Announcing Stata Release 11 STATE 11

Stata 11 ships July 27. Order now at www.stata.com.

Highlights of What's New in Stata 11

Multiple imputation (MI)

- Univariate imputation
- Multivariate imputation
- MI Control Panel
- Manage MI datasets (see page 2)

GMM

- Linear and nonlinear models
- One-step, two-step, and iterative estimators
- Cross-sectional, time-series, and panel data
- Panel-style instruments (see page 6)

Competing-risks regression

- Time-varying covariates
- Cumulative incidence graphs
- Subhazard ratios (see page 4)



Multivariate time series

- State-space models
- Dynamic-factor models
- Multivariate GARCH (see page 5)

PDF documentation

- Comes with every copy of Stata
- · Includes all manuals
- Integrated with help files (see page 6)

Factor variables

- Interactions
- · Polynomial terms
- Create indicators automatically (see page 4)



Panel data and mixed models

- Error covariance structures
- Unit-root tests
- · Standard errors for BLUPs

Marginal analysis

- Estimated marginal means
- Predictive margins
- Average marginal effects (see page 5)

Fonts in graphics

- Italics and bold
- Greek letters
- Mathematical symbols
- Superscripts and subscripts
- Multiple fonts (see page 3)

More

- Ten random-number generators
- More density and distribution functions
- DFBETAs, likelihood displacement values, and LMAX statistics after Cox regression
- Normal, GED, and Student's *t* errors in ARCH
- Multivariate tests
- Numerical derivative functions
- Associative arrays
- Soundex name-matching functions
- Object-oriented programming

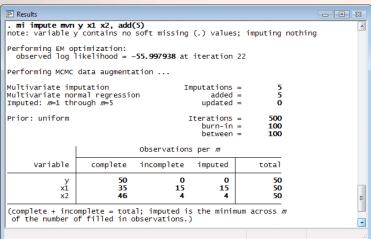
New interface

- Variables Manager
- Live view on data
- Data and variable filters
- Syntax highlighting
- Code folding
- Bookmarks (see page 3)

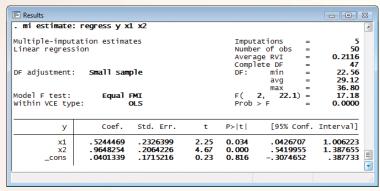
Multiple imputation for missing data

Stata's new mi command provides a full suite of multiple-imputation methods for the analysis of incomplete data, data for which some values are missing. mi provides both the imputation and the estimation steps. mi's estimation step encompasses both estimation on individual datasets and pooling in one easy-to-use procedure. Features are provided to enable examining the pattern of missing values in the data. Flexible imputation methods are provided, including five univariate imputation methods that can be used as building blocks for multivariate imputation, as well as MVN.

We want to study the linear relationship between y and predictors x1 and x2. Our data contain missing values, however, and standard casewise deletion would result in a 40% reduction in sample size! We will fit the model using multiple imputation. First, we impute missing values and arbitrarily create five imputation datasets:



That done, we can fit the model:



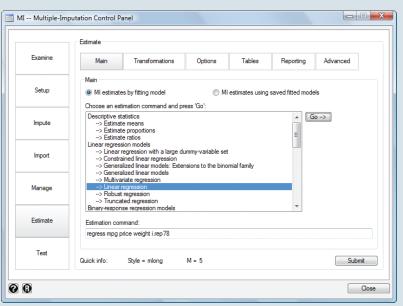
mi estimate fits the specified model (linear regression here) on each of the imputation datasets (five here) and then combines the results into one multiple-imputation inference.

mi can import already imputed data, including data from NHANES or ice, or you can form imputations yourself. Either way, dealing with multiple copies of the data is the bane of multiple-imputation analysis, and Stata solves that problem because mi automatically keeps all datasets in sync. You can create or drop variables or observations just as if you were working with one dataset! You can merge datasets, append datasets, and even reshape your data. The fact that the data management you perform might need to be carried out consistently over 5, 50, or even 500 datasets is irrelevant.

Multiple-Imputation Control Panel

The Multiple-Imputation Control Panel unifies mi's capabilities and guides you from the beginning of your analysis to the end, from examining the pattern of missing values to performing estimation and inference with multiple imputation.

