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
Power curves for log-rank test
(for survival curves of experimental vs. standard treatment)

Assumptions: alpha = 0.05 (two sided); equal number of subjects to be assigned to each group; 70% of patients in the control group survive to end of study.
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

```plaintext
Source | SS     | df | MS      | Number of obs = 74
-------+--------+----+---------+-----------------------------
Model  | 0.0087 | 1  | 0.0087  | F( 1, 72) = 194.71
       | 1      | 72 | 0.0004  | Prob > F = 0.0000
Residual| 0.0032 | 72 | 0.0004  | R-squared = 0.7300
       | 0.0032 | 72 | 0.0004  | Adj R-squared = 0.7263
Total  | 0.0119 | 73 | 0.0001  | Root MSE = 0.0067
       | 0.0119 | 73 | 0.0001  |

      | Coef.  | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
-------+--------+-----------+-------+-----+----------------------|
      | gpm    |          |       |     |                      |
-------+--------+-----------+-------+-----+----------------------|
      | weight | 0.000014 | 1.00e-06 | 13.95 | 0.000 | 0.0000121 0.0000161 |
      | _cons  | 0.0077    | 0.0031  | 2.45 | 0.017 | 0.0014431 0.0139723 |
```
Better: Copy Text, change to fixed-width font
Best: Copy as Picture

![Image of ANOVA table from Stata 10](image-url)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>0.008729651</td>
<td>1</td>
<td>0.008729651</td>
</tr>
<tr>
<td>Residual</td>
<td>0.003227977</td>
<td>72</td>
<td>0.000044833</td>
</tr>
<tr>
<td>Total</td>
<td>0.011957628</td>
<td>73</td>
<td>0.000163803</td>
</tr>
</tbody>
</table>

Number of obs = 74
F( 1, 72) = 194.71
Prob > F = 0.0000
R-squared = 0.7300
Adj R-squared = 0.7263
Root MSE = 0.0067

| gpm      | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------|-------|-----------|-------|-----|----------------------|
| weight _cons | 0.000141 | 1.01e-06 | 13.95 | 0.000 | 0.000121, 0.000161 |
| _cons     | 0.0077077  | 0.0031426 | 2.45  | 0.017 | 0.0014431, 0.0139723 |
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
Kaplan–Meier survival estimates

Number at risk

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Test drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>drug</td>
<td>20</td>
<td>14</td>
</tr>
<tr>
<td>analysis time</td>
<td>0 20 30 40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 8 4 1 0</td>
<td>0 10 4 1 0</td>
</tr>
</tbody>
</table>

- blue: drug = 1
- red: drug = 2
- green: drug = 3
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
Report to Users

Stata Growth

Bobby’s Manuals

A. Riley (StataCorp)
### Lines of Code

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

### Components of source code

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

  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
Statalist Subscriptions
08oct2003 – 22jun2006

7,861 total subscriptions