

# THE STATA NEWS

Volume 18 Number 3

July/August/September 2003

www.stata.com

## FDA file format now supported

Stata 8 can now read and write files in the format required for submissions to the U.S. Food and Drug Administration (FDA)—SAS XPORT format. Two new commands provide this ability: `fdasave` and `fdause`. The primary intent of these commands is to assist people in making submissions to the FDA, but the commands are general enough to use in transferring data between SAS and Stata.

These new commands are included in the latest free update to Stata 8, which you obtain by launching Stata, typing `update`, and following the instructions. You can access online documentation after updating, by typing `whelp fdasave`.

`fdasave` saves the data in the FDA's official format for submitting and archiving new drug and device applications (NDAs). To format your Stata dataset for submission, type

```
. fdasave filename
```

Stata creates an FDA format file `filename.xpt` containing the data. If the data includes value labels, Stata automatically creates an additional FDA format file, `formats.xpf`, containing the value-label definitions. These files are in exactly the format expected by the FDA and can easily be read into Stata or into SAS Statistical Software (SAS Institute Inc., Cary, North Carolina).

`fdause` reads FDA format files into Stata. You type

```
. fdause filename
```

Stata reads the FDA format file `filename.xpt`, and if the file `formats.xpf` also exists, Stata also reads the value-label definitions.

For more information about FDA NDA submissions, see <http://www.fda.gov/cder/guidance/2867fnl.pdf>.

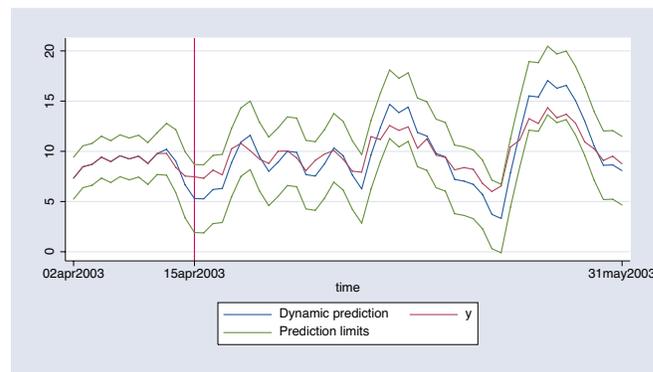
## Time-series graphs in Stata 8

Stata 8 has new features for the graphical display of time-series data. Stata's graphics engine has an enhanced time axis and associated options that allow you to specify date strings instead of numeric values. The latest update also contains new time-series plottypes, `tsline` and `tsrline`, and a graph command for plotting panel data, `xtline`.

When you `tsset` your data and identify the scale of the time variable (e.g., `daily`, `weekly`, `monthly`, ...), Stata creates a time axis when plotting variables over time. You can now specify date strings when adding tick marks (both major and minor, labeled and unlabeled), vertical lines, and text boxes.

`tsline` and `tsrline` automatically identify the time variable set by `tsset`. `tsline` generates line plots of one or more time series, and `tsrline` generates range plots with lines. These two plottypes can be combined; for example, you can plot a forecast with error bounds.

```
. tsset time, daily
. tsline ypred y || tsrline ll ul ||
  in -60/1 , ytitle("")
  tlabel(minmax 15apr2003) tline(15apr2003)
```



`xtline` automatically identifies the time and panel variables set by `tsset`. By default, `xtline` plots several variables versus time for each panel separately. Optionally, `xtline` can also produce a single graph with line plots overlaid by panel.

You can obtain online documentation after updating by typing `whelp tsline`, `whelp xtline`, and `whelp twoway_options`.

## Multiple-language dataset support

Just added to Stata 8 is label language, which allows you to create datasets that contain data, variable, and value labels in different languages. A dataset might contain one set of labels in English, another in German, and a third in Spanish. Or, a dataset might contain labels all in the same language, one set of long labels and another set of shorter ones. A dataset may contain up to 100 sets of labels.

This new capability is available for free to all Stata 8 users—just launch Stata, type `update`, and follow the instructions.