- Use Examine to check missing-value patterns so that you can determine the appropriate imputation method
- Use Setup to set up your data when you need to impute values. If your data already contain imputations, skip Setup and use Import instead.
- Use Impute to create imputations. You can impute one variable or multiple variables simultaneously.
 Regression, predictive mean matching, logistic, ordered logistic, multinomial logistic, and MVN are all available. With monotone missingness, you can combine imputation methods.
- Use Manage to perform data management.
- Use Estimate to fit your model and combine results.
- Use **Test** to perform tests of hypotheses.



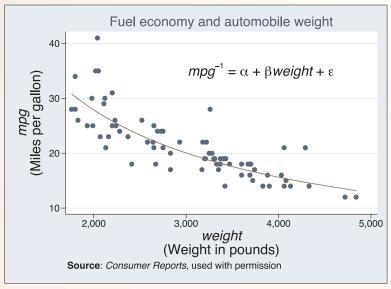
Graphs now support multiple fonts and symbols

You can now include multiple fonts in regular, bold, and italics; the full Greek alphabet in upper- and lowercase; and over 70 mathematical symbols anywhere that text appears in graphs.

Graphs now accept an extended version of Stata's text markup and control language, SMCL. Greek letters and mathematical symbols have intuitive, easy-to-remember names (tags). To make the title of the x axis of your graph say " β -carotene", you just specify xtitle ("{&beta}-carotene"). To include χ^2 in your text, specify {&chi}{superscript:2}, or even shorter, {&chi} {sup:2}.

Additional tags make it easy to select from among sans serif, serif, monospace, and symbol fonts.

Of course, these new features are also available via the dialog boxes and through the interactive Graph Editor.



New Data Editor, Do-file Editor, and Variables Manager

70

72

70

69

69

71

69

71

70

69

70

70

85

253

328

367

378

397

449

468 571

634

641

681

698

739

836

36

53

60

62

64

72

99

112

115

125

128

132

51

51

53

52

50

Filter Observations

Filter by observation

Use in-range qualifie

Vars: 21 Obs: 557 of 28,534 Filter: On Mode: Edit

18

18

18

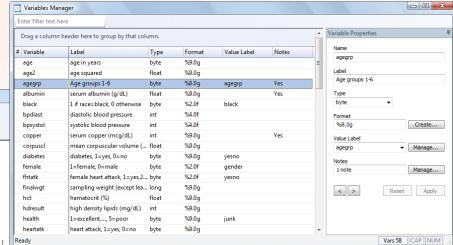
18

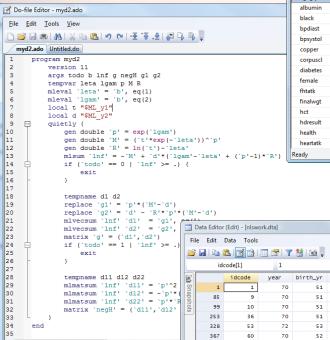
18

18

Stata's new Data Editor, Do-file Editor, and Variables Manager make your data-management and programming tasks simpler than ever.

The new Do-file Editor under Windows includes syntax highlighting, code folding (so you can hide blocks of code while you focus on other parts), the ability to work on multiple files simultaneously, and no limit on file size.





mlmatsum 'lnf' 'd12' = 'p'^2
mlmatsum 'lnf' 'd12' = -'p'*
mlmatsum 'lnf' 'd22' = 'p'*)

matrix 'negH' = ('d11','d12'

The new Variables Manager allows you to change variable names, storage types, formats, labels, and notes with the click of your mouse. You can apply a filter to view only variables that meet criteria you specify.

The new Data Editor allows you to sort variables, apply filters to view subsets of your data, take snapshots (so you can 12 easily undo changes you make), and 12 enter dates and times. The Data Editor 12 is a live view onto your data: change your 12 12 data by running a Stata command, and see the results reflected in the Data Editor 10 instantly.

