1 Read this—it will help

Contents

1.1 Getting started with Stata

To get started with Stata, you should

1. Read the *Getting Started* (GSM, GSU, or GSW) manual for your operating system.

2. Then turn to the other manuals; see [U] **1.2 The Stata Documentation**.

1.2 The Stata Documentation

The *User's Guide* is divided into three sections: *Stata basics*, *Elements of Stata*, and *Advice*. The table of contents lists the chapters within each of these sections. Click on the chapter titles to see the detailed contents of each chapter.

The *Guide* is full of a lot of useful information about Stata; we recommend that you read it. If you only have time, however, to read one or two chapters, then read [U] **11 Language syntax** and [U] **12 Data**.

The other manuals are the *Reference* manuals. The Stata *Reference* manuals are each arranged like an encyclopedia—alphabetically. Look at the *Base Reference Manual*. Look under the name of a command. If you do not find the command, look in the subject index in [I] *Stata Index*. A few commands are so closely related that they are documented together, such as ranksum and median, which are both documented in [R] **ranksum**.

Not all the entries in the *Base Reference Manual* are Stata commands; some contain technical information, such as [R] **Maximize**, which details Stata's iterative maximization process, or [R] **Error messages**, which provides information on error messages and return codes.

Like an encyclopedia, the *Reference* manuals are not designed to be read from cover to cover. When you want to know what a command does, complete with all the details, qualifications, and pitfalls, or when a command produces an unexpected result, read its description. Each entry is written at the level of the command. The descriptions assume that you have little knowledge of Stata's features when they are explaining simple commands, such as those for using and saving data. For more complicated commands, they assume that you have a firm grasp of Stata's other features.

If a Stata command is not in the *Base Reference Manual*, you can find it in one of the other *Reference* manuals. The titles of the manuals indicate the types of commands that they contain. The *Programming Reference Manual*, however, contains commands not only for programming Stata but also for manipulating matrices (not to be confused with the matrix programming language described in the *Mata Reference Manual*).

The complete list of Stata Documentation is as follows:

[GSM]	*Getting Started with Stata for Mac*
[GSU]	*Getting Started with Stata for Unix*
[GSW]	*Getting Started with Stata for Windows*
[U]	*Stata User's Guide*
[R]	*Stata Base Reference Manual*
[ADAPT]	*Stata Adaptive Designs: Group Sequential Trials Reference Manual*
[BAYES]	*Stata Bayesian Analysis Reference Manual*
[BMA]	*Stata Bayesian Model Averaging Reference Manual*
[CAUSAL]	*Stata Causal Inference and Treatment-Effects Estimation Reference Manual*
[CM]	*Stata Choice Models Reference Manual*
[D]	*Stata Data Management Reference Manual*
[DSGE]	*Stata Dynamic Stochastic General Equilibrium Models Reference Manual*
[ERM]	*Stata Extended Regression Models Reference Manual*
[FMM]	*Stata Finite Mixture Models Reference Manual*
[FN]	*Stata Functions Reference Manual*
[G]	*Stata Graphics Reference Manual*
[H2OML]	*Machine Learning in Stata Using H2O: Ensemble Decision Trees Reference Manual*
[IRT]	*Stata Item Response Theory Reference Manual*
[LASSO]	*Stata Lasso Reference Manual*
[XT]	*Stata Longitudinal-Data/Panel-Data Reference Manual*
[META]	*Stata Meta-Analysis Reference Manual*
[ME]	*Stata Multilevel Mixed-Effects Reference Manual*
[MI]	*Stata Multiple-Imputation Reference Manual*
[MV]	*Stata Multivariate Statistics Reference Manual*
[PSS]	*Stata Power, Precision, and Sample-Size Reference Manual*
[P]	*Stata Programming Reference Manual*
[RPT]	*Stata Reporting Reference Manual*
[SP]	*Stata Spatial Autoregressive Models Reference Manual*
[SEM]	*Stata Structural Equation Modeling Reference Manual*
[SVY]	*Stata Survey Data Reference Manual*
[ST]	*Stata Survival Analysis Reference Manual*
[TABLES]	*Stata Customizable Tables and Collected Results Reference Manual*
[TS]	*Stata Time-Series Reference Manual*
[I]	*Stata Index*
[M]	*Mata Reference Manual*

In addition, installation instructions may be found in the *Installation Guide*.

