THE STATA NEWS

Volume 22, Number 4 **October/November/December 2007** www.stata.com

Check out the *Stata Journal*

The *Stata Journal* is a quarterly publication containing articles about statistics, data analysis, teaching methods, and effective use of Stata's language. The Journal publishes peer-reviewed papers together with shorter notes and comments, user-written commands, regular columns, book reviews, and other material of interest to researchers applying statistics in a variety of disciplines.

Edited by H. Joseph Newton and Nicholas J. Cox, the *Stata Journal* is indexed on two of Thomson Scientific's citation indexes—*The Science Citation Index Expanded* and the *CompuMath Citation Index*. To subscribe or purchase past issues of the *Stata Journal*, visit **www.stata-journal.com**.

Highlights of the *Stata Journal*, volume 7, number 3

Each issue of the *Stata Journal* contains a selection of articles from a variety of disciplines, so regardless of your field you will certainly find something that appeals to you. To give you a glimpse of what is available, here we summarize, in order of appearance, the most recent issue's articles.

Robust standard errors for panel regressions with cross-sectional dependence

Daniel Hoechle
Department of Finance
University of Basel
Basel, Switzerland

In many disciplines analyzing large panel datasets is now common, because panels contain more information than simple cross-sectional datasets and thus allow for greater precision in estimation. However, ignoring possible correlation of regression disturbances both within and across panels can lead to incorrect statistical inference. Many researchers use cluster–robust standard errors available with Stata's `cluster` VCE type, but that does not account for the cross-sectional dependence.

In this article, Hoechle presents a new Stata program, `xtscc`, that estimates pooled ordinary least-squares and weighted least-squares regression, as well as fixed-effects (within) regression models with Driscoll and Kraay (*Review of Economics and Statistics* 80: 549–560) standard errors. His Monte Carlo simulations indicate that Driscoll–Kraay standard errors are well calibrated when cross-sectional dependence is present, while erroneously ignoring cross-sectional correlation in the estimation of panel models can lead to severely biased statistical results. He illustrates the `xtscc` program by considering an application from empirical finance. He also proposes a Hausman-type test for fixed effects that is robust to general forms of cross-sectional and temporal dependence.

Estimating parameters of dichotomous and ordinal item response models with gllamm

Xiaohui Zheng and Sophia Rabe-Hesketh
Graduate School of Education
University of California, Berkeley
Berkeley, CA

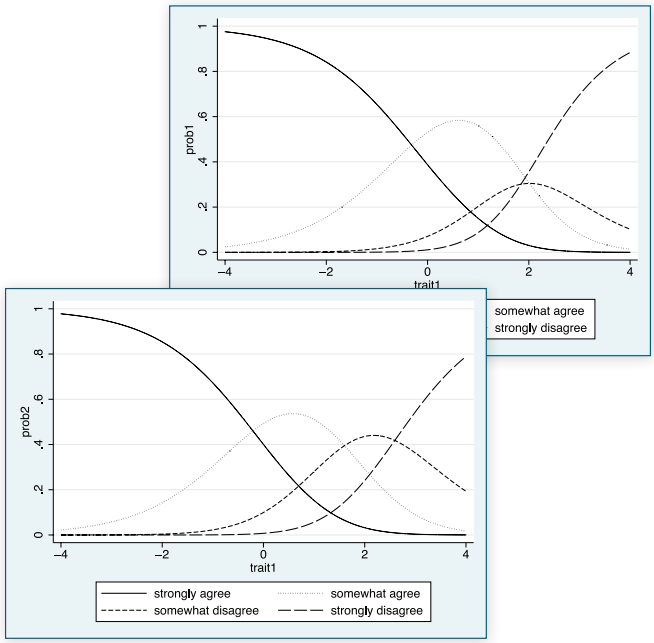
Item response theory models are measurement models for categorical responses. Traditionally, the models are used in educational testing, where responses to test items can be viewed as indirect measures of

Inside this issue:

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

latent ability. A latent variable is a characteristic that is not directly observable. Examples include intelligence, happiness, satisfaction, and attitudes. Latent variables can be measured indirectly through their effects on observable indicators, such as items in achievement tests or psychological questionnaires. The test items are scored either dichotomously (correct/incorrect) or by using an ordinal scale (a grade from poor to excellent). Item response models also apply equally for measurement of other latent traits.