Apply Filter

Remove Filter

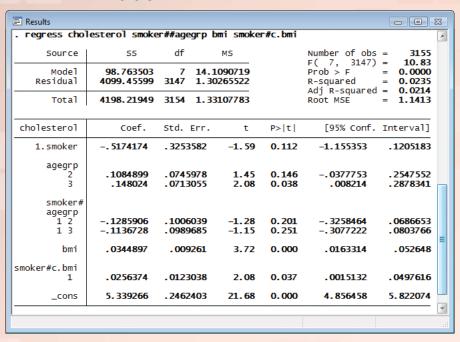
Close

Factor variables

Stata now handles factor (categorical) variables elegantly. You can now prefix a variable with i. to specify indicators for each level (category) of the variable. You can put a # between two variables to create an interaction—indicators for each combination of the categories of the variables. You can put ## instead to specify a full factorial of the variables—main effects for each variable and an interaction. If you want to interact a continuous variable with a factor variable, just prefix the continuous variable with c. You can specify up to eight-way interactions.

We run a linear regression of cholesterol level on a full factorial of age group and whether the person smokes along with a continuous body mass index (bmi) and its interaction with whether the person smokes.

. regress cholesterol smoker##agegrp bmi smoker#c.bmi



We could have used parenthesis binding, to type the same model more briefly:

. regress cholesterol smoker##(agegrp c.bmi)

Base levels can be changed on the fly: i.agegrp uses the default base level of 1, whereas b3.agegrp makes 3 the base level.

The level indicator variables are not created in your dataset, saving lots of space.

Factor variables are integrated deeply into Stata's processing of variable lists, providing a consistent way of interacting with both estimation and postestimation commands.

Competing-risks regression

Competing-risks regression provides a useful alternative to Cox regression in the presence of one or more competing risks. For example, say that you are studying the time from initial treatment for cancer to recurrence of cancer in relation to the type of treatment administered and demographic factors. Death is a competing event: the person under treatment may die, impeding the occurrence of the event of interest, recurrence of cancer. Unlike censoring, which merely obstructs you from viewing the event, a competing event prevents the event of interest from occurring altogether, and your analysis should adjust accordingly.

Stata's new stcrreg command implements competing-risks regression based on Fine and Gray's proportional subhazards model.

In Cox regression, you focus on the survivor function, which indicates the probability of surviving beyond a given time. In competing-risks regression, you instead focus on the cumulative incidence function, which indicates the probability of the event of interest happening before a given time. Competing-risks regression is semiparametric in that the baseline subhazard of the event of interest is left unspecified and the effects of covariates are assumed to be proportional. Time-varying covariates and coefficients are allowed.

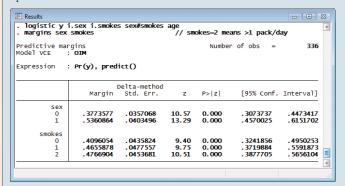
Stata's stcurve command has been fully updated to work with stcrreg and to produce cumulative incidence function graphs, and you can use predict to obtain an array of statistics.

Marginal means, adjusted predictions, and marginal effects

Stata 11 does margins. Margins are statistics calculated from predictions of a previously fit model at fixed values of some covariates and averaging or otherwise integrating over the remaining covariates. Included are estimated marginal means, least-squares means, average and conditional marginal/partial effects (as derivatives or elasticities), average and conditional adjusted predictions, predictive margins, and more.

Predictive margins are particularly apt for survey data and other samples that represent a population (either by weighting or random sampling). In such cases, the margins or marginal effects can be considered estimates of the same statistics in the population or data-generating process (DGP), and their standard errors and confidence intervals can be used to make inferences about the population or DGP.

You can compute the statistics for each subgroup or subpopulation designated by unique values of varlist by using the by(varlist) option.



Because of Stata 11's new factor-variable features, margins understands squared variables and other polynomial forms, so you can get the overall marginal effect for age even when age enters as age and age squared.

Stata 11's margins includes options so that the reported standard errors account for complex survey sampling including weights, sampling units, pre- and poststratification, and subpopulations.

After margins, you can test equality or any linear or nonlinear combination of estimated results.

You can even form linear and nonlinear combinations, so you can examine risk ratios, risk differences, etc., and perform tests with them.

margins works after almost every Stata estimation command.

State-space and dynamic-factor models

Stata's new space and dfactor commands make it easy to fit a wide variety of multivariate time-series models by casting them as linear state-space models, including vector autoregressive moving-average (VARMA) models, structural time-series (STS) models, and dynamic-factor models.

