Spotlight on import excel and export excel: Easing the exchange of data

My favorite feature added in Stata 12 is the ability to import and export Microsoft Excel® files. Nearly every day, I work on a project that requires me to transfer data from a spreadsheet into Stata. Previously, there were two alternatives: I could copy-and-paste the data into the Data Editor or I could export the data as a text file and then use `insheet`. The copy-and-paste method works fine for small datasets, but for larger datasets I typically went with the latter method. However, when exporting the data as a text file, I would still have to verify that the top line contained valid variable names, or else I would have to specify my own variable names within Stata. Moreover, the process was clumsy. Why do I need to export the data in an intermediate format, and why do I invariably need to fire up my text editor to inspect that text file? The Excel import and export features added in Stata 12 are a godsend to me.

Spreadsheets are nearly ubiquitous in business and government, and they allow Stata users to exchange data with those who are less fortunate. Many commercial data providers supply an Excel plug-in that allows you to retrieve their data directly into spreadsheets. Countless more providers distribute their data as Excel file downloads. Accountants and financial analysts are accustomed to

working with Excel files but typically do not have Stata on their machines, so being able to import and export Excel files is a convenience to me.

There are several reasons why I prefer to use a spreadsheet as part of my data collection process. It allows me to do an initial scan of the data using the same program that created the file to make sure it contains the data I expected. In addition, spreadsheet-format data files often contain rows and columns that include descriptions, comments, and other textual information; seeing the data in a spreadsheet helps me decide what area to import into Stata. Finally, some websites give the option of downloading data as a spreadsheet or a text file. If there are many data fields, looking at the text file can sometimes be confusing, especially if it contains missing data, so I will opt to download the spreadsheet version.

Recently, I was looking at the distribution of sales of Stata across the United States. As part of my analysis, I needed to look at economic growth in each of the 366 metropolitan statistical areas in the country. I therefore proceeded to the Bureau of Economic Analysis website, quickly found gross metropolitan product data, and proceeded to download the dataset as a spreadsheet file.

I then opened the file and noticed that the actual data were in rows 7 through 373 (including a row for all metropolitan areas). Row 6 contained column headings that included entries like "2010". Knowing that valid Stata variable names could not start with numbers, I renamed those entries to, for example, "Y2010". I then opened Stata, filled in the **import excel** dialog box as shown in the figure below, and had the data I needed to proceed.
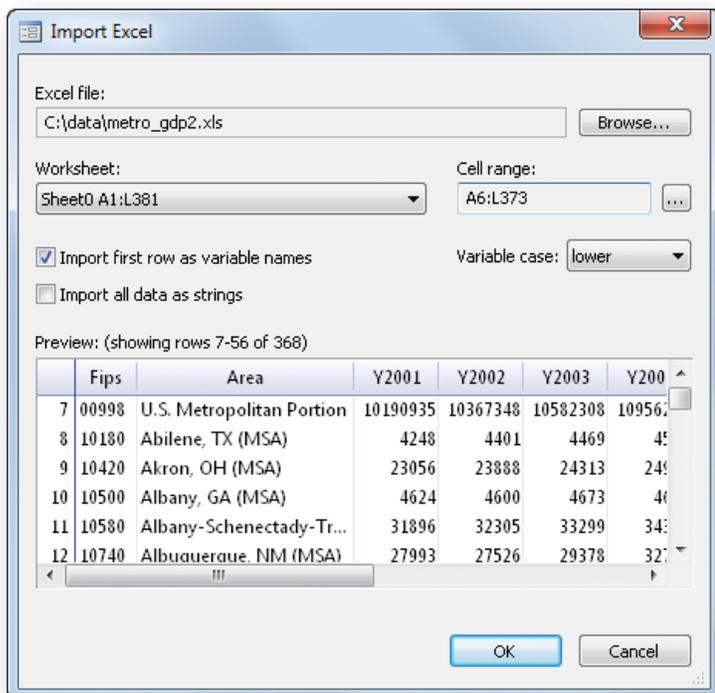
The previous example was straightforward, and I probably could have just as easily worked with a raw text file or else just copied the data from the spreadsheet and pasted it into the Data Editor. The next example would have been impossible without **import excel** to save the day.

An associate emailed me with a problem. He had downloaded financial information on 6,400 companies from the Bloomberg Professional service. The nearly 200 megabytes of raw data were stored in a set of 10 Excel spreadsheets, one for each industry he was studying. Within each spreadsheet file, each company's data was stored on an individual sheet; one spreadsheet had about 300 sheets while another had over 1,000. Moreover, each sheet had rows of headers and footers of varying sizes. One alternative would have been to write a Visual Basic script to go through each sheet and export the relevant range of data as a text file. However, I hadn't written a VB script in several years, and I would still have had to write another program in Stata to import each text file, create a valid date variable, and check for obvious errors.