This article describes the one- and two-parameter logit models for dichotomous items, the partial-credit and rating scale models for ordinal items, and an extension of these models where the latent variable is regressed on explanatory variables. Zheng and Rabe-Hesketh show how these models can be expressed as generalized linear latent and mixed models and fitted using the popular user-written command `gllamm`.

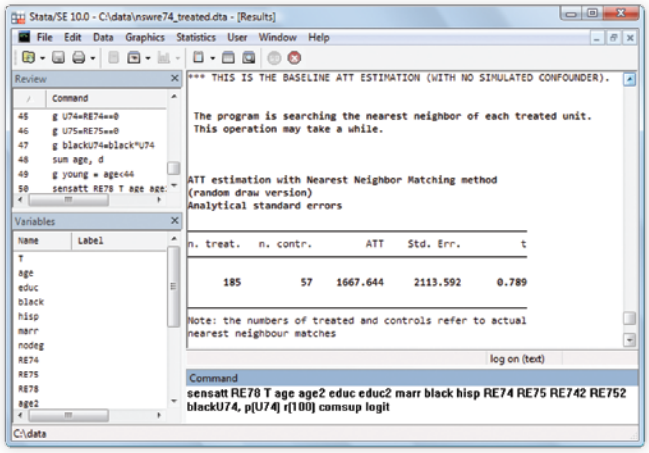


Simulation-based sensitivity analysis for matching estimators

Tommaso Nannicini
Universidad Carlos III de Madrid
Madrid, Spain

The use of matching estimators to estimate treatment effects has skyrocketed in recent years. Early evidence showed that these estimators offered promising performance, though later studies showed that they are only effective when particular conditions are met. Past articles in the *Stata Journal* by Abadie, Drukker, Herr, and Imbens; and by Becker and Ichino presented user-written Stata commands that implement matching estimators. Leuven and Sianesi also have a command available through the SSC archive at Boston College. Because these estimators are appropriate only under certain conditions such as the conditional independence assumption, it is important that users verify that these conditions are met.

This article presents a Stata program, `sensatt`, that implements the sensitivity analysis for matching estimators proposed by Ichino, Mealli, and Nannicini (*Journal of Applied Econometrics*, forthcoming). The analysis simulates effects with respect to deviations from the conditional independence assumption. The program uses the commands for propensity-score matching (`att*`) developed by Becker and Ichino. Nannicini provides an example by using the National Supported Work demonstration, widely known in the program evaluation literature.

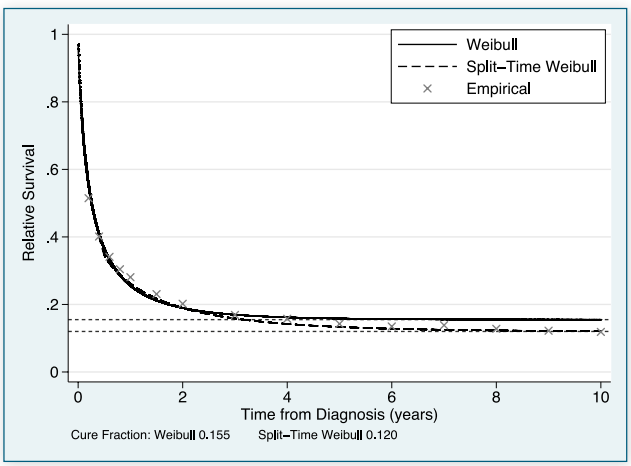


Modeling of the cure fraction in survival studies

Paul C. Lambert
Centre for Biostatistics and Genetic Epidemiology
Department of Health Sciences
University of Leicester
Leicester, UK

In certain types of survival models, some of the subjects may never experience the event of interest. For example, when studying the recurrence of a disease in previously treated patients, some of those patients may not have a second episode. In these *cure models*, the survival curve eventually reaches a plateau, and interest lies in estimating the proportion of subjects who do not experience the event, known as the cure fraction.