1.2.1 PDF manuals

Every copy of Stata comes with Stata's complete PDF documentation.

The PDF documentation may be accessed from within Stata by selecting **Help** > **PDF documentation**. Even more convenient, every help file in Stata links to the equivalent manual entry. If you are reading **help regress**, simply click on (`View complete PDF manual entry`) below the title of the help file to go directly to the [R] **regress** manual entry.

We provide some tips for viewing Stata's PDF documentation at https://www.stata.com/support/faqs/resources/pdf-documentation-tips/.

Video example

PDF documentation in Stata

1.2.2 Example datasets

Various examples in this manual use what is referred to as the automobile dataset, `auto.dta`. We have created a dataset on the prices, mileages, weights, and other characteristics of 74 automobiles and have saved it in a file called `auto.dta`. (These data originally came from the April 1979 issue of *Consumer Reports* and from the United States Government EPA statistics on fuel consumption; they were compiled and published by Chambers et al. [1983].)

In our examples, you will often see us type

```
. use https://www.stata-press.com/data/r19/auto
```

We include the `auto.dta` file with Stata. If you want to use it from your own computer rather than via the internet, you can type

```
. sysuse auto
```

See [D] **sysuse**.

You can also access `auto.dta` by selecting **File > Example datasets...**, clicking on *Example datasets installed with Stata*, and clicking on `use` beside the `auto.dta` filename.

There are many other example datasets that ship with Stata or are available over the web. Here is a partial list of the example datasets included with Stata:

auto.dta	1978 automobile data
bplong.dta	Fictional blood-pressure data, long form
bpwide.dta	Fictional blood-pressure data, wide form
cancer.dta	Patient survival in drug trial
census.dta	1980 Census data by state
citytemp.dta	US city temperature data
educ99gdp.dta	Education and gross domestic product
gnp96.dta	US gross national product, 1967–2002
lifeexp.dta	1998 life expectancy
network1.dta	Fictional network diagram data
nlsw88.dta	1988 US National Longitudinal Survey of Young Women (NLSW), extract
pop2000.dta	2000 US Census population, extract
sandstone.dta	Subsea elevation of Lamont sandstone in an area of Ohio
sp500.dta	S&P 500 historic data
surface.dta	NOAA sea surface temperature
tsline1.dta	Simulated time-series data
uslifeexp.dta	US life expectancy, 1900–1999
voter.dta	1992 US presidential voter data

All of these datasets may be used or described from the **Example datasets...** menu listing.

Even more example datasets, including most of the datasets used in the reference manuals, are available at the Stata Press website (https://www.stata-press.com/data/). You can download the datasets with your browser, or you can use them directly from the Stata command line:

```
. use https://www.stata-press.com/data/r19/nlswork
```

An alternative to the use command for these example datasets is webuse. For example, typing

```
. webuse nlswork
```

is equivalent to the above use command. For more information, see [D] **webuse**.

Video example

Example datasets included with Stata

1.2.3 Cross-referencing

The *Getting Started* manual, the *User's Guide*, and the *Reference* manuals cross-reference each other.

[R] **regress**
[D] **reshape**
[XT] **xtreg**

The first is a reference to the regress entry in the *Base Reference Manual*, the second is a reference to the reshape entry in the *Data Management Reference Manual*, and the third is a reference to the xtreg entry in the *Longitudinal-Data/Panel-Data Reference Manual*.

[GSW] **B Advanced Stata usage**
[GSM] **B Advanced Stata usage**
[GSU] **B Advanced Stata usage**

are instructions to see the appropriate section of the *Getting Started with Stata for Windows*, *Getting Started with Stata for Mac*, or *Getting Started with Stata for Unix* manual.

1.2.4 The index

The *Stata Index* contains a combined index for all the manuals.

To find information and commands quickly, you can use Stata's `search` command; see [R] **search**. At the Stata command prompt, type `search geometric mean`. `search` searches Stata's keyword database and the internet to find more commands and extensions for Stata written by Stata users.

1.2.5 The subject table of contents

A subject table of contents for the *User's Guide* and all the *Reference* manuals is located in the *Stata Index*. This subject table of contents may also be accessed by clicking on **Contents** in the PDF bookmarks.

