

THE STATA NEWS

Volume 15

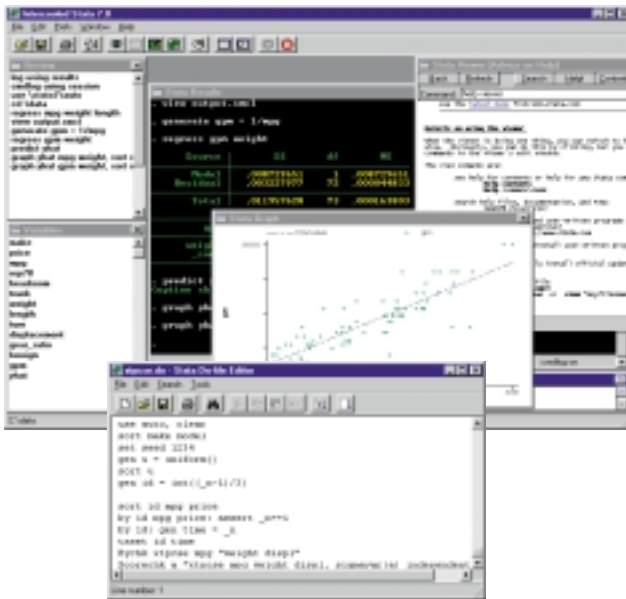
October/November/December 2000

www.stata.com

Stata 7 to ship

Stata 7 updates virtually every aspect of Stata, from statistics to programming language, to the way output is displayed, and even including the new bottom layer on which future graphics will be developed.

We will begin shipping Stata 7 on December 21, 2000, but we recommend ordering early! See ordering details below.



Ordering Stata 7

We will be happy to take your new or upgrade order now. Call
800-782-8272 (US)
800-248-8272 (Canada)
979-696-4600 (worldwide)
979-696-4601 (FAX)

or place your order on-line at www.stata.com.

Orders will be shipped in the order in which they are received.

Inside this issue:

Stata 7 to ship	1
Highlights of Stata 7	1
Features of Stata 7	2
NC-101. Introduction to Stata	5
NC-151. Introduction to Stata Programming	5
New release of Stat/Transfer	6
We are moving!	6
Holiday hours	6

Highlights of Stata 7

Statistics

New survival analysis models and methods
frailty/heterogeneity, continuously time-varying covariates, stratification of distributional parameters, linear models for distributional parameters, person-time, incidence rates, SMR ratios, and more

New panel data estimators
Arellano-Bond, autoregression, instrumental variables, and more

Cluster analysis

Nested logit

Receiver operating characteristics (ROC) curves and analysis

Pharmacokinetics analysis

Marginal effects and elasticities for all estimators

Linear constraints for most maximum likelihood estimators

HAC covariance matrices and standard errors:
heteroskedastic and autocorrelation consistent (HAC) covariance matrices for regression, probit, logit, Poisson, and many other estimators

Dramatic enhancements to generalized linear models (GLMs)
IRLS and ML estimation, HAC standard errors, user-defined links, HAC kernels, and more

Output and interface

Improved output
Output looks better and becomes interactive due to new Stata Markup and Control Language (SMCL)

Windowed interface for Unix

Net search for user-written programs

Language and programming

Long variable names – 32 characters

Ado-files are now byable

Sort stability for all commands

New loop controls for iterating over numbers and lists

8 to 12% faster

Graphics

Line styles

Low-level programming additions for development of new graphs

And, there is much more. See the rest of this issue for an overview and more details. See <http://www.stata.com> for a complete description.

Features of Stata 7

Statistics

Survival analysis (st)

`stcox` will now estimate Cox proportional hazard models with *continuously* time-varying covariates, and you do not need to modify your data to obtain the estimates.

`streg` can now estimate parametric survival models with individual-level frailty (unobserved heterogeneity). Two forms of the frailty distribution are allowed: gamma and inverse gaussian. Frailty is allowed with all the parametric distributions currently available, including exponential, weibull, gompertz, lognormal, loglogistic, and gamma.

`streg` now allows you to estimate stratified models, meaning that the distributional parameters (ancillary parameters and intercept) are allowed to differ across strata.

