

2007 Report to Users

Alan Riley

Vice President, Software Development
StataCorp LP

2007 Nordic and Baltic Stata Users Group Meeting, Stockholm



1. Stata 10
2. Stata Growth
3. Stata Press
4. User Meetings

Began shipping in June 2007

Features

- Mixed models for binary/count data
- Exact logistic and exact Poisson
- Power analysis for survival data
- Survey *everything*
- Instrumental variables
- Dynamic panel data
- Multivariate methods
- Graph editor
- Interface enhancements
- and more ...

Began shipping in June 2007

Features

- Mixed models for binary/count data
- Exact logistic and exact Poisson
- Power analysis for survival data
- Survey *everything*
- Instrumental variables
- Dynamic panel data
- Multivariate methods
- Graph editor
- Interface enhancements
- and more ...

New commands `xtmelogit` and `xtmepoisson`

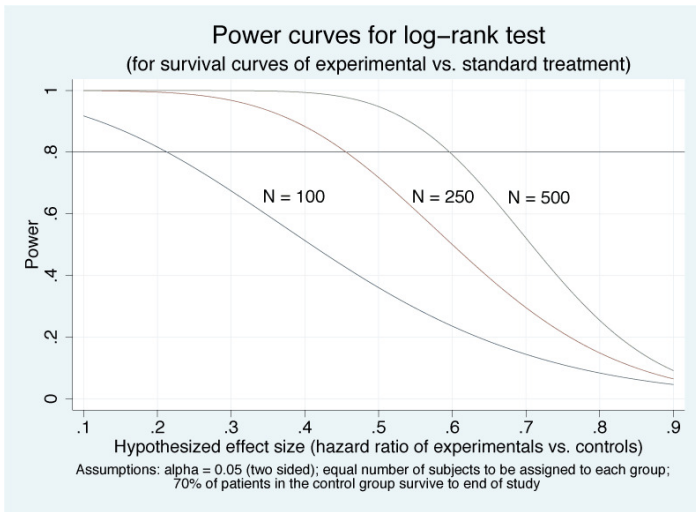
- Mixed effects for binary and count responses
- Syntax and output similar to `xtmixed`
- Random intercepts and coefficients
- Crossed-effects models
- Predict random effects and their standard errors
- Adaptive Gaussian quadrature and scalability via the quicker Laplace approximation

New commands `exlogistic` and `expoisson`

- Small-sample alternative to ML estimation
- Inference not based on asymptotic theory, but instead on exact enumerations of the sufficient-statistics space
- Think covariate-adjusted exact binomial or exact Poisson as obtained from `ci`, say.
- Can estimate parameters even in the case of perfect prediction
- More exact methods to come

Power analysis for survival studies

- Command `stpower`
- Log-rank tests, Cox models, and exponential regression
- Solve for sample size, power, or effect size
- Calculations take into account censoring, withdrawal, and unequal allocation
- Flexible syntax makes creating tables easy
- Can save output as a dataset for graphing
- GUI



Closing the book on survey

- 27 additional estimation commands made to work with survey data
- This includes `streg` and `stcox`
- Support for strata with one PSU
- Survey calculations parallelized in Stata/MP

Command `ivregress`

- Complete overhaul of the previous `ivreg`
- Implements 2SLS, LIML, and GMM estimators
- Provides robust, cluster robust, and HAC standard errors
- Several postestimation tools for tests of instrumental relevance and of overidentifying restrictions
- Similar to the widely used `ivreg2` command of Baum, Schaffer, and Stillman

New suite of commands

- `xtabond` uses lagged levels of the endogenous variables and lagged differences of exogenous variables as instruments
- `xtdpdsys` uses lagged differences of the endogenous variables as additional instruments, improving performance with highly persistent autoregressive processes
- `xtdpd` provides greatest flexibility in determining what to use for instruments, at the cost of a more complicated syntax
- One- and two-step estimators with conventional GMM, robust, and bias-corrected robust standard errors

A complete toolkit

- Discriminant Analysis: LDA, QDA , Logistic, and KNN
- MCA (Multiple Correspondence Analysis)
- Modern (Nonmetric) MDS, in addition to classical MDS in Stata 9
- CA now allows crossed (stacked) variables
- Cluster and MDS now have the Gower dissimilarity measure for a mix of binary and continuous variables