1.2.6 Typography

We mix the ordinary typeface that you are reading now with a typewriter-style typeface `that looks like this`. When something is printed in the typewriter-style typeface, it means that something is a command or an option—it is something that Stata understands and something that you might actually type into your computer. Differences in typeface are important. If a sentence reads, "You could list the result …", it is just an English sentence—you *could* list the result, but the sentence provides no clue as to how you might actually do that. On the other hand, if the sentence reads, "You could `list` the result …", it is telling you much more—you could list the result, and you could do that by using the `list` command.

We will occasionally lapse into periods of inordinate cuteness and write, "We `described` the data and then `listed` the data." You get the idea. `describe` and `list` are Stata commands. We purposely began the previous sentence with a lowercase letter. Because `describe` is a Stata command, it must be typed in lowercase letters. The ordinary rules of capitalization are temporarily suspended in favor of preciseness.

We also mix in words printed in italic type, such as "To perform the rank-sum test, type `ranksum` *varname*, by(*groupvar*)". Italicized words are not supposed to be typed; instead, you are to substitute another word for them.

We would also like users to note our rule for punctuation of quotes. We follow a rule that is often used in mathematics books and British literature. The punctuation mark at the end of the quote is included in the quote only if it is a part of the quote. For instance, the pleased Stata user said she thought that Stata was a "very powerful program". Another user simply said, "I love Stata."

In this manual, however, there is little dialogue, and we follow this rule to precisely clarify what you are to type, as in, type "`cd c:`". The period is outside the quotation mark because you should not type the period. If we had wanted you to type the period, we would have included two periods at the end of the sentence: one inside the quotation and one outside, as in, type "the orthogonal polynomial operator, `p.`".

We have tried not to violate the other rules of English. If you find such violations, they were unintentional and resulted from our own ignorance or carelessness. We would appreciate hearing about them.

We have heard from Nicholas J. Cox of the Department of Geography at Durham University, UK, and express our appreciation. His efforts have gone far beyond dropping us a note, and there is no way with words that we can fully express our gratitude.

1.2.7 Vignette

If you look, for example, at the entry [R] **brier**, you will see a brief biographical vignette of Glenn Wilson Brier (1913–1998), who did pioneering work on the measures described in that entry. A few such vignettes were added without fanfare in the Stata 8 manuals, just for interest, and many more were added in Stata 9, and even more have been added in each subsequent release. A vignette could often appropriately go in several entries. For example, George E. P. Box deserves to be mentioned in entries other than [TS] **arima**, such as [R] **boxcox**. However, to save space, each vignette is given once only, and an index of all vignettes is given in the *Stata Index*.

Most of the vignettes were written by Nicholas J. Cox, Durham University, and were compiled using a wide range of reference books, articles in the literature, internet sources, and information from individuals. Especially useful were the dictionaries of Upton and Cook (2014) and Everitt and Skrondal (2010) and the compilations of statistical biographies edited by Heyde and Seneta (2001) and Johnson and Kotz (1997). Of these, only the first provides information on people living at the time of publication.

1.3   What's new

There are a lot of new features in Stata 19.

For a thorough overview of the most important new features, visit

<div align="center">https://www.stata.com/new-in-stata/</div>

For a brief overview of all the new features that were added with the release of Stata 19, in Stata type

```
. help whatsnew18to19
```

Stata is continually being updated. For a list of new features that have been added since the release of Stata 19, in Stata type

```
. help whatsnew19
```

1.4   References

Chambers, J. M., W. S. Cleveland, B. Kleiner, and P. A. Tukey. 1983. *Graphical Methods for Data Analysis.* Belmont, CA: Wadsworth.

Everitt, B. S., and A. Skrondal. 2010. *The Cambridge Dictionary of Statistics.* 4th ed. Cambridge: Cambridge University Press.

Gould, W. W. 2014. Putting the Stata Manuals on your iPad. *The Stata Blog: Not Elsewhere Classified.* https://blog.stata.com/2014/10/28/putting-the-stata-manuals-on-your-ipad/.

Heyde, C. C., and E. Seneta, eds. 2001. *Statisticians of the Centuries.* New York: Springer.

Johnson, N. L., and S. Kotz, eds. 1997. *Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present.* New York: Wiley.

Pinzon, E., ed. 2015. *Thirty Years with Stata: A Retrospective.* College Station, TX: Stata Press.

Upton, G. J. G., and I. T. Cook. 2014. *A Dictionary of Statistics.* 3rd ed. Oxford: Oxford University Press.

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