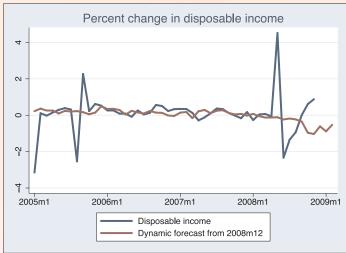
We have data on industrial production (ipman), real disposable income (dsp), weekly hours worked (awhi), and the unemployment rate (unrate). We suspect there exists a latent factor that can explain all four of these series, and we conjecture that the latent factor follows an AR(2) process.

First, we fit our model by typing

```
. dfactor (ipman dsp awhi unrate = , noconstant)  (f = , ar(1/2))
```

With our model fit, let's obtain dynamic forecasts for disposable income beginning in December 2008:

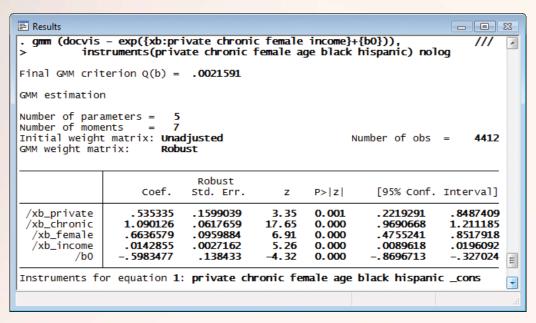
```
. predict dsp_f, dynamic(tm(2008m12))
. tsline dsp dsp f if month >= tm(2005m1)
```



Generalized method of moments (GMM)

Stata's new gmm command makes generalized method of moments estimation as simple as nonlinear least-squares regression and nonlinear seemingly unrelated regression. Just specify your residual equations by using substitutable expressions, list your instruments, select a weight matrix, and obtain your results.

Here we fit a Poisson model of the number of doctor visits as a function of gender, income, and whether a person has a chronic disease or private health insurance. We have reason to believe that income is endogenous, so we use age and race as instruments.



By default, gmm uses the two-step estimator and a weight matrix that assumes the errors are independent but not necessarily identically distributed.

Whether our instruments are valid is certainly open for debate—age likely influences the number of doctor visits—and we can test their validity by using estat overid.

```
. estat overid
Test of overidentifying restriction:
Hansen's J chi2(2) = 9.52598 (p = 0.0085)
```

Hansen's *J* statistic indicates that one or more of our instruments are not valid.

gmm allows for weight matrices that assume observations are independent and identically distributed, independent but not identically distributed, clustered, or autocorrelated. Automatic bandwidth selection is available for heteroskedasticity- and autocorrelation-consistent weight matrices.

In addition to standard instruments, gmm allows you to create panel-style instruments used in dynamic and other panel models with endogenous regressors.

For more complicated analyses, gmm allows you to write a program to evaluate your residual equations instead of using substitutable expressions. These programs are structured like those that ml, nl, and nlsur use. Both the interactive and programmable versions of gmm allow you to specify analytic derivatives.

PDF documentation

Stata now ships with complete PDF documentation! Stata 11's new PDF manuals are fully integrated with Stata's interactive help system. Click and jump to the right manual, the right page! PDF manuals ship with every copy of Stata. Of course, printed documentation is still available.

Order now

To place your order, visit

www.stata.com



Stata Conference DC 09

Mark your calendar now for the 2009 Stata Conference in Washington, DC, on July 30 and 31. The conference will be

your best chance to meet with other Stata users and your only chance this year to talk with Stata developers at a meeting in the United States.

In addition to user contributions, the conference includes overviews and tutorials on new Stata 11 features presented by StataCorp developers.

The tentative schedule for the conference is as follows. It looks to be busy and exciting. For abstracts, please visit

www.stata.com/meeting/dcconf09/abstracts.html

Dates: Thursday and Friday, July 30-31, 2009

Venue: Hotel Monaco, Washington, DC

700 F St. NW

Washington, DC 20004

USA

	Price	Student price
Single day	\$125	\$50
Both days	\$195	\$75
Optional dinner	\$38	