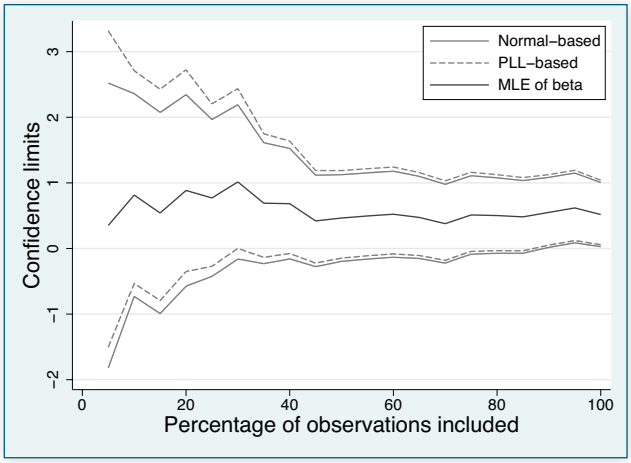
In population-based cancer studies, cure is said to occur when the mortality (hazard) rate in the diseased group of individuals returns to the same level as that expected in the general population. Patients have an obvious interest in the cure fraction, and it can be used to monitor trends and differences in survival of curable disease. In this article, Lambert describes the `strsmix` and `strsnmix` commands, which fit the two main types of cure fraction model, namely, the mixture and nonmixture cure fraction models. These models allow incorporation of the expected background mortality rate and thus enable the modeling of relative survival when cure is a possibility. The article includes an example to illustrate the commands.



Profile likelihood for estimation and confidence intervals

Patrick Royston
Cancer and Statistical Methodology Groups
MRC Clinical Trials Unit
London, UK

The standard method of confidence-interval construction is based on the asymptotic normality of the maximum likelihood estimate of a parameter. However, these confidence intervals are inaccurate when the sampling distribution of the estimate is nonnormal. The technique known as profile likelihood can produce confidence intervals with better coverage. This technique derives confidence intervals based on the asymptotic chi-squared distribution of the likelihood-ratio test statistic and may be used when the model includes only the variable of interest or several other variables in addition. Profile-likelihood confidence intervals are particularly useful in nonlinear models. The command `pllf` computes and plots the maximum likelihood estimate and profile likelihood-based confidence interval for one parameter in a wide variety of regression models.



Fitting mixed logit models by using maximum simulated likelihood

Arne Risa Hole
National Primary Care Research and Development Centre
Centre for Health Economics
University of York
York, UK

The mixed or random parameters logit model is a variant of the standard multinomial logit model that incorporates individual-level heterogeneity by treating some or all of the regression parameters of the model as random coefficients. This article describes the `mixlogit` Stata command for fitting mixed logit models by the method of maximum simulated likelihood. The command allows for variables to have either normally or lognormally distributed coefficients, and it allows the random coefficients to be correlated. Postestimation commands allow you to obtain predicted probabilities and the elements of the coefficient covariance matrix when the coefficients are specified to be correlated.

An exact and a Monte Carlo proposal to the Fisher–Pitman permutation tests for paired replicates and for independent samples

Johannes Kaiser
Laboratory for Experimental Economics
University of Bonn
Bonn, Germany