The only saving grace was that my associate did have a list of the companies' ticker symbols separated by industry.

Because of Stata 12's **import excel** function, I was able to write a do-file to read in each company's data, **append** that data to a master Stata dataset, and provide error checking, all in just 50 lines of code! A do-file to read in the text files and perform the same integrity checks would likely have been longer. Going that route, I probably would have spent at least half a day brushing up on my VB skills to write a program to export the data as text as well. **import excel** proved invaluable here.

I have referred to Excel spreadsheet files, but in fact I do not even use Microsoft Excel. Many of StataCorp's internal applications are UNIX-based, so I use a Linux computer at work. I also use Linux at home so that I do not have to remember the idiosyncrasies of two operating systems. Instead of Excel, I use LibreOffice Calc, an open-source alternative, and I almost never run into problems working with Excel files in Calc. Even if you do not use Microsoft Excel, the ability to work with Excel files in Stata can still prove useful.

Most of the time, I need to import spreadsheet data into Stata to perform analyses. However, Stata also has an **export excel** command that allows you to create Excel spreadsheets based on Stata datasets.

Of course, if I were working on a detailed transaction-level dataset with millions of observations on hundreds of variables, I would not want to use a spreadsheet program at all. Use the right tool for the job. But despite the increasing interest in "big data", many common tasks involve moderately sized datasets. In those cases, **import excel** and **export excel** are valuable additions to your toolbox.

— Brian Poi
Senior Economist

New from Stata Press

Interpreting and Visualizing Regression Models Using Stata
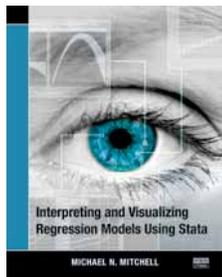
Author:	Michael N. Mitchell
Publisher:	Stata Press
Copyright:	2012
ISBN-13:	978-1-59718-107-5
Pages:	588; paperback
Price:	$58.00

Michael Mitchell's *Interpreting and Visualizing Regression Models Using Stata* is a clear treatment of how to carefully present results from model-fitting in a wide variety of settings. It is a boon to anyone who has to present the tangible meaning of a complex model in a clear fashion, regardless of the audience. As an example, many experienced researchers start to squirm when asked to give a simple explanation of the practical meaning of interactions in nonlinear models such as logistic regression. The techniques presented in Mitchell's book make answering those questions easy. The overarching theme of the book is that graphs make interpreting even the most complicated models containing interaction terms, categorical variables, and other intricacies straightforward.

Using a dataset based on the General Social Survey, Mitchell starts with basic linear regression with a single independent variable, and then illustrates how to tabulate and graph predicted values. While illustrating, Mitchell focuses on Stata's `margins` and `marginsplot` commands, which play a central role in the book and which greatly simplify the calculation and presentation of results from regression models. In particular, through use of the `marginsplot` command, Mitchell shows how you can graphically visualize every model presented in the book. Gaining insight into results is much easier when you can view them in a graph rather than in a mundane table of results.

Mitchell then proceeds to more-complicated models where the effects of the independent variables are nonlinear. After discussing how to detect nonlinear effects, he presents examples using both standard polynomial terms (squares and cubes of variables) as well as fractional polynomial models, where independent variables can be raised to powers like -1 or $1/2$. In all cases, Mitchell again uses the `marginsplot` command to illustrate the effect that changing an independent variable has on the dependent variable. Piecewise-linear models are presented as well; these are linear models in which the slope or intercept is allowed to change depending on the range of an independent variable. Mitchell also uses the `contrast` command when discussing categorical variables; as the name suggests, this command allows you to easily contrast predictions made for various levels of the categorical variable.

Interaction terms can be tricky to interpret, but Mitchell shows how graphs produced by `marginsplot` greatly clarify results. Individual chapters are devoted to two- and three-way interactions containing all continuous or all categorical variables and include many practical examples. Raw regression output including interactions of continuous and categorical variables can be nigh impossible to interpret, but again Mitchell makes this a snap through judicious use of the `margins` and `marginsplot` commands in subsequent chapters.

The first two-thirds of the book is devoted to cross-sectional data, while the final third considers longitudinal data and complex survey data. A significant difference between this book and most others on regression models is that Mitchell spends quite some time on fitting and visualizing discontinuous models—models where the outcome can change value suddenly at thresholds. Such models are natural in settings such as education and policy evaluation, where graduation or policy changes can make sudden changes in income or revenue.

This book is a worthwhile addition to the library of anyone involved in statistical consulting, teaching, or collaborative applied statistical environments. Graphs greatly aid the interpretation of regression models, and Mitchell's book shows you how.

You can find the table of contents and online ordering information at stata-press.com/books/interpreting-visualizing-regression-models.