`streg` has also been modified to allow you to specify any linear-in-the-parameters equation for any of the distributional parameters, which allows you to create various forms of stratification, as well as allowing distributional parameters to be linear functions of other covariates.

`stptime` calculates person-time (person-years) and incidence rates and computes the standardized mortality/morbidity ratios (SMR).

`sts test` has been modified to include additional tests for comparing survivor distributions, including the Tarone–Ware test, the Fleming–Harrington test, and the Peto–Peto–Prentice test. Also new is a test for trend.

`stci` calculates and reports the level and confidence intervals of the survivor function, as well as computing and reporting the mean survival time and confidence interval.

`stsplit` is now much faster and now allows for splitting on failure times, as well as providing some additional convenience options. See [R] `stsplit`, but remember that `stcox` can now estimate with continuous time-varying covariates without you having to `stsplit` the data beforehand.

`stcurve` has a new `outfile` option.

Panel data/Cross-sectional time-series analysis (xt)

`xtabond` produces the Arellano–Bond one-step, one-step robust, and two-step estimators for dynamic panel-data models, models in which there are lagged dependent variables. `xtabond` can be used with exogenously unbalanced panels and, uniquely, handles embedded gaps in the time series as well as opening and closing gaps. Predetermined covariates are allowed and you can select between using the full instrument matrix or a pared down version. Both the Sargan and autocorrelation tests derived by Arellano and Bond are reported.

`xtregar` estimates cross-sectional time-series models in which e_{it} is assumed to follow an AR(1) process and reports both the within estimator and a GLS random-effects estimator. The estimator can handle unequally spaced observations and exogenously unbalanced panels.

`xtregar` uniquely reports the modified Bhargava et al. Durbin–Watson statistic and the Baltagi–Wu locally best invariant test statistic for autocorrelation.

`xtregiv` estimates cross-sectional time-series regressions with (generalized) instrumental variables, or, said differently, estimates two-stage least squares time-series cross-sectional models. `xtregiv` can estimate such models using the between-2SLS estimator, the within-2SLS estimator, the first-differenced 2SLS estimator, the Balestra–Varadharajan–Krishnakumar G2SLS estimator, or the Baltagi EC2SLS estimator. All the estimators allow use of balanced or (exogenously) unbalanced panels.

`xtpcse` produces panel-corrected standard errors (PCSE) for linear cross-sectional time-series models where the parameters are estimated by OLS or Prais–Winsten regression. When computing the standard errors and the variance–covariance estimates, the disturbances are, by default, assumed to be heteroskedastic and contemporaneously correlated across panels.

Estimation commands (other than xt or st)

`nlogit` estimates nested logit models. In a nested logit model, multiple outcomes are grouped into a nested tree structure, and nested logit has the advantage over multinomial and conditional logistic models of allowing you to parameterize away the assumption of independence of irrelevant alternatives (IIA).

`glm` has also been rewritten and its facilities for estimating generalized linear models dramatically extended. It now offers an expanded choice of link functions and also allows user-specified link and variance functions. Newey–West or heteroskedastic autocorrelation consistent (HAC) covariance matrices and standard errors can be estimated. These include options for three standard HAC kernels as well as user-defined kernels. `glm` will now report maximum-likelihood based estimates of standard errors, IRLS based estimates, robust linearization based estimates, bootstrapped and jackknifed estimates, and several others.

`heckman`, `intreg`, `poisson`, `streg` and almost all of Stata’s maximum likelihood estimators will now accept linear constraints. The constraints are defined using the `constraint` command and are specified to the estimator using the new `constraint()` option.

Heteroskedasticity and autocorrelation consistent (HAC) covariance matrices and standard error estimates can now be produced for probit, logit, Poisson, negative binomial, and many other cross-sectional maximum likelihood models. These variance estimates are accessed through the enhancements to `glm` discussed above.

`treatreg` estimates the treatment effects model using either a two-step estimator or a full maximum-likelihood estimator. The treatment effects model considers the effect of an endogenously chosen binary treatment on another endogenous continuous variable, conditional on two sets of independent variables.

`boxcox` has been rewritten. It now produces maximum-likelihood estimates of the coefficients and the Box–Cox transform parameter(s). Box–Cox models may be estimated in various forms, with the transform on the left, on the right, or on both sides.