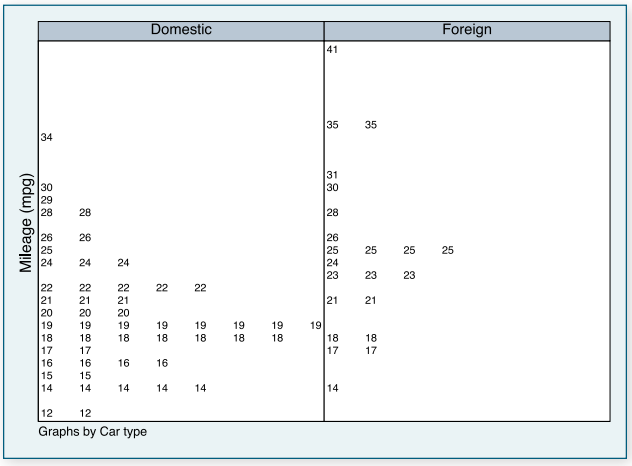
Fisher–Pitman permutation tests for paired replicates allow one to test the difference in means of two outcomes or the difference in means for two independent groups. After outlining the theory of exact tests, Kaiser derives Monte Carlo simulations for both of these Fisher–Pitman tests. Simulations can be useful if one deals with many observations because of the complexity of the algorithms used to compute the exact tests with respect to sample sizes. The tests are designed to be more powerful alternatives to the Wilcoxon signed-rank test and the Wilcoxon–Mann–Whitney rank-sum test if the observations are given on at least an interval scale. The results gained by Monte Carlo versions of the tests are sufficiently accurate in comparison with the exact versions.

Speaking Stata: Turning over a new leaf

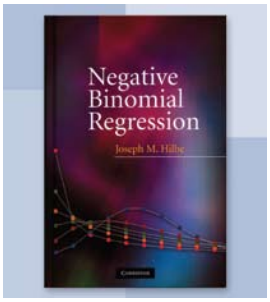
Nicholas J. Cox
Department of Geography
Durham University
Durham City, UK

Nicholas Cox's *Speaking Stata* columns help users become fluent in Stata and emphasize how much users can accomplish with just a few basic commands. In this installment Cox shows how to use Stata's scatterplot facilities to produce stem-and-leaf displays. These displays have been widely taught since John W. Tukey publicized them energetically in the 1970s. They remain useful for many distributions of small or modest size, especially for showing fine structure such as digit preference.

Although Stata has long had the command `stem` for displaying stem-and-leaf plots, that command does not utilize Stata's modern graphics engine. In this article, Cox shows how to re-create stem-and-leaf displays by using scatterplots of stem variable versus position on line with leaves shown as marker labels. Using the `by()` option with the `scatter` plottype, one can easily compare stem-and-leaf displays for different groups, and back-to-back presentation of paired displays is also possible. Cox also discusses variants on standard stem-and-leaf displays in which each distinct value is a stem, each distinct value is its own leaf, or axes are swapped. The command `stemplot` embodies the lessons of this article to automate the production of stem-and-leaf displays.



From the Stata Bookstore



Title: *Negative Binomial Regression*
Author: Joseph M. Hilbe
Publisher: Cambridge University Press
Copyright: 2007
ISBN-10: 0-521-85772-4
ISBN-13: 978-0-521-85772-7
Pages: 251; hardcover
Price: $72.00

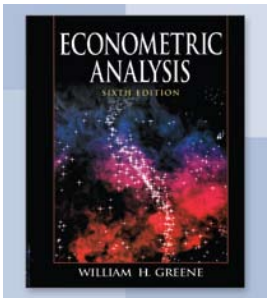
Negative Binomial Regression, by Joseph M. Hilbe, offers an exclusive review of the negative binomial model and its variations. Negative binomial regression has, in recent years, gained popularity as an alternative to Poisson regression. It is used to account for overdispersion, often encountered in many real-world applications with count responses.

The book covers a catalog of the count-response models, their estimation methods, and the algorithms used to fit these models. The author discusses the problem of overdispersion and ways to handle it in detail. The emphasis is on the application of the negative binomial models to various research problems involving overdispersed count data. A fair amount of the material is devoted to the discussion of model-selection techniques, the interpretation of the results, regression diagnostics, and methods of assessing goodness-of-fit.

The author also describes various extensions of the negative binomial model that handle excess zeros, censored and truncated data, panel/longitudinal data, and data arising from sample selection. Stata is used extensively throughout.