### Inside this issue:

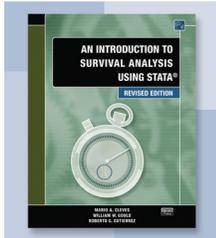
FDA file format now supported	1
Time-series graphs in Stata 8	1
Multiple-language dataset support	1
New books from Stata Press	2
From the Stata Bookstore	3
Latest NetCourse schedule	3

One set of labels is in use at any instance. When other Stata commands produce output, such as `describe`, `tabulate`, etc., they use the currently set language. You may change the set by typing

```
. label language languagename
```

For more information, install the free update and type `whelp label_language`.

## New books from Stata Press



Title: *An Introduction to Survival Analysis Using Stata, Revised Edition*

Authors: Mario Cleves, William Gould, & Roberto Gutierrez

Publisher: Stata Press

Copyright: 2004

Pages: 308; paperback

ISBN: 1-881228-84-3

Price: \$52.00

### An Introduction to Survival Analysis Using Stata, Revised Edition

*An Introduction to Survival Analysis Using Stata, Revised Edition* is the ideal tutorial for professional data analysts who want to learn survival analysis for the first time or who are well versed in survival analysis but not as dexterous in using Stata to analyze survival data. This text also serves as a valuable reference to those who already have experience using Stata's survival analysis routines.

Survival analysis is a field of its own requiring specialized data management and analysis procedures. Toward this end, Stata provides the `st` family of commands for organizing and summarizing survival data. The authors of this text are also the authors of Stata's `st` commands, and they are also the authors of Stata's NetCourse 631—An Introduction to Survival Analysis. This text is an outgrowth of the lecture notes for that course, and those who have taken the course will find in this text the companion text that many participants lamented not having.

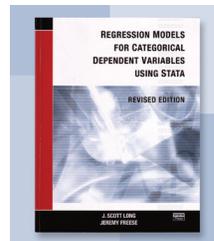
This book provides statistical theory, step-by-step procedures for analyzing survival data, an in-depth usage guide for Stata's most widely used `st` commands, and a collection of tips and pointers for using Stata to analyze survival data and present the results. This book develops from first principles the statistical concepts unique to survival data and assumes only a knowledge of basic probability and statistics and a working knowledge of Stata.

The first three chapters of the text cover basic theoretical concepts: hazard functions and cumulative hazard functions and their interpretations, survivor functions, hazard models, and a comparison of nonparametric, semiparametric, and parametric methodologies. Chapter 4 deals with censoring and truncation. The next three chapters cover the formatting, manipulation, `stsetting`, and error-checking involved in preparing survival data for analysis using Stata's `st` analysis commands. Chapter 8 covers nonparametric methods, including the Kaplan–Meier and Nelson–Aalen estimators, and the various nonparametric tests for the equality of survival experience.

Chapters 9, 10, and 11 discuss Cox regression and include various examples of fitting a Cox model, obtaining predictions, interpreting results, building models, and model diagnostics. The final four chapters cover parametric models, which are fit using Stata's `streg` command. These chapters include detailed derivations of all six parametric models currently supported in Stata, methods for determining which model is appropriate, obtaining predictions, stratification, and advanced topics such as frailty models.

The revised edition uses the new Stata graphics system throughout. It also covers areas of most recent development in Stata's survival analysis suite, namely the estimation of hazard functions via kernel smoothing and Cox models with shared frailty.

A complete table of contents, as well as online ordering information, can be found at <http://www.stata-press.com/books/saus.html>. You can also order using the enclosed bookstore order form.



Title: *Regression Models for Categorical Dependent Variables Using Stata, Revised Edition*

Authors: J. Scott Long & Jeremy Freese

Publisher: Stata Press

Copyright: 2003

Pages: 368; paperback

ISBN: 0-881228-82-7

Price: \$52.00

### Regression Models for Categorical Dependent Variables Using Stata, Revised Edition

While regression models for categorical dependent variables are ubiquitous, a discussion of how to interpret these models has been sorely lacking. *Regression Models for Categorical Dependent Variables Using Stata, Revised Edition* by Scott Long and Jeremy Freese fills this void. This book discusses how to fit and interpret regression models for categorical data with Stata and includes some commands written by the authors. Hypothesis testing and goodness-of-fit statistics are also discussed.