Multilevel and Longitudinal Modeling Using Stata, Third Edition
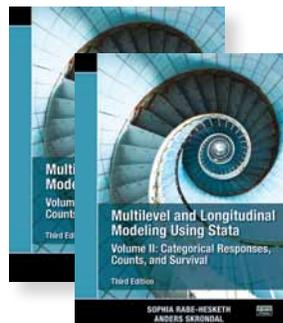
Authors:	Sophia Rabe-Hesketh and Anders Skrondal
Publisher:	Stata Press
Copyright:	2012
ISBN-13:	978-1-59718-108-2
Pages:	974; paperback
Price:	$109.00 (both volumes)

Multilevel and Longitudinal Modeling Using Stata, Third Edition, by Sophia Rabe-Hesketh and Anders Skrondal, looks specifically at Stata's treatment of generalized linear mixed models, also known as multilevel or hierarchical models. These models are "mixed" because they allow fixed and random effects, and they are "generalized" because they are appropriate for continuous Gaussian responses as well as binary, count, and other types of limited dependent variables.

The material in the third edition consists of two volumes, a result of the substantial expansion of material from the second edition, and has much to offer readers of the earlier editions.

Volume I is devoted to continuous Gaussian linear mixed models and has nine chapters organized into four parts. The first part reviews the methods of linear regression. The second part provides in-depth coverage of two-level models, the simplest extensions of a linear regression model.

Rabe-Hesketh and Skrondal begin with the comparatively simple random-intercept linear model without covariates, developing the mixed model from principles and thereby familiarizing the reader with terminology, summarizing

and relating the widely used estimating strategies, and providing historical perspective. Once the authors have established the mixed-model foundation, they smoothly generalize to random-intercept models with covariates and then to a discussion of the various estimators (between, within, and random effects). The authors then discuss models with random coefficients.

The third part of volume I describes models for longitudinal and panel data, including dynamic models, marginal models (a new chapter), and growth-curve models (a new chapter). The fourth and final part covers models with nested and crossed random effects, including a new chapter describing in more detail higher-level nested models for continuous outcomes.

The mixed-model foundation and the in-depth coverage of the mixed-model principles provided in volume I for continuous outcomes make it straightforward to transition to generalized linear mixed models for noncontinuous outcomes, which are described in volume II.

Volume II is devoted to generalized linear mixed models for binary, categorical, count, and survival outcomes. The second volume has seven chapters also organized into four parts. The first three parts in volume II cover models for categorical responses, including binary, ordinal, and nominal (a new chapter); models for count data; and models for survival data, including discrete-time and continuous-time (a new chapter) survival responses. The fourth and final part in volume II describes models with nested and crossed-random effects with an emphasis on binary outcomes.

The book has extensive applications of generalized mixed models performed in Stata. Rabe-Hesketh and Skrondal developed `gllamm`, a Stata program that can fit many latent-variable models, of which the generalized linear mixed model is a special case. As of version 10, Stata contains the `xtmixed`, `xtmelogit`, and `xtmepoisson` commands for fitting multilevel models, in addition to other `xt` commands for fitting standard random-intercept models. The types of models fit by these commands sometimes overlap; when this happens, the authors highlight the differences in syntax, data organization, and output for the two (or more) commands that can be used to fit the same model. The authors also point out the relative strengths and weaknesses of each command when used to fit the same model, based on considerations such as computational speed, accuracy, available predictions, and available postestimation statistics.

In summary, this book is the most complete, up-to-date depiction of Stata's capacity for fitting generalized linear mixed models. The authors provide an ideal introduction for Stata users wishing to learn about this powerful data analysis tool.

Multilevel and Longitudinal Modeling Using Stata, Third Edition may be purchased as a two-volume set for $109. Alternatively, either volume may be purchased individually for $62.

- *Volume I: Continuous Responses*
- *Volume II: Categorical Responses, Counts, and Survival*

You can find the table of contents and online ordering information at stata-press.com/books/multilevel-longitudinal-modeling-stata.

A Gentle Introduction to Stata, Revised Third Edition
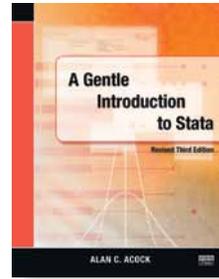
Author:	Alan C. Acock
Publisher:	Stata Press
Copyright:	2012
ISBN-13:	978-1-59718-109-9
Pages:	401; paperback
Price:	$48.00

Alan C. Acock's *A Gentle Introduction to Stata, Revised Third Edition* is aimed at new Stata users who want to become proficient in Stata. After reading this introductory text, new users not only will be able to use Stata well but also will learn new aspects of Stata easily.