The book is aimed at the statisticians, econometricians, and practicing researchers analyzing count-response data. It is written to be comprehensible to a reader with a general background in maximum likelihood estimation and generalized linear models. At the same time, the text has enough mathematical details to also satisfy the more theoretically minded reader.

You can find the table of contents and online ordering information at **www.stata.com/bookstore/nbr.html**. You can also order using the enclosed bookstore order form.



Title: *Econometric Analysis, 6th Edition*
Author: William H. Greene
Publisher: Prentice Hall
Copyright: 2008
ISBN-10: 0-13-513245-2
ISBN-13: 978-0-13-513245-6
Pages: 1,178; hardcover
Price: $149.00

William Greene's *Econometric Analysis* has served as the standard reference for econometrics among economists, political scientists, and other social scientists for nearly two decades, and the newly released sixth edition is certain to carry on that tradition. The book's abundance of examples and Greene's emphasis on how to put econometric theory to practical use make the book valuable not just to graduate students taking their first course in econometrics but also to students and professionals who engage in empirical research.

As with most econometrics texts, Greene's *Econometric Analysis* begins by introducing the linear regression model. Part I of the book, consisting of seven chapters, covers the properties of the least-squares estimator, inference and prediction, and tests for functional form and specification. Part II of the book generalizes the linear regression model to allow for heteroskedasticity; then, with the generalized least-squares estimator already discussed in the context of nonspherical disturbances, presents the fixed- and random-effects panel-data models as straightforward extensions of least-squares analysis. A chapter also discusses systems of regression equations, another application of GLS. Stata's `nl` and `nlsur` commands make fitting nonlinear models as easy as fitting linear models, and it is refreshing to see nonlinear models introduced relatively early in *Econometric Analysis*. Part III discusses instrumental variables and simultaneous equations. The chapter on instrumental variables has been updated to include highlights of the recent research on weak instruments, and it includes a lucid discussion of dynamic panel-data models.

Part I lays out an estimation principle (least-squares analysis), and parts II and III show how that methodology can be applied to a wide variety of applications. Having a firm grasp of a general estimation framework makes learning new applications much easier, and this style of teaching continues through the rest of the book.

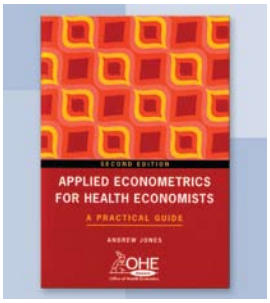
Aside from least squares, maximum likelihood, generalized methods of moments, and, increasingly, simulation-based and Bayesian estimation frameworks represent the four most commonly used techniques in econometrics; and in part IV, a chapter is devoted to each of these. Each chapter strikes a good balance between theoretical rigor and applications illustrating their use. Many newer discrete-choice models require evaluation of multivariate normal probabilities; thus, chapter 17 includes a detailed discussion of the GHK simulator.

Part V is devoted to time-series techniques, including estimation in the presence of serial correlation, models with lagged variables including vector autoregressions, stochastic processes and ARIMA models, and

nonstationarity, unit roots, and cointegration. The chapters in part V frequently make use of the results obtained in part IV on estimation frameworks.

Econometric Analysis has long been recognized for its extensive coverage of limited dependent variable models, and part VI of the sixth edition continues in that tradition. Binomial, multinomial, and ordered outcomes for both cross-sectional and panel data are covered in one chapter. Two chapters are devoted to truncation, censoring, and sample selection, as well as count and duration models.

You can find the table of contents and online ordering information at **www.stata.com/bookstore/ea.html**. You can also order using the enclosed bookstore order form.