`truncreg` estimates truncated regression models. Truncated regression refers to regressions estimated on samples drawn based on the dependent variable, and therefore for which (sometimes) neither the dependent nor independent variables are observed.

Cluster analysis

`cluster` performs partitioning and hierarchical cluster analysis using a variety of methods. Two partitioning cluster methods are provided—`kmeans` and `kmedians`—and three hierarchical-cluster methods are provided—single linkage, average linkage, and complete linkage. Included are 14 binary measures and 7 different continuous measures (counting things such as the Minkowski distance # as one).

More than one cluster analysis can be saved simultaneously, so that the results of different analyses may be compared. `cluster` allows adding notes to analyses and, of course, dropping analyses. `cluster` also provides post-clustering commands that can, for instance display the dendrogram (clustering tree) from a hierarchical analysis or produce new grouping variables in the data based on the analysis.

`cluster` is actually a complete system and has been designed to be extended. Users may program extensions for new cluster methods, new cluster management routines, and new post-analysis summary methods.

Commands for epidemiologists

Five new commands are provided for the analysis of Receiver Operating Characteristic (ROC) curves.

`roctab` is used to perform nonparametric ROC analyses. By default, `roctab` calculates the area under the curve. Optionally, `roctab` can plot the ROC curve, display the data in tabular form, and produce Lorenz-like plots.

`rocfite` estimates maximum-likelihood ROC models assuming a binormal distribution of the latent variable. `rocplot` may be used after `rocfite` to plot the fitted ROC curve and simultaneous confidence bands.

`roccomp` tests the equality of two or more ROC areas obtained from applying two or more test modalities to the same sample or to independent samples.

`rocgold` independently tests the equality of the ROC area of each of several test modalities against a “gold” standard ROC curve. For each comparison, `rocgold` reports the raw and the Bonferroni adjusted significance probability. Optionally, Sidak’s adjustment for multiple comparisons can be obtained.

Pharmacokinetics

There are four new estimation commands and two new utilities intended for the analysis of pharmacokinetic data.

`pkexamine` calculates pharmacokinetic measures from time-and-concentration subject-level data. `pkexamine` computes and displays the maximum measured concentration, the time at the maximum measured concentration, the time of the last measurement, the elimina-

tion rate, the half-life, and the area under the concentration-time curve (AUC).

`pksumm` obtains the first four moments from the empirical distribution of each pharmacokinetic measurement and tests the null hypothesis that the measurement is normally distributed.

`pkcross` analyzes data from a crossover design experiment. When analyzing pharmaceutical trial data, if the treatment, carryover, and sequence variables are known, the omnibus test for separability of the treatment and carryover effects is calculated.

`pkequiv` performs bioequivalence testing for two treatments. By default, `pkequiv` calculates a standard confidence interval symmetric about the difference between the two treatment means. Optionally, `pkequiv` calculates confidence intervals symmetric about zero and intervals based on Fieller’s theorem. Additionally, `pkequiv` can perform interval hypothesis tests for bioequivalence.

`pkshape` and `pkcollapse` help in reshaping the data into the form that the above commands need.

Marginal effects

`mfx` reports marginal effects after estimation of *any* model. Marginal effects refers to $df()/dx_i$ evaluated at \mathbf{x} , where $f()$ is any function of the data and the model’s estimated parameters, \mathbf{x} are the model’s covariates, and x_i is one of the covariates. For instance, the model might be probit and $f()$ the cumulative normal distribution, in which case $df()/dx_i$ is the change in the probability of a positive outcome with respect to a change in one of the covariates. \mathbf{x} might be specified as the mean, so that the change would be evaluated at the mean, or, optionally, \mathbf{x} might be specified by the user to be any point in the covariate space.

`mfx` works after estimation of any model in Stata and is capable of producing marginal effects for anything `predict` can produce. For instance, after `tobit`, you could get the marginal effect of the probability of an outcome being uncensored, or the expected value of the uncensored outcome, or the expected value of the censored outcome.

`mfx` can uniquely compute results as derivatives or elasticities.