Acock assumes that the user is not familiar with any statistical software. This assumption of a blank slate is central to the structure and contents of the book. Acock starts with the basics; for example, the portion of the book that deals with data management begins with a careful and detailed example of turning survey data on paper into a Stata-ready dataset on the computer. When explaining how to go about basic exploratory statistical procedures, Acock includes notes that will help the reader develop good work habits. This mixture of explaining good Stata habits and good statistical habits continues throughout the book.

Acock is quite careful to teach the reader all aspects of using Stata. He covers data management, good work habits (including the use of basic do-files), basic exploratory statistics (including graphical displays), and analyses using the standard array of basic statistical tools (correlation, linear and logistic regression, and parametric and nonparametric tests of location and dispersion). Acock teaches Stata commands by using the menus and dialog boxes while still stressing the value of do-files. In this way, he ensures that all types of users can build good work habits. Each chapter has exercises that the motivated reader can use to reinforce the material.

The tone of the book is friendly and conversational without ever being glib or condescending. Important asides and notes about terminology are set off in boxes, which makes the text easy to read without any convoluted twists or forward-referencing. Rather than splitting topics by their Stata implementation, Acock arranges the topics as they would appear in a basic statistics textbook; graphics and postestimation are woven into the material in a natural fashion. Real datasets, such as the General Social Surveys from 2002 and 2006, are used throughout the book.

The focus of the book is especially helpful for those in psychology and the social sciences, because the presentation of basic statistical modeling is supplemented with discussions of effect sizes and standardized coefficients. Various selection criteria, such as semipartial correlations, are discussed for model selection.

The revised third edition of the book has been updated to reflect the new features available in Stata 12 and Stata 11. The ANOVA chapter has been revised to incorporate the `pwmeans` command, to do mean comparisons,

and the `marginsplot` command, which simplifies the construction of graphs showing interaction effects. Menus and screenshots have also been updated. As in the third edition, an entire chapter is devoted to the analysis of missing data and the use of multiple-imputation methods. Factor-variable notation is introduced as an alternative to the manual creation of interaction terms. The new Variables Manager and revamped Data Editor are featured in the discussion of data management.

You can find the table of contents and online ordering information at stata-press.com/books/gentle-introduction-to-stata.

....

New from the Stata Bookstore

## A Short Introduction to Stata for Biostatistics (Updated to Stata 12)
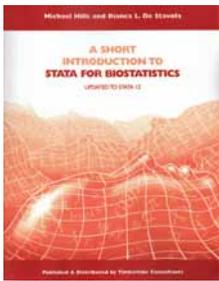
Authors:	Michael Hills and Bianca L. De Stavola
Publisher:	Timberlake Consultants
Copyright:	2012
ISBN-13:	978-0-9571708-0-3
Pages:	181; paperback
Price:	$52.00

A Short Introduction to Stata for Biostatistics bridges the information gap between Stata's *Getting Started* manual and *Reference* manuals by providing a more detailed introduction to the most often used analytic methods in biomedical research. Although the book is written specifically for biostatisticians, epidemiologists, and health professionals new to Stata, it is also useful for more-experienced users wanting more in-depth knowledge of both Stata commands and biostatistical issues. The book is hands on, intended to be used while working with Stata, and includes a CD-ROM containing the datasets and several author-written programs.

The first four chapters provide an overview of data entry and management commands, including those used to create, label, and drop variables and those used to sort observations. The next two chapters cover graphics. Then comes the bulk of the book, which details methods used in data description and analysis. Beginning with commands used to create frequency tables and summary statistics, the authors proceed to describe commands used for univariate and multivariate analyses, including linear regression, Poisson regression, logistic regression, survival data analysis (proportional hazards models and competing-risks models), and meta analysis. Included among the final chapters is a useful tutorial on report generation.

You can find the table of contents and online ordering information at stata.com/bookstore/short-intro-stata-biostatistics.

....

Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models, Second Edition
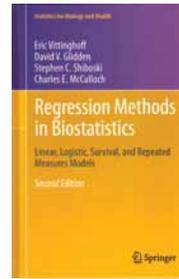
Authors:	Eric Vittinghoff, David V. Glidden, Stephen C. Shiboski, and Charles E. McCulloch
Publisher:	Springer
Copyright:	2012
ISBN-13:	978-1-4614-1352-3
Pages:	509; hardcover
Price:	$74.75

Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models, Second Edition is intended as a teaching text for a one-semester or two-quarter secondary statistics course in biostatistics. The book's focus is multipredictor regression models in modern medical research. The authors recommend as a prerequisite an introductory course in statistics or biostatistics, but the first three chapters provide sufficient review material to make this requirement not critical.

Vittinghoff, Glidden, Shiboski, and McCulloch take a unified approach to regression models. They begin with linear regression and then discuss issues such as model statement and assumptions, types of regressors (for example, categorical versus continuous), interactions, causation and confounding, inference and testing, diagnostics, and alternative models for when assumptions are violated. Then they discuss these same issues in the contexts of other multipredictor regression models, namely, logistic regression, the Cox model, and generalized linear models (GLMs). The authors then cover generalized estimating equations (GEE) and the analysis of survey data. Almost all analyses are performed using Stata.