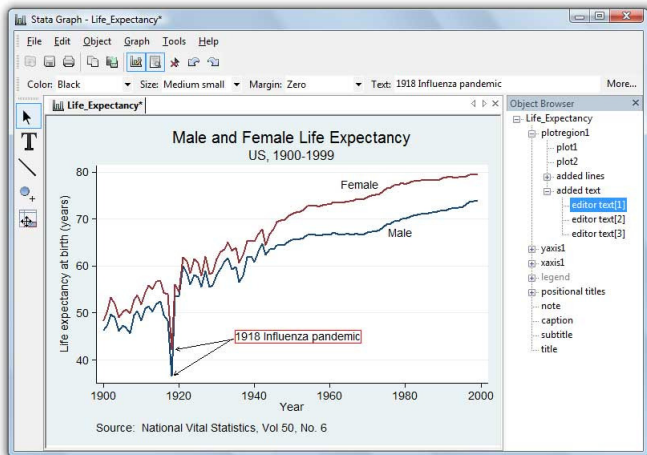
Editing graphs

- Point-and-click interface
- Click on those objects you wish to edit
- Right-clicking gives contextual menus
- Makes adding lines, arrows, and text easy
- You can undo your changes
- And its darn fun to play with

Editing graphs

- Point-and-click interface
- Click on those objects you wish to edit
- Right-clicking gives contextual menus
- Makes adding lines, arrows, and text easy
- You can undo your changes

- And its darn fun to play with



Interface enhancements

- Tabbed graph windows
- Redesigned viewer with a Forward button
- More Review window features
- More Variables window features
- WYSIWYG Copy/Paste output

Copy Text

Document - WordPad

File Edit View Insert Format Help

Arial 10 Western **B** / U ☺

Number of obs = 74
 F(1, 72) = 194.71
 Prob > F = 0.0000
 R-squared = 0.7300
 Adj R-squared = 0.7263
 Root MSE = .0067

gpm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
weight	.0000141	1.01e-06	13.95	0.000	.0000121 .0000161
_cons	.0077077	.0031426	2.45	0.017	.0014431 .0139723

For Help, press F1

Better: Copy Text, change to fixed-width font

The screenshot shows a WordPad window titled "Document - WordPad" with a menu bar (File, Edit, View, Insert, Format, Help) and a toolbar. The font is set to "Lucida Console", size 8, and style "Western". The text content is Stata output, displayed in a fixed-width font. The output includes summary statistics and regression coefficients.

Source	SS	df	MS			
Model	.008729651	1	.008729651	Number of obs =	74	
Residual	.003227977	72	.000044833	F(1, 72) =	194.71	
Total	.011957628	73	.000163803	Prob > F =	0.0000	
				R-squared =	0.7300	
				Adj R-squared =	0.7263	
				Root MSE =	.0067	

gpm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
weight	.0000141	1.01e-06	13.95	0.000	.0000121	.0000161
_cons	.0077077	.0031426	2.45	0.017	.0014431	.0139723

For Help, press F1

Best: Copy as Picture

Document - WordPad

File Edit View Insert Format Help

Arial 10 Western **B** / U ☺

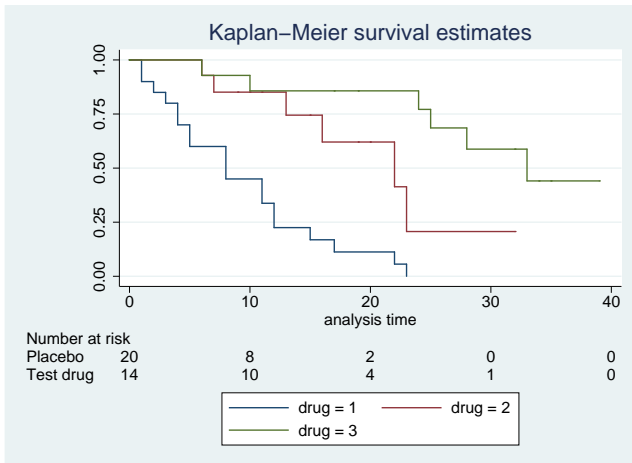
Source	SS	df	MS			
Model	.008729651	1	.008729651	Number of obs =	74	
Residual	.003227977	72	.000044833	F(1, 72) =	194.71	
Total	.011957628	73	.000163803	Prob > F =	0.0000	
				R-squared =	0.7300	
				Adj R-squared =	0.7263	
				Root MSE =	.0067	

gpm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
weight	.0000141	1.01e-06	13.95	0.000	.0000121	.0000161
_cons	.0077077	.0031426	2.45	0.017	.0014431	.0139723