The book begins with a lucid introduction to Stata and then provides a general treatment of estimation, testing, fit, and interpretation in this class of models. Binary outcomes, ordinal outcomes, nominal outcomes, and count outcomes are covered in detail in separate chapters. The final chapter discusses how to fit and interpret models with special characteristics such as ordinal and nominal independent variables, interaction, and nonlinear terms. One appendix discusses the syntax of the author-written commands, and a second gives details of the datasets used by the authors in the book.

This book is filled with concrete examples. Because all the examples, datasets, and author-written commands are available from the authors at their web site, readers can easily replicate the examples using Stata. This book is ideal for students or applied researchers who want to know how to fit this type of model and understand its output.

The revised edition uses the new Stata graphics system throughout the book. In addition, the revised edition discusses multiple missing-value codes and contains updated output throughout the text.

A complete table of contents, as well as online ordering information, can be found at <http://www.stata-press.com/books/regmodcdvs.html>. You can also order using the enclosed bookstore order form.

### From the Stata Bookstore



Title: *Essential Medical Statistics*, 2d ed  
 Authors: B. R. Kirkwood & J. A. C. Sterne  
 Publisher: Blackwell Science Ltd.  
 Copyright: 2003  
 Pages: 501; paperback  
 ISBN: 0-865-42871-9  
 Price: \$49.75

### Essential Medical Statistics, 2d ed

*Essential Medical Statistics* by Betty Kirkwood and Jonathan Sterne gives an excellent overview of the basics of medical statistics and achieves a good balance between classic methodology, such as contingency tables, and more sophisticated models, such as Cox regression. As personal computers become more powerful, the use of complex models in leading medical journals is becoming more prevalent, and the second edition of this text follows this shift by focusing more on these models. A very nice section on basic meta-analysis is also included.

Covered in the text are the basics of normal-based tests and inference, including linear regression, contingency tables and logistic regression, longitudinal analysis, Poisson regression, survival analysis (including the Cox model), likelihood theory, meta-analysis, Bayesian statistics, basic design, sample-size calculations, and measurement error.

A complete table of contents, as well as online ordering information, can be found at <http://www.stata.com/bookstore/ems.html>. You can also order using the enclosed bookstore order form.



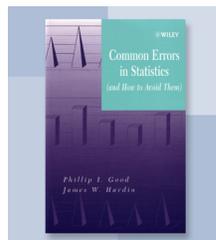
Title: *Survival Analysis*, 2d ed  
 Authors: J. Klein & M. Moeschberger  
 Publisher: Springer  
 Copyright: 2003  
 Pages: 536; hardcover  
 ISBN: 0-387-95399-X  
 Price: \$89.75

### Survival Analysis: Techniques for Censored and Truncated Data, 2d ed

*Survival Analysis: Techniques for Censored and Truncated Data* by John Klein and Melvin Moeschberger is an essential reference for any researcher using techniques of survival analysis. Ideal for self-study or for a two-term graduate sequence in survival analysis, this book maintains a technical level suited for both the medical researcher and the professional statistician. Most impressive are the number and scope of real-data examples that are presented in the first chapter and used subsequently throughout the text.

Covered topics include base terminology (hazards, censoring, mean residual life, etc.), nonparametric estimation, life tables, hazard estimation, nonparametric tests, Cox regression (including time-varying covariates and stratification), the additive hazards model, Cox regression diagnostics, parametric regression models, and multivariate analysis, including frailty models.

A complete table of contents, as well as online ordering information, can be found at <http://www.stata.com/bookstore/tctd.html>. You can also order using the enclosed bookstore order form.