The second edition provides two new chapters and substantially expands some of the existing chapters. Specifically, a new chapter on strengthening causal inference describes the fundamentals of causal inference and concentrates on two estimation methods—inverse probability weighting and what the authors call potential outcomes estimation. This chapter also covers propensity scores, time-dependent treatments, instrumental variables, and principal stratification. The other new chapter is on missing data. The authors describe the missing-data problem and its impact on statistical inference. They then discuss three approaches for handling missing data: maximum likelihood estimation, multiple imputation, and inverse weighting. Among the substantially revised chapters are chapters on logistic regression, now including categorical outcomes; on survival analysis, now including competing risks; on generalized linear models, now including negative binomial and zero-truncated and zero-inflated count models; and more. All the Stata examples used in the book have been updated for Stata 12.

You can find the table of contents and online ordering information at stata.com/bookstore/regression-methods-biostatistics.

....

STATA® Conference

About the Conference

Join us in sunny San Diego for the 2012 Stata Conference. The Stata Conference is enjoyable and rewarding for Stata users at all levels and from all disciplines. This year's program will include presentations by users and invited speakers, and it will also include the ever-popular "Wishes and grumbles" session. Representatives from StataCorp include Bill Gould, President and Head of Development; Chuck Huber, Senior Statistician; and Kristin MacDonald, Senior Statistician.

Accommodations

Rooms at the Manchester Grand Hyatt are available at the discounted rate of $229 per night. For reservations, call 1-888-421-1442 and identify yourself as a guest with the group Stata, or reserve online (see stata.com/sandiego12 for details). Please make your reservation by Monday, June 25, 2012, to receive the discounted rate.

Scientific committee

- A. Colin Cameron, University of California–Davis
- Xiao Chen, University of California–Los Angeles
- Phil Ender (chair), University of California–Los Angeles
- Estie Hudes, University of California–San Francisco
- Michael Mitchell, U.S. Department of Veterans Affairs

Logistics organizer

Sarah Marrs, StataCorp LP
smarrs@stata.com

Dates	July 26–27
Venue	Manchester Grand Hyatt One Market Place San Diego, CA 92101
Cost	$195 regular; $75 student
Register	**stata.com/sandiego12**

Program

Thursday, July 26

Custom Stata commands for semi-automatic confidentiality screening of Statistics Canada data
Jesse McCrosky, University of Saskatchewan

scdensity: A program for self-consistent density estimation
Joerg Luedicke, University of Florida and Yale University

TMPM: The trauma mortality prediction model is robust to ICD-9, ICD-10, and AIS Coding lexicons
Alan Cook, Baylor University Medical Center

Adoption: A new Stata routine for consistently estimating population technological adoption parameters
Aliou Diagne, African Rice Center

Graphics (and numerics) for univariate distributions
Nicholas J. Cox, Durham University

Binary choice models with endogenous regressors
Christopher Baum, Boston College and DIW Berlin
Yingying Dong, University of California–Irvine
Arthur Lewbel, Boston College

An application of multiple imputation and sampling-based estimation
Haluk Gedikoglu, Lincoln University of Missouri

The application of Stata's multiple-imputation techniques to analyze a design of experiments with multiple responses
Clara Novoa, Texas State University

EFA within a CFA context
Phil Ender, Unversity of California–Los Angeles

Structural equation modeling using the SEM builder and the sem command
Kristin MacDonald, StataCorp LP

Imagining a Stata/Python combination
James Fiedler, Universities Space Research Association

Friday, July 27

Issues for analyzing competing-risks data with missing or misclassification in causes
Ronny Westerman, Philipps-University of Marburg

Generating survival data for fitting marginal structural Cox models using Stata
Mohammed Ehsanul Karim, University of British Columbia

Computing optimal strata bounds using dynamic programming
Eric Miller, Summit Consulting

Correct standard errors for multistage regression-based estimators: A guide for practitioners with illustrations
Joseph Terza, University of North Carolina–Greensboro

Shrinkage estimators for structural parameters
Tirthankar Chakravarty, University of California–San Diego

Stata implementation of the nonparametric spatial heteroskedasticity- and autocorrelation-consistent covariance matrix estimator
P. Wilner Jeanty, Rice University

Big data, little spaces, high speed: Using Stata to analyze the determinants of broadband access in the United States
David Beede, U.S. Department of Commerce
Brittany Bond, U.S. Department of Commerce

A comparative analysis of lottery-, charter-, and traditional-based elementary schools within the Anchorage school district
Matthew McCauley, University of Alaska–Anchorage

