1.74	0.082	-201.9233	3374.442
/sigma	983.7262	94.44303	10.42	0.000	798.6213	1168.831

If we assume that our data were censored, the tobit model is

```
. tobit whrs kl6 k618 wa we, ll(0)
```

(output omitted)

Tobit regression	Number of obs = 250
	Uncensored = 150
Limits: Lower = 0	Left-censored = 100
Upper = +inf	Right-censored = 0
	LR chi2(4) = 23.03
	Prob > chi2 = 0.0001
Log likelihood = -1367.0903	Pseudo R2 = 0.0084

whrs	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
kl6	-827.7655	214.7521	-3.85	0.000	-1250.753	-404.7781
k618	-140.0191	74.22719	-1.89	0.060	-286.221	6.182766
wa	-24.97918	13.25715	-1.88	0.061	-51.09118	1.13281
we	103.6896	41.82629	2.48	0.014	21.30625	186.0729
_cons	589.0002	841.5952	0.70	0.485	-1068.651	2246.652
var(e.whrs)	1715859	216775.7			1337864	2200650

◀

□ Technical note

Whether truncated regression is more appropriate than the ordinary least-squares estimation depends on the purpose of that estimation. If we are interested in the mean of wife's working hours conditional on the subsample of market laborers, least-squares estimation is appropriate. However if we are interested in the mean of wife's working hours regardless of market or nonmarket labor status, least-squares estimates could be seriously misleading.

Truncation and censoring are different concepts. A sample has been censored if no observations have been systematically excluded but some of the information contained in them has been suppressed. In a truncated distribution, only the part of the distribution above (or below, or between) the truncation points is relevant to our computations. We need to scale it up by the probability that an observation falls in the range that interests us to make the distribution integrate to one. The censored distribution used by tobit, however, is a mixture of discrete and continuous distributions. Instead of rescaling over the observable range, we simply assign the full probability from the censored regions to the censoring points. The truncated regression model is sometimes less well behaved than the tobit model.

Davidson and MacKinnon (1993) provide an example where truncation results in more inconsistency than censoring.



Stored results

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

Scalars

<code>e(N)</code>	number of observations
<code>e(N_bf)</code>	number of observations before truncation
<code>e(chi2)</code>	model χ^2
<code>e(k_eq)</code>	number of equations in <code>e(b)</code>
<code>e(k_eq_model)</code>	number of equations in overall model test
<code>e(k_aux)</code>	number of auxiliary parameters
<code>e(df_m)</code>	model degrees of freedom
<code>e(ll)</code>	log likelihood
<code>e(ll_0)</code>	log likelihood, constant-only model
<code>e(N_clust)</code>	number of clusters
<code>e(sigma)</code>	estimate of sigma
<code>e(p)</code>	p -value for model test
<code>e(rank)</code>	rank of <code>e(V)</code>
<code>e(ic)</code>	number of iterations
<code>e(rc)</code>	return code
<code>e(converged)</code>	1 if converged, 0 otherwise

Macros

<code>e(cmd)</code>	<code>truncreg</code>
<code>e(cmdline)</code>	command as typed
<code>e(llopt)</code>	contents of <code>ll()</code> , if specified
<code>e(ulopt)</code>	contents of <code>ul()</code> , if specified
<code>e(depvar)</code>	name of dependent variable
<code>e(wtype)</code>	weight type
<code>e(wexp)</code>	weight expression
<code>e(title)</code>	title in estimation output
<code>e(clustvar)</code>	name of cluster variable
<code>e(offset1)</code>	offset
<code>e(chi2type)</code>	Wald or LR; type of model χ^2 test
<code>e(vce)</code>	<i>vcetype</i> specified in <code>vce()</code>
<code>e(vcetype)</code>	title used to label Std. err.
<code>e(opt)</code>	type of optimization
<code>e(which)</code>	max or min; whether optimizer is to perform maximization or minimization
<code>e(ml_method)</code>	type of ml method
<code>e(user)</code>	name of likelihood-evaluator program
<code>e(technique)</code>	maximization technique
<code>e(properties)</code>	<code>b V</code>
<code>e(predict)</code>	program used to implement <code>predict</code>
<code>e(marginsok)</code>	predictions allowed by <code>margins</code>
<code>e(asbalanced)</code>	factor variables <code>fvset</code> as <code>asbalanced</code>
<code>e(asobserved)</code>	factor variables <code>fvset</code> as <code>asobserved</code>

Matrices

<code>e(b)</code>	coefficient vector
<code>e(Cns)</code>	constraints matrix
<code>e(ilog)</code>	iteration log (up to 20 iterations)
<code>e(gradient)</code>	gradient vector
<code>e(V)</code>	variance-covariance matrix of the estimators
<code>e(V_modelbased)</code>	model-based variance
<code>e(means)</code>	means of independent variables
<code>e(dummy)</code>	indicator for dummy variables

Functions

<code>e(sample)</code>	marks estimation sample
------------------------	-------------------------

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

Matrices	
<code>r(table)</code>	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

Greene (2018, 918–924) and Davidson and MacKinnon (1993, 534–537) provide introductions to the truncated regression model.

Let $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ be the model. \mathbf{y} represents continuous outcomes either observed or not observed. Our model assumes that $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$.

Let a be the lower limit and b be the upper limit. The log likelihood is

$$\ln L = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^n (y_j - \mathbf{x}_j\boldsymbol{\beta})^2 - \sum_{j=1}^n \log \left\{ \Phi\left(\frac{b - \mathbf{x}_j\boldsymbol{\beta}}{\sigma}\right) - \Phi\left(\frac{a - \mathbf{x}_j\boldsymbol{\beta}}{\sigma}\right) \right\}$$

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](#).

`truncreg` also supports estimation with survey data. For details on VCEs with survey data, see [SVY] [Variance estimation](#).

References

- Berndt, E. R. 1996. *The Practice of Econometrics: Classic and Contemporary*. New York: Addison–Wesley.
- Canette, I. 2016. Understanding truncation and censoring. *The Stata Blog: Not Elsewhere Classified*. <http://blog.stata.com/2016/12/13/understanding-truncation-and-censoring/>.
- Davidson, R., and J. G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- Gray, L. A., and M. Hernández-Alava. 2018. A command for fitting mixture regression models for bounded dependent variables using the beta distribution. *Stata Journal* 18: 51–75.
- Greene, W. H. 2018. *Econometric Analysis*. 8th ed. New York: Pearson.

Also see

- [R] [truncreg postestimation](#) — Postestimation tools for `truncreg`
- [R] [regress](#) — Linear regression
- [R] [tobit](#) — Tobit regression
- [BAYES] [bayes: truncreg](#) — Bayesian truncated regression
- [FMM] [fmm: truncreg](#) — Finite mixtures of truncated linear 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

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.