For Help, press F1

Just to name a few

- Save estimation results to disk
- Nonlinear seemingly unrelated regression
- `optimize()` in Mata
- Easier syntax for choice models
- New date/time formats with millisecond resolution
- Automation (also known as OLE Automation)
- At-risk tables in survival graphs



Stata 1, January 1985

- 44 commands
- 175 pages of documentation

Stata 8, January 2003

- over 600 commands
- 4652 pages of documentation

Stata 1, January 1985

- 44 commands
- 175 pages of documentation

Stata 8, January 2003

- over 600 commands
- 4652 pages of documentation

Stata 9, April 2005

- over 700 commands including new matrix language Mata
- 6413 pages of documentation

Stata 10, June 2007

- 924 commands
- 8035 pages of documentation

Stata 9, April 2005

- over 700 commands including new matrix language Mata
- 6413 pages of documentation

Stata 10, June 2007

- 924 commands
- 8035 pages of documentation



Lines of Code

Category	Stata 8	Stata 9	Stata 9/MP	Stata 10
Source	928,629	1,262,173	1,407,474	1,596,350
Abs. Diff.		333,544	145,301	188,876
Rel. Diff.		35.9%	11.5%	13.4%
Help	132,844	255,956	274,279	323,673
Abs. Diff.		123,112	18,323	49,394
Rel. Diff.		92.7%	7.2%	18.0%
Certification	249,188	510,562	563,856	697,701
Abs. Diff.		261,374	53,294	133,845
Rel. Diff.		104.9%	10.4%	23.7%
Total	1,310,661	2,028,691	2,245,609	2,617,724
Abs. Diff.		718,030	216,918	372,115
Rel. Diff.		54.8%	10.7%	16.6%

Category	Stata 8	Stata 9	Stata 9/MP	Stata 10
Source	928,629	1,262,173	1,407,474	1,596,350
Abs. Diff.		333,544	145,301	188,876
Rel. Diff.		35.9%	11.5%	13.4%
C code	492,385	687,658	820,071	864,602
Abs. Diff.		195,273	132,413	44,531
Rel. Diff.		39.7%	19.3%	5.4%
Mata code	0	10,265	12,391	37,266
Abs. Diff.		10,265	2,126	24,875
Rel. Diff.		–	20.7%	200.8%
Ado code	263,833	354,108	362,669	420,148
Abs. Diff.		90,275	8,561	57,479
Rel. Diff.		34.2%	2.4%	15.8%
Control Files	172,411	210,142	212,343	274,334
Abs. Diff.		37,731	2,201	61,991
Rel. Diff.		21.9%	1.0%	29.2%

Books published in the last year

- **Generalized Linear Models and Extensions, 2nd Edition**
by James Hardin and Joe Hilbe
- **An Introduction to Modern Econometrics Using Stata**
by Christopher F. Baum
- **Thirty-three Stata Tips**
by Joseph Newton and Nicholas Cox (eds)

Forthcoming books, 2007

- **Workflow in Data Analysis Using Stata**
by J. Scott Long
- **Applied Microeconometrics Using Stata**
by A. Colin Cameron and Pravin K. Trivedi
- **A Visual Guide to Stata Graphics, 2nd Edition**
by Michael N. Mitchell
- **An Introduction to Forecasting Time Series Using Stata**
by Robert Yaffee

Forthcoming books, 2008

- **An Introduction to Stata Programming**
by Christopher F. Baum
- **A Gentle Guide to Advanced Statistics Using Stata**
by Alan Acock and Peter Lachenbruch
- **A Guide to Stochastic Frontier Models: Specification and Estimation**
by Subai Kumbhakar and Hung-Jen Wang

A record 7 meetings this year

- **German: Essen**
April 2
- **North American: Boston**
August 13–14
- **Nordic and Baltic: Stockholm**
September 7
- **UK: London**
September 10-11
- **Italian: Rome**
September 24-25
- **West Coast: Los Angeles**
October 25-26
- **Seminars on Stata: Washington, DC**
November 2