Matching individuals in the Current Population Survey: A distance-based approach
Stuart Craig, Yale University

Dialog programming for automating the African Transformation Index (ATI): Challenges in using Stata
Kwaku Damoah, African Center for Economic Transformation

Allocative efficiency analysis using DEA in Stata
Choonjoo Lee, Korea National Defense University

Psychometric analysis using Stata
Chuck Huber, StataCorp LP

Report to users
Bill Gould, StataCorp LP

Wishes and grumbles: User feedback and Q&A

Register today!
stata.com/sandiego12

2012 Stata Users Group meetings

Germany: June 1

The 10th German Stata Users Group meeting will be held at the WZB Social Science Research Center in Berlin on Friday, June 1, 2012. Patrick Royston, University College–London; Willi Sauerbrei, University of Freiburg; and Maurizio Pisati, University of Milano–Bicocca will present this year's keynote speeches. Bill Rising, Director of Educational Services, and Bill Gould, President and Head of Development, from StataCorp will also attend.

Program

Handling interactions in Stata, especially with continuous predictors
> Patrick Royston, University College London
> Willi Sauerbrei, University of Freiburg

Exploratory spatial data analysis using Stata
> Maurizio Pisati, University of Milano–Bicocca

leebounds: Lee's treatment effect bounds for samples with nonrandom sample selection
> Harald Tauchmann, Rheinisch-Westfälisches Institut für Wirtschaftsforschung

Comparing observed and theoretical distributions
> Maarten Buis, University of Tübingen

A simple alternative to the linear probability model for binary choice models with endogenous regressors
> Christopher F. Baum, Boston College and DIW Berlin
> Yingying Dong, University of California–Irvine
> Arthur Lewbel, Boston College
> Tao Yang, Boston College

Robust regression in Stata
> Ben Jann, University of Bern

Working in the margins to plot a clear course
> Bill Rising, StataCorp LP

Can multilevel multiprocess models be estimated using Stata? A case for the cmp command
> Tamás Bartus, Corvinus University of Budapest

Rescaling results of mixed nonlinear probability models to compare regression coefficients or variance components across hierarchically nested models
> Dirk Enzmann, University of Hamburg
> Ulrich Kohler, Social Science Research Center Berlin

Multilevel tools
> Katja Möhring, University of Cologne
> Alexander Schmidt, University of Cologne

Modular programming in Stata
> Daniel Schneider, University of Frankfurt/Main

Report to users
> Bill Gould, StataCorp

Wishes and grumbles

Venue:	WZB Social Science Research Center
	Reichpietschufer 50
	D-10785 Berlin
Cost:	Meeting only: €45 regular; €25 student
	Workshop only: €65; Both meeting and workshop: €85
Details:	stata.com/meeting/germany12

Registration

Participants are asked to travel at their own expense. The conference fee covers costs for coffee, tea, and lunch. There will also be an optional informal meal at a restaurant in Berlin on Friday evening at additional cost.

You can register by emailing Anke Mrosek (anke.mrosek@dpc.de) or by writing, phoning, or faxing to

Anke Mrosek
Dittrich & Partner Consulting GmbH
Prinzenstr. 2
42697 Solingen

Tel: +49 (0) 212 260 66-24
Fax: +49 (0) 212 260 66-66

Portugal: September 7

Save the date! This year's meeting will be held at the Nova School of Business and Economics in Lisbon, Portugal. For upcoming details, go to stata.com/meeting/portugal12.

SPAIN

UK: September 13–14

Venue:	Cass Business School
	Bunhill Row
	London EC1Y 8TZ
Cost:	Both days: £96 regular; £66 student
	Single day: £66 regular; £48 student
	Dinner (optional): £36
Details:	stata.com/meeting/uk12

Bill Gould, President and Head of Development, and Bill Rising, Director of Educational Services, from StataCorp will attend.

Registration

To register for this meeting, contact Timberlake Consultants:

info@timberlake.co.uk
timberlake.co.uk
+44 (0) 20 8697 3377

Spain: September 12

Venue:	Universitat de Barcelona (UB)
	Facultat d'Economia i Empresa
	Avda. Diagonal, 690. Barcelona 08034
Cost:	€60 regular; €30 student
Details:	stata.com/meeting/spain12

Isabel Cañette and Gustavo Sanchez, Senior Statisticians from StataCorp, will attend. The meeting will include the popular "Wishes and grumbles" session.

Call for presentations

The conference will be conducted mostly in English. Papers will be accepted in both Spanish and English, with English being the preferred language. Presentations are welcome on any Stata-related topic, including user-written Stata programs, case studies of research or teaching using Stata, discussions of data-management problems, and surveys or critiques of Stata facilities in specific fields. The submission deadline is June 22. For submission guidelines, see stata.com/meeting/spain12.