Other statistical commands

`jkknife` performs jackknife estimation, which is (1) an alternative, first-order unbiased estimator for a statistic; (2) a data-dependent way to calculate the standard error of the statistic and to obtain significance levels and confidence intervals; and (3) a way of producing measures reflecting the observation’s influence on the overall statistic.

`ml` can now perform estimation with linear constraints. All that is required is that you specify the `constraint()` option on the `ml maximize` command.

`statsby` creates a dataset of the results of a command executed by `varlist`: . Results can be any of the saved results, and, if it is an estimation command, can be the coefficients and standard errors.

`lfit`, `lroc`, `lsens`, and `lstat` now work after `probit` just as they do after `logit` or `logistic`.

`drawnorm` draws random samples from a multivariate normal distribution with specified means and covariance matrix.

`corr2data` creates fictional datasets with the specified means and covariance matrix (correlation structure). Thus, you can take published results and duplicate and modify them if the estimator is solely a function of the first two moments of the data, such as `regress`, `ivreg`, `anova`, or `factor`.

`median` performs a nonparametric test that K samples were drawn from populations with the same median.

`tabstat` displays tables of summary statistics, possibly broken down (conditioned) on another variable.

Introducing SMCL

Stata 7 features a new formatting language known as SMCL (pronounced *smicle*), which stands for Stata Markup and Control Language. SMCL delivers enhanced output including solid lines, font control (such as underlining, italics, and bold), intelligent paragraph-wrapping, colors, and, most importantly, the ability to add HTML-like interactivity to output. Log files may be saved in SMCL format so that they retain all attributes of Stata output such as color and font.

All Stata 7 output and help files have been written in SMCL. Stata *ado*-file programmers can produce exciting results using this new language in their output and in help files. Links can be created that jump to other SMCL files, search for and install new Stata commands from the web, or even execute Stata commands. Since all Stata output is SMCL, these links need not be restricted to help files; Stata output can become interactive.

The Viewer

Stata's help window has been transformed into a browser-like window known as the Viewer. In addition to allowing the user to view help files, the Viewer allows you to view and print any SMCL or ASCII file and to convert SMCL files to ASCII if desired. Links in SMCL files in the Viewer can load other files in the Viewer or can even execute Stata commands with output displayed in the Results window. All of Stata's Internet features are accessible from within the Viewer, such as the capability to search for, download, and install new commands with a few mouse-clicks.

The Viewer can be accessed by pulling down **File**, or you can use the new `view` command, which provides some additional features.

Command logs and simultaneous logging

You can now create command logs (ASCII text logs containing only what you type, which used to be called `noprocc` logs) using the new `cmdlog` command. Even better, you can create command logs and full session logs simultaneously.

Windowed Stata now across all platforms

Stata for Unix users now have the same windowed interface that Stata for Windows and Stata for Macintosh users have. We include both the new, windowed interface Stata for Unix and the old console Stata for Unix in Stata 7. The old console version is still useful in batch situations, but Stata(console), as it is now called, can no longer render graphs.

Graphics

Stata now offers line styles. More importantly, Stata now has a new graphics' engine, and all Stata graphs run through the engine. The new graphics' engine includes a graphics' programming environment that will allow development of future Stata graphics additions.

Language enhancements and other new features

There are many other brand new features and new features for existing commands, here we list just a few.

Ado-files now byable and other improvements to `by`

`by varlist`: is now allowed with virtually every Stata command, including commands implemented as *ado*-files, such as `egen`.

`by varlist`: now has a `sort` option, so you no longer have to first sort your data and then type the `by` command. Instead, for instance, you can type `by foreign, sort: summarize mpg`, or, equivalently, `bysort foreign: summarize mpg`.

`by` has a new parenthesis notation: `'by id (time): ...'` means to perform ... *by id*, but first verify that the data are sorted by *id* and *time*. `'by id (time), sort: ...'` says to sort the data by *id* and *time* and then perform ... *by id*.

`by varlist`: also has a new `rc0` option which says to keep on going even if one of the *by*-groups results in an error.

Programmers: making your previously written *ado*-files byable is easy.