Title: *Common Errors in Statistics*  
 Authors: Phillip Good & James Hardin  
 Publisher: John Wiley & Sons, Inc.  
 Copyright: 2003  
 Pages: 221; paperback  
 ISBN: 0-471-46068-0  
 Price: \$44.75

### Common Errors in Statistics

*Common Errors in Statistics* by Phillip Good and James Hardin contains common-sense, minimally technical advice on how to improve experimental design, analysis of data, and presentation of results. It provides a guide for experienced scientists as well as students learning to design and complete experiments and statistical analysis.

The text begins with a section on foundations that covers sources of error, hypotheses, and data collection. The second section, on hypothesis testing and parameter estimation, takes a harder look at statistical evaluation of the data, including strengths and limitations of various statistical procedures, and guidelines for reporting results from what information to include to how to create an informative and easy-to-understand graph. The final section covers building a model, with topics on univariate and multivariable regression, as well as validation of the model chosen.

A complete table of contents, as well as online ordering information, can be found at <http://www.stata.com/bookstore/ceis.html>. You can also order using the enclosed bookstore order form.

### Latest NetCourse™ schedule

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

- “I thought the material was excellent. I was also impressed by the opportunity the students had to ask questions and the efforts made by the Course Leaders to answer them, regardless of how complicated the question.”
- “I thought the NetCourse was superior because of the quality of the notes and the emphasis on intuition.”

A brief summary of the upcoming NetCourses is listed below. For details on how NetCourses work and for course syllabi, visit <http://www.stata.com/info/products/netcourse/>.

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

## Introduction to Stata

### Prerequisites

Stata 8

### Dates offered

October 17 – November 28

### Course Leaders

James Hassell, Allen McDowell, and Derek Wagner

### Enrollment Deadlines

October 13

### Price

\$95

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

## Introduction to Stata programming

### Prerequisites

Stata 8; basic knowledge of using Stata interactively

### Dates offered

October 17 – November 28

### Course Leaders

Allen McDowell and Kevin Crow

### Enrollment Deadlines

October 13

### Price

\$125

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

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



## HOW TO REACH US

### Stata Corporation

4905 Lakeway Drive  
College Station TX 77845  
USA

**PHONE** 979-696-4600

**FAX** 979-696-4601

**EMAIL** stata@stata.com

**WEB** www.stata.com

Please include your Stata serial number with all correspondence.

## INTERNATIONAL RESELLERS

### CreActive snc

*Serving Italy*  
tel: +39 0575 333297  
email: staff@creactive.net

### IEM

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

### Informatique Inc

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

### SOFTWARE shop Inc

*Serving Bolivia, Chile, Colombia,  
Ecuador, Peru, Venezuela*  
tel: 425-651-4090  
email: ventas@software-shop.com

### Timberlake Consultores Brasil

*Serving Brazil*  
tel: +55-11-3263-1287  
email: info@timberlake.com.br

### Timberlake Consultants Polska

*Serving Poland*  
tel: +48 600 370 435  
email: info@timberlake.pl

## INTERNATIONAL DISTRIBUTORS

### Chips Electronics

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

### Cosinus Computing BV

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

### Dittrich & Partner Consulting

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

### Ixon Technology Company Ltd

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

### Mercostat Consultores

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

### Metrika Consulting

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

### MultiON Consulting S.A. de C.V.

*Serving Belize, Costa Rica,  
El Salvador, Guatemala,  
Honduras, Mexico,  
Nicaragua, Panama*  
tel: 52 (55) 55 59 40 50  
email: info@multion.com.mx

### Ritme Informatique

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

### Scientific Solutions S.A.

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

### Survey Design & Analysis Services

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

### Timberlake Consultants

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

### Timberlake Consulting S.L.

*Serving Spain*  
tel: +34 (9) 5 560 14 30  
email:  
timberlake@timberlakeconsulting.com

### Timberlake Consultores, Lda.

*Serving Portugal*  
tel: +351 214 702 869  
email:  
timberlake.co@mail.telepac.pt

### TStat S.r.l.

*Serving Italy*  
tel: +39 0864 210101  
email: tstat@tstat.it

### Vishvas Marketing-Mix Services

*Serving India*  
tel: 91-22-25892639  
email: bandya@vmmsmeet.com