Registration

To register, please send an email to Timberlake Consulting S.L. at info@timberlakeconsulting.com; they will email you the registration form that you will need to fill out and return before September 7.

Italy: September 20–21

Venue:	Grand Hotel Majestic "Giá Baglioni"
	Via Indipendenza, 8, 40121 Bologna
Cost:	Meeting only: €95 regular; €71 student
	Meeting and training: €400 regular; €300 student
Details:	stata.com/meeting/italy12

The first day of the meeting will comprise five sessions; the second day of the meeting will comprise two training courses (one in Italian and one in English). Bill Rising, StataCorp's Director of Educational Services, will attend.

Call for presentations

As in previous years, the emphasis will be on the development of new commands or procedures currently unavailable in Stata. Also encouraged are proposals based on the use of Stata in previously unpublished empirical research and other general-interest applications of Stata, such as data management or teaching with Stata. The submission deadline is June 30. For submission guidelines, see stata.com/meeting/italy12.

Registration

Submit your completed registration form, found at stata.com/meeting/italy12, to TStat S.r.l. by September 10.

TStat S.r.l.
statausers@tstat.it
Tel: +39 0864 210101
Fax: +39 0864 206014

ITALY

Public training schedule

Course	Dates	Location	Cost
Multilevel/Mixed Models Using Stata	August 30–31, 2012 October 4–5, 2012	Washington, DC Washington, DC	$1,295
Survey Data Analysis Using Stata	May 30–31, 2012	Washington, DC	$1,295
Using Stata Effectively: Data Management, Analysis, and Graphics Fundamentals	June 19–20, 2012 August 28–29, 2012 October 2–3, 2012 November 1–2, 2012	Chicago, IL Washington, DC Washington, DC San Francisco, CA	$950

Multilevel/Mixed Models Using Stata

This two-day course taught by Bill Rising, StataCorp's Director of Educational Services, is an introduction to using Stata to fit multilevel/mixed models. Mixed models contain both fixed effects analogous to the coefficients in standard regression models and random effects not directly estimated, but instead summarized through the unique elements of their variance–covariance matrix. Mixed models may contain more than one level of nested random effects, and hence these models are also referred to as multilevel or hierarchical models, particularly in the social sciences. Stata's approach to linear mixed models is to assign random effects to independent panels where a hierarchy of nested panels can be defined for handling nested random effects. Exercises will supplement the lectures and Stata examples.

The course will cover the following topics:

- Random-intercept linear model
- Random coefficients and the various covariance structures that can be imposed with multiple random-effects terms
- Methods for fitting more complex models, including crossed-effects models, growth curve models, and models with complex and grouped constraints on covariance structures
- Predictions, model diagnostics, and other postestimation tasks
- Binary and count responses

Survey Data Analysis Using Stata

This course taught by Stas Kolenikov, Senior Survey Statistician at Abt SRBI and Adjunct Assistant Professor at University of Missouri–Columbia, covers how to use Stata for survey data analysis assuming a fixed population. We will begin by reviewing the sampling methods used to collect survey data, and how they act in the estimation of totals, ratios, and regression coefficients. We will then cover the three variance estimators implemented in Stata's survey estimation commands. Strata with a single sampling unit, certainty sampling units, subpopulation estimation, and poststratification will also be covered in some detail.

The course will cover the following topics:

- Sampling
- Sampling design characteristics
  - › Cluster sampling
  - › Stratified sampling
  - › Sampling without replacement
- Regression with survey data
- Variance estimation
  - › Linearization
  - › Balanced repeated replication (BRR)
  - › Jackknife
- Special types of sampling units
  - › Strata with a single sampling unit
  - › Certainty units
- Restricted sample and subpopulation estimation
- Poststratification

Using Stata Effectively

Become intimately familiar with all three components of Stata: data management, analysis, and graphics. This two-day course, taught by Bill Rising (StataCorp's Director of Educational Services), is aimed at both new Stata users and those who want to use Stata more effectively. You will learn to use Stata efficiently and to make your work reproducible and self-explanatory. As a result, collaborative changes and follow-up analyses will become much simpler.

The course will cover the following topics:

- Stata basics, including overviews of the following elements and how to use them together: Stata's GUI, command line, and scripting
- Data management
- Workflow
- Analysis, including basic statistical commands and common postestimation commands for predictions, hypothesis tests, marginal effects, and more
- Graphics, including the Graph Editor

For course details or to enroll, visit stata.com/public-training or contact us at training@stata.com. Want to know when a new course is announced? Sign up for an email alert at stata.com/alerts.

Enroll today at stata.com/public-training.

NetCourse schedule

NetCourse 101, Introduction to Stata

An introductory 6-week course that teaches how to use Stata interactively.

       July 6–August 17, 2012 $95.00

NetCourse 151, Introduction to Stata Programming

An introductory 6-week course that teaches Stata data-analysis programming to those who have a basic knowledge of using Stata interactively.