Register: www.stata.com/meeting/dcconf09/

Schedule	
Thursday,	
8:30 AM	Easy approaches to GMM David Drukker, StataCorp
9:20 AM	Mixed-process models with cmp David Roodman, <i>Center for Global Development</i>
10:10 AM	Coffee break
10:25 AM	Multivariate time series David Drukker, <i>StataCorp</i>
11:15 AM	Survey statistics in Stata Jeff Pitblado, StataCorp
12:05 PM	Lunch break
1:00 PM	Regression diagnostics for survey data Rick Valliant, <i>University of Maryland</i>
1:45 PM	Using Stata for subpopulation analysis of complex sample survey data Brady West, <i>University of Michigan</i>
2:30 PM	Coffee break
2:45 PM	Implementing econometric estimators with Mata Christopher F. Baum, <i>Boston College</i>
3:10 PM	Estimating high-dimensional fixed-effects models Paulo Guimaraes, <i>University of South Carolina</i>
3:45 PM	Data envelopment analysis in Stata Choonjoo Lee and Ji Yong-bae, <i>Korea National Defense University</i>
4:15 PM	Coffee break
4:30 PM	Estimating the fractional response model with an endogenous count variable Hoa Nguyen, <i>Michigan State University</i>
4:50 PM	Threshold regression with threg Mei-Ling Ting Lee and Tao Xiao, <i>University of Maryland</i>
5:20 PM	Causal inference Austin Nichols, <i>Urban Institute</i>
6:00 PM	Adjourn

7:00 PM Optional dinner at Oyamel (401 7th St. NW)

Friday, July	y 31
8:30 AM	Handling categorical covariates gracefully Jeff Pitblado, <i>StataCorp</i>
9:20 AM	Between tables and graphs Nicholas J. Cox, <i>Durham University (UK)</i>
10:10 AM	Coffee break
10:25 AM	Integrating Stata into your workflow Bill Rising, StataCorp
11:15 AM	Stata in large-scale development Michael Lokshin, <i>The World Bank</i>
12:05 PM	Lunch break
1:00 PM	Stata for microtargeting using C++ and ODBC Masahiko Aida, <i>Greenberg Quinlan Rosner</i>
1:20 PM	A Stata regression-space database server module
	Mario Alberto Barabino, <i>Bergamo University (Italy)</i>
1:45 PM	Meta-analytic depiction of ordered categorical diagnostic test accuracy in ROC space Ben Dwamena, <i>University of Michigan</i>
2:30 PM	Coffee break
2:45 PM	Stata commands for moving data between PHASE and HaploView Chuck Huber, Texas A&M Health Science Center School of Rural Public Health
3:05 PM	Automated individualized student assessment Stas Kolenikov, <i>University of Missouri</i>
3:30 PM	Altruism squared: The economics of Statalist exchanges Martin Weiss, <i>University of Tuebingen (Germany)</i>
3:50 PM	Implementing custom graphics in Stata Sergiy Radyakin, <i>The World Bank</i>
4:15 PM	Coffee break
4:30 PM	Report to users Bill Gould, StataCorp
5:15 PM	Wishes and grumbles: User feedback and Q&A

Scientific committee

Austin Nichols (chair), *Urban Institute* (austinnichols@gmail.com) Frauke Kreuter, *University of Maryland* (**fkreuter@survey.umd.edu**) Michael Lokshin, World Bank (mlokshin@worldbank.org) Mei-Ling Ting Lee, *University of Maryland* (**mltlee@umd.edu**)

Logistics organizers

Chris Farrar, StataCorp LP (cfarrar@stata.com) Gretchen Farrar, StataCorp LP (gfarrar@stata.com)

Go to www.stata.com/alerts/ and check the box for "Stata Conferences and Users Group meetings" to receive updates about the conference.

One-day course

Introduction to the new features of Stata 11

Learn what is new in Stata 11. This one-day course is aimed at users of Stata 10 and earlier who are interested in learning about the new features in Stata 11. You will learn about diverse topics such as efficient reproducible data management using point-and-click tools, simpler do-file development using the new Do-file Editor, and extended and unified marginal analysis using the margins command and Stata's new factor variables. You will also learn about Stata's new multipleimputation tools. Finally, we will briefly cover some new additions and extensions to Stata's estimation and programming tools.

This promises to be an intense session where you can quickly learn about many of Stata 11's new tools. For more information, see

www.stata.com/training/public.html

International Distributors

Columbia CP. Ltd.