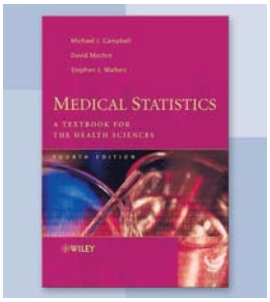


Title: *Applied Econometrics for Health Economists, 2nd Edition*
Author: Andrew Jones
Publisher: Radcliffe Publishing
Copyright: 2007
ISBN-10: 1-84619-171-8
ISBN-13: 978-1-84619-171-8
Pages: 120; paperback
Price: $37.75

Applied Econometrics for Health Economists: A Practical Guide, Second Edition, by Andrew Jones, intuitively discusses the major methods used in applied health economics. In each discussion, Jones uses Stata to analyze real data and interprets the results. Because the discussions are concise, are intuitive, and relegate all equations to an appendix, this book is an excellent choice for students or researchers who need quick, but surprisingly complete, introductions to or reviews of the main health-econometric methods for individual-level data.

As is clear from the table of contents, the book's coverage is thorough. Each chapter is short and quickly covers an important topic for applied health economics. The combination of these first-rate short discussions and the illustrative Stata examples ensure that both students and researchers will widely use this book.

You can find the table of contents and online ordering information at **www.stata.com/bookstore/aefhe.html**. You can also order using the enclosed bookstore order form.



Title: *Medical Statistics: A Textbook for the Health Sciences, 4th Edition*
Authors: Michael J. Campbell, David Machin, and Stephen J. Walters
Publisher: Wiley
Copyright: 2007
ISBN-10: 0-470-02519-0
ISBN-13: 978-0-470-02519-2
Pages: 344; paperback
Price: $32.75

Medical Statistics: A Textbook for the Health Sciences, Fourth Edition, by Michael J. Campbell, David Machin, and Stephen J. Walters, is one of many suitable books for a one- to two-semester statistics course for health professionals. However, what sets it apart is that it is both complete and easy to read. The standard tools relevant to health professions are present: odds ratios, survival analysis, observational studies, etc. Another helpful feature in this text is a section at the end of each chapter titled *Points when reading the literature*, which lists items the reader should consider when encountering the use of the discussed methods in research publications.

You can find the table of contents and online ordering information at **www.stata.com/bookstore/medstat.html**. You can also order using the enclosed bookstore order form.

Upcoming NetCourse™ schedule

Stata NetCourses are a convenient, inexpensive, web-based way to learn how to exploit the full power of Stata. Courses are divided into a series of lectures that are posted to the web site each Friday, and participants can post questions and comments to a discussion board. We also offer NetCourseNow, a way for you to choose the time and pace of a NetCourse.

Enroll by visiting **www.stata.com/netcourse** or use the enclosed order form.

NC101: Introduction to Stata

Content:	An introduction to using Stata interactively
Prerequisites:	Stata 10
Course leaders:	Theresa Boswell, Kevin Crow, and Kerry Kammire
Course length:	6 weeks (4 lectures)
Dates:	January 18–February 29, 2008
Enrollment deadline:	January 17, 2008
Next dates:	March 7–April 18, 2008
Enrollment deadline:	March 6, 2008
Price:	$95
Course syllabus:	**www.stata.com/netcourse/nc101.html**

NC151: Introduction to Stata Programming

Content:	An introduction to Stata programming dealing with what most statistical software users mean by programming, namely, the careful performance of reproducible analyses
Prerequisites:	Stata 10; basic knowledge of using Stata interactively
Course leaders:	Theresa Boswell, Kevin Crow, and Kerry Kammire
Course length:	6 weeks (4 lectures)
Dates:	January 18–February 29, 2008
Enrollment deadline:	January 17, 2008
Next dates:	March 7–April 18, 2008
Enrollment deadline:	March 6, 2008
Price:	$125
Course syllabus:	**www.stata.com/netcourse/nc151.html**

NC152: Advanced Stata Programming

Content: This course teaches you how to create and debug new commands that are indistinguishable from those of official Stata. The course assumes that you know why and when to program and, to some extent, how. You will learn how to parse both standard and nonstandard Stata syntax by using the intuitive `syntax` command, how to manage and process saved results, how to process by-groups, and more.

Prerequisites: Stata 10; course content of NetCourse 151 or equivalent knowledge.