       July 6–August 17, 2012 $125.00

NetCourse 152, Advanced Stata Programming

An advanced 7-week course that teaches how to add new commands

       October 12–November 30, 2012 $150.00

NetCourse 461, Introduction to Univariate Time Series with Stata

An introductory 7-week course that teaches the practical aspects of time series that are most needed by practitioners and applied researchers.

       October 12–November 30, 2012 $295.00

NetCourseNow

Would you prefer to choose the time and set the pace of a NetCourse? Want to have a personal NetCourse instructor? This the course for you.

For details on all NetCourses, visit stata.com/netcourse.

What our users love about Stata

We recently had a contest on our Facebook page. To enter, contestants posted their favorite Stata command, feature, or just a comment telling us why they love Stata. Contestants then asked their friends, colleagues, and fellow Stata users to vote for their entry by 'Like'-ing the post. The grand prize, a copy of Stata/MP 12 (8-core). The response was overwhelming! Congratulations to Rodrigo Briceno, whose entry garnered an amazing 2,235 Likes:

> One of the most remarkable experiences with Stata was when I learned to use loops. Making repetitive procedures in so short amounts of time is really amazing! I LIKE STATA!

Second place, with 1,464 Likes, went to Juan Jose Salcedo:

> My Favorite STATA command is by far COLLAPSE! Getting descriptive statistics couldn't be any easier!

Third place, with 140 Likes, went to Tymon Sloczynski

> My favourite command is '`oaxaca`', a user-written command (by Ben Jann from Zurich) which can be used to carry out the so-called Oaxaca–Blinder decomposition. I often use it in my research and it saves a lot of time—which easily makes it favourite!

Read all the entries at blog.stata.com/our-users-favorite-commands.

Upcoming events

APS 2012

Chicago, IL, May 24–27, 2012

The Association for Psychological Science (APS) will have their annual meeting in Chicago, IL, from May 24–27. For more information, go to psychologicalscience.org/convention.

Stata representatives, including Kristin MacDonald, Senior Statistician, and Vince Wiggins, Vice President of Scientific Development, will be available to answer your questions about all things Stata. (Booth #401)

JSM 2012

San Diego, CA, July 28–August 2, 2012

The 2012 Joint Statistical Meetings (JSM) will be held at the San Diego Convention Center from July 28–August 2. For more information, go to amstat.org/meetings/jsm/2012.

Jeff Pitblado, Director of Statistical Software, and Bill Rising, Director of Educational Services, among other representatives from StataCorp, will be on hand to answer any questions you have about Stata. (Booth #201)

APA 2012

Orlando, FL, August 2–5, 2012

The 2012 American Psychological Association (APA) annual convention will take place in Orlando, FL, from August 2–5. For more information, go to www.apa.org/convention.

Among the attendees from StataCorp will be Kristin MacDonald, Senior Statistician, and Vince Wiggins, Vice President of Scientific Development. They look forward to meeting with you and answering any questions you have about Stata. (Booth #1014)

APSA 2012

New Orleans, LA, August 30–September 2, 2012

The American Political Science Association will conduct its annual meeting and exhibition in New Orleans, LA, from August 30–September 2. For more information, go to apsanet.org/content_77049.cfm.

Brian Poi, Senior Economist from StataCorp will be available to demonstrate Stata's features. (Booth #139)

Stop by our booths for a demonstration of the new features in Stata 12, including SEM, and to get 20% off your next purchase of Stata Press books and *Stata Journal* subscriptions. We'll also hold a drawing for a copy of Stata/MP (4-core) at each of these conferences. Don't miss your opportunity to win.

We look forward to seeing you there!

STATA®

StataCorp
4905 Lakeway Drive
College Station, TX 77845-4512
USA

Return service requested.

Contact us

979-696-4600   979-696-4601 (fax)

service@stata.com   stata.com

Please include your Stata serial number with all correspondence.

Find a Stata distributor near you
stata.com/worldwide

f **facebook.com/StataCorp**

t **twitter.com/Stata**

b **blog.stata.com**

STATA® Conference

Dates	July 26–27
Venue	Manchester Grand Hyatt One Market Place San Diego, CA 92101
Cost	$195 regular; $75 student
Register	**stata.com/sandiego12**

Join us in sunny San Diego for the 2012 Stata Conference. The Stata Conference is enjoyable and rewarding for Stata users at all levels and from all disciplines. This year's program will include presentations by users and invited speakers, and it will also include the ever-popular "Wishes and grumbles" session. Representatives from StataCorp include Bill Gould, President and Head of Development; Chuck Huber, Senior Statistician; and Kristin MacDonald, Senior Statistician.

San Diego 2012

See inside for the complete program!

stata.com/sandiego12