New varlist abbreviation rules Varlists now understand `*` when used as other than a suffix, so you can type `pop*99` or `*1999`, in addition to `pop*`. `*` means zero or more changes go here. Also understood is the new `~` abbreviation character. `~` works the same as `*` except it adds the condition that only one variable matches the pattern.

The other new abbreviation character is `?`, which means one character goes here, so `pop?9` would match `pop99` but not `pop1999`.

Sort stability. Commands that report results of calculations (commands not intended to change the data) no longer change the sort order of the data. This includes *ado*-files.

European decimal format. Stata now understands output formats such as `%9,2f` as well as `%9.2f`, so that you can elect to display the number 500.5 as 500,5.

`egen` has 14 new functions, and new options have been added to `group()` and `rank()`.

New programming commands `foreach`, `forvalues`, and

continue `make` stepping through loops easier and give alternatives to `while` and `for`.

The new `hexdump` command gives a hexadecimal dump of files.

`destring` makes converting variables from string to numeric easier.

`net search` searches the web for user-written additions to Stata, displays information and help files for these additions and allows these additions to be installed or uninstalled with just a mouse click. This facility provides direct access to the hundreds of Stata commands published by users in the Stata Technical Bulletin, on Statalist, or on personal websites. `net search` is the latest incarnation of `webseek`, a command not included in Stata 6, but which was made available during the release.

Faster

Although Stata 7 has more features, it is also faster. Ado-files execute between 8.8 and 11.8 percent faster. Some programs, we have observed, execute 13 percent faster.

New documentation

The Stata 7 documentation consists of a *Getting Started* manual (for Windows, Macintosh, or Unix), a *User's Guide*, a *Graphics Manual*, a *Programming Manual*, and a 4-volume *Reference Manual set*.

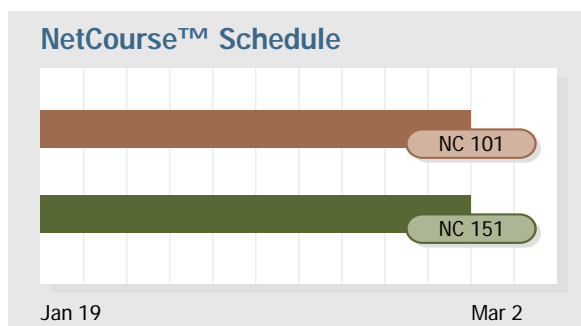
Especially for students is the updated *Stata Reference Manual Extract*, which is a 600+ page subset of the *Stata Reference Manual set*, containing the most important words about the most frequently used commands.

Latest NetCourse™ schedule

The Stata NetCourses™ have been well-received by participants as is reflected in the following remarks:

- “I thought the NetCourse was superior because of the quality of the notes and the emphasis on intuition.”
- “The NetCourses are simply great.”

A brief schedule and summary of the upcoming NetCourses are listed below. For more details on how NetCourses work and for course syllabi, visit www.stata.com/info/products/netcourse.



An enrollment form for the two upcoming NetCourses has been enclosed with the Stata News. You can also enroll online at www.stata.com/info/products/netcourse/enrollment.html.

NC-101. Introduction to Stata

NC-101 is designed to take smart, knowledgeable people and turn them into proficient interactive users of Stata. The course covers not just the obvious such as getting data into Stata, but also covers lots of detailed techniques and tricks to make you a powerful Stata user. From web update features and match-merging to using `by` groups and explicit subscribing, many of Stata's key concepts are explored.

Introduction to Stata	
Prerequisites	Stata 6 is OK, but we recommend Stata 7
Date offered	January 19 – March 2
Course Leaders	Shannon Driver and Jeremy Wernow
Enrollment Deadline	January 15
Price	\$95

NC-151. Introduction to Stata Programming

NC-151 is intended for all Stata users. Through a combination of lectures, example applications, and carefully chosen problems, the course addresses the full range of methods and techniques necessary to be most productive in the Stata environment. Beginning with effective ways to organize both simple and complicated analyses in Stata, NetCourse 151 then moves into programming elements that can be used to work more efficiently. Key programming topics include macro processing, program flow of control, using do-files, programming ado-files, Monte-Carlo simulations, and bootstrapped standard errors.