Course leaders: Theresa Boswell, Kevin Crow, and Kerry Kammire

Course length: 7 weeks (5 lectures)

Dates: January 18–March 7, 2008

Enrollment deadline: January 17, 2008

Price: $150

Course syllabus: **www.stata.com/netcourse/nc152.html**

NC461: Introduction to Univariate Time Series with Stata

Content: This course introduces univariate time-series analysis that emphasizes the practical aspects most needed by practitioners and applied researchers. The course is written to appeal to a broad array of users, including economists, forecasters, financial analysts, managers, and anyone who encounters time-series data.

Prerequisites: Stata 10; course content of NetCourse 101 or equivalent knowledge. Familiarity with basic cross-sectional summary statistics and linear regression.

Course leaders: Brian Poi and Gustavo Sanchez

Course length: 7 weeks (4 lectures plus overview of multivariate methods)

Dates: January 18–March 7, 2008

Enrollment deadline: January 17, 2008

Price: $295

Course syllabus: **www.stata.com/netcourse/nc461.html**

International Distributors

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

Columbia CP, Ltd.
China, Hong Kong, Indonesia, Malaysia, Philippines, Singapore, Thailand, Vietnam
tel: +86-131-47501748 China
+852-28580080 Hong Kong
+62-21-633 3456 ext. 401 Indonesia
+603-56389268 Malaysia
+63-2-484-0021 Philippines
+65-68425538 Singapore
+66-8-17209515 Thailand
+84-4-7870931 Vietnam
email: sales@columbiacp.com

Dittrich & Partner Consulting
Czech Republic, Germany, Hungary
tel: +49 2 12 / 26 066-0
email: sales@dpc.de

DPC Geschäftsstelle Wien
Austria
tel: +59-212-2606624
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, Malaysia, Singapore, Thailand, Vietnam
tel: +65 62199062 Singapore,
Malaysia, Vietnam, Indonesia
662 6279000 Thailand
email: stata@eastasiatc.com.sg

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

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

Scientific Formosa, Inc.
Taiwan
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
+ 57-(1)-619-4000 Colombia
+ 58-(212)-771-5899 Venezuela
+ 51-(1)-705-9789 Peru
+ 56-(2)-595-2954 Chile
email: Ventas@SOFTWARE-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)208 697 3377
email: info@timberlake.co.uk

Timberlake Consultants Polska
Poland
tel: +48 (0)22-8252927
email: info@timberlake.pl

Timberlake Consulting S.L.
Spain
tel: +34 955601430
email: timberlake@timberlakeconsulting.com

Timberlake Consultores Brasil
Brazil
tel: +55-11-3040-3042
email: info@timberlake.com.br

Timberlake Consultores, Lda.
Portugal
tel: +351 245 308 549
email: info@timberlake.pt

TStat S.r.l.
Italy
tel: +39 0864 210101
email: tstat@tstat.it

Vishvas Marketing-Mix Services
India
tel: 91-22-2589 26 39
email: vishvas@vsnl.com

International Resellers

Axoft
Russia, Armenia, Azerbaijan, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Moldova, Tajikistan, Turkmenistan, Ukraine, Uzbekistan
tel: +7-495 232-5215
email: info@axoft.ru

BockyTech, Inc.
Taiwan
tel: +886-2-23618050
email: raymond@bockytech.com.tw

Informatique, Inc.
Japan
tel: +81-3-3505-1250
email: sales@informatiq.co.jp

Katalogo Software
Brazil
tel: +55-11-3405 4507
email: caio@katalogo.com.br

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

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

Tashtit Scientific Consultants, Ltd.
Israel
tel: +972-3-523-0825
email: info@tashtitsoft.co.il

Token Communication, Ltd.
Romania
tel: +40 364 103256
email: statasales@token.com.ro

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

How to reach us

StataCorp
4905 Lakeway Drive
College Station, TX 77845
USA

Phone 979-696-4600
Fax 979-696-4601
Email service@stata.com
Web www.stata.com

Please include your Stata serial number with all correspondence.