China, Hong Kong, Malaysia, Philippines, Singapore, Thailand, Vietnam

tel: +86-131-47501748 China

- +852-28580080 Hong Kong
- +603-56389268 Malaysia
- +63-2-5647757 Philippines +65-6842 5538 Singapore
- +66-8-17209515 Thailand
- +84-4-7870931 Vietnam
- email: sales@columbiacp.com

Dittrich & Partner Consulting

Czech Republic, Germany, Hungary +49 2 12 / 26 066-0

email: sales@dpc.de

DPC Geschäftsstelle Wien

tel: +49-212-2606624

Austria

email: info@dpc.co.at

DPC Nederland

The Netherlands

tel: +31 (0)76 530 2330 email: sales@dpc-software.nl

East Asia Training & Consultancy

Indonesia, Singapore, Thailand tel: +65 62199062 Singapore,

Indonesia

662 6279000 Thailand email: stata@eastasiatc.com.sq

iXon Technology Company, Ltd.

Taiwan

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

JasonTG

South Korea

tel: +82 (02)470-4143

email: info@jasontg.com

MercoStat Consultores

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

Metrika Consulting

the Baltic States, Denmark, Finland, Iceland Norway, Sweden

tel: +46 (0)8-792 47 47 email: sales@metrika.se

MultiON Consulting S.A. de C.V.

Belize, Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama

tel: 52 (55) 5559 4050

email: direccion@multion.com.mx Quantec Research (Pty), Ltd.

tel: +27-12-3615154 email: software@quantec.co.za

Ritme Informatique

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

Scientific Formosa, Inc.

tel: 886-2-25050525 email: amanda@sciformosa.com.tw

Scientific Solutions, S.A.

Switzerland

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

SOFTWARE Shop, Inc.

Bolivia, Chile, Colombia, Ecuador, Peru, Venezuela

tel: +1-425-651-4090 Bolivia, Ecuador

+ 56-(2)-899-0455 Chile + 57-(1)-619-4000 Colombia

+ 58-(212)-771-5899 Venezuela

+ 51-(1)-706-8197 Peru email: Ventas@SOFTWARF-shop.com

Survey Design & Analysis Services

Australia, New Zealand

tel: +61 (0)3 9878 7373 email: sales@survey-design.com.au

Timberlake Consultants

Eire, Ireland, UK

tel: +44 (0)20 8697 3377

email: info@timberlake.co.uk

Timberlake Consultants Polska Poland

tel: +48 (0)22-8252927

email: info@timberlake.pl

Timberlake Consulting S.L.

Snain

tel: +34 955601430 email: timberlake@timberlakeconsulting.com

Timberlake Consultores Brasil

tel: +55-11-3170-3123 email: info@timberlake.com.br

Timberlake Consultores, Lda.

Portugal

tel: +351 245 308 549 email: info@timberlake.pt

TStat S.r.I. Italy

+39 0864 210101 email: tstat@tstat.it

Vishvas Marketing-Mix Services

India

tel: 91-22-2589 26 39 email: vishvas@vsnl.com

International Resellers

Russia, Armenia, Azerbaijan, Belarus Georgia, Kazakhstan, Kyrgyzstan, Moldova, Tajikistan, Turkmenistan, Ukraine, Uzbekistan tel: +7-495 232-0023 ext. 232

BockvTech, Inc.

tel: +886-2-23618050 email: ravmond@bockvtech.com.tw

email: annakuri@axoft.ru

CHIPS Electronics

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

Informatique, Inc.

Japan

tel: +81-3-3505-1250 email: sales@informatiq.co.jp

Katalogo Software

Brazil tel: +55-11-3405 4507 email: patricias@katalogo.com.bi

MP & Associates

Greece, Cyprus tel: +30-210-7600955 email: info@mpassociates.gr

NFLICA

Japan tel: 81-3-5307-1133 email: softinfo@univcoop.or.jp

Tashtit Scientific Consultants, Ltd.

+972-3-523-0825

email: info@tashtitsoft.co.il Token Communication, Ltd.

tel: +40 364 103256 email: statasales@token.com.rd

TurnTech China Beijing Shiji Tianyan Software

China tel: +86-10-62978511 +86-10-62669193 email: info@sciencesoftware.com.cn

tel: +90 312 446 1866 email: info@uytes.com.tr

How to reach us

StataCorp

USA

4905 Lakeway Drive College Station, TX 77845

Phone 979-696-4600 Fax

Email Web

979-696-4601 service@stata.com www.stata.com

Please include your Stata serial number with all correspondence.