Introduction to Stata Programming	
Prerequisites	Stata 6 is OK; but we recommend Stata 7
Date offered	January 19 – March 2
Course Leaders	Pete Huckelba and Allen McDowell
Enrollment Deadline	January 15
Price	\$125

New release of Stat/Transfer

Due to the longer names allowed by Stata 7, Stata's dataset format changed in release 7. Of course, Stata 7 will read datasets created by earlier versions.

In addition to Stata Release 7, Stat/Transfer 6.0 adds support for SAS Versions 7 and 8 and MatLab 5. New features are wildcards for tables and worksheets that allow you to move multiple tables in a single command and web self-updating for the Windows version.

Stat/Transfer 6.0 is available for Windows or for Unix : DEC Alpha (OSF/1), HP-9000 (HP-UX), IBM RS/6000 (AIX), Intel 486/Pentium (Linux), SGI MIPS (Irix), Sun SPARC (Solaris).

Holiday hours

StataCorp will be closed on December 25th and January 1st.

We would like to wish everyone a happy and safe holiday season!

We are moving!

Phase I of Stata's new headquarters is almost complete, so we will be moving to 4905 Lakeway Drive, College Station, Texas 77845 sometime in January. The relocation will be accomplished in stages so you may use either the University Drive address or the Lakeway Drive address to send correspondence. Phone numbers remain unchanged.



THE STATA NEWS is published 4 times a year. It is free to all registered users of Stata.



HOW TO REACH US

StataCorp
4905 Lakeway Drive
College Station TX 77845
USA

PHONE 979-696-4600
FAX 979-696-4601
EMAIL stata@stata.com
WEB www.stata.com

Please include your Stata serial number with all correspondence.

INTERNATIONAL DISTRIBUTORS

Applied Statistics & Systems Consultants

Serving Israel
tel: +972 (0)6 6100101
email: assc@netvision.net.il

Axon Technology Company Ltd

Serving Taiwan
tel: +886-(0)2-27045535
email: hank@axon.axon.com.tw

Chips Electronics

Serving Brunei, Indonesia, Malaysia, Singapore
tel: 62 - 21 - 452 17 61
email: puyuh23@indo.net.id

Dittrich & Partner Consulting

Serving Austria, Czech Republic, Germany, Hungary, Poland
tel: +49 2 12 / 26 066 - 0
email: sales@dpc.de

IEM

Serving Botswana, Lesotho, Namibia, Mozambique, South Africa, Swaziland, Zimbabwe
tel: +27-11-8286169
email: iem@hot.co.za

JasonTech Inc

Serving Korea
tel: +82-(0)2-420-6700
email: info@jat.co.kr

Mercostat Consultores

Serving Argentina, Brazil, Paraguay, Uruguay
tel: 598-2-613-7905
email: mercost@adinet.com.uy

Metrika Consulting

Serving the Baltic States, Denmark, Finland, Iceland, Norway, Sweden
tel: +46-708-163128
email: sales@metrika.se

MultION Consulting S.A. de C.V.

Serving Mexico
tel: 52 (5) 559-4050 Ext 190
email: multion@multion.com.mx

Ritme Informatique

Serving Belgium, France, Luxembourg
tel: +33 (0)1 42 46 00 42
email: info@ritme.com

Scientific Solutions S.A.

Serving Switzerland
tel: 41 (0)21 711 15 20
email: info@scientific-solutions.ch

Smit Consult

Serving The Netherlands
tel: +31 416-378 125
email: info@smitconsult.nl

Survey Design & Analysis Services

Serving Australia, New Zealand
tel: +61 (0)3 9878 7373
email: sales@survey-design.com.au

Timberlake Consultants

Serving Eire, U.K.
tel: +44 (0)208 697 3377
email: info@timberlake.co.uk

Timberlake Consultants s.r.l.

Serving Italy
tel: +39 (0)864 210101
email: timberlake@arc.it

Timberlake Consulting S.L.

Serving Spain
tel: +34 (9) 5 422 0648
email: timberlake@zoom.es

Timberlake Consultores, Lda.

Serving Portugal
tel: +351 (0)1 471 73 47
email: timberlake.co@mail.telepac.pt

Vishvas Marketing-Mix Services

Serving India
tel: 91-251-440087
email: vishvas@vsnl.com