### Highlights of What’s New in Stata 11

<table>
<thead>
<tr>
<th>Multiple imputation (MI)</th>
<th>GMM</th>
<th>Competing-risks regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Univariate imputation</td>
<td>• Linear and nonlinear models</td>
<td>• Time-varying covariates</td>
</tr>
<tr>
<td>• Multivariate imputation</td>
<td>• One-step, two-step, and iterative estimators</td>
<td>• Cumulative incidence graphs</td>
</tr>
<tr>
<td>• MI Control Panel</td>
<td>• Cross-sectional, time-series, and panel data</td>
<td>• Subhazard ratios</td>
</tr>
<tr>
<td>• Manage MI datasets</td>
<td>• Panel-style instruments</td>
<td>(see page 4)</td>
</tr>
<tr>
<td>(see page 2)</td>
<td>(see page 6)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multivariate time series</th>
<th>PDF documentation</th>
<th>Factor variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>• State-space models</td>
<td>• Comes with every copy of Stata</td>
<td>• Interactions</td>
</tr>
<tr>
<td>• Dynamic-factor models</td>
<td>• Includes all manuals</td>
<td>• Polynomial terms</td>
</tr>
<tr>
<td>• Multivariate GARCH</td>
<td>• Integrated with help files</td>
<td>• Create indicators automatically</td>
</tr>
<tr>
<td>(see page 5)</td>
<td>(see page 6)</td>
<td>(see page 4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel data and mixed models</th>
<th>Marginal analysis</th>
<th>More</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Error covariance structures</td>
<td>• Estimated marginal means</td>
<td>• Ten random-number generators</td>
</tr>
<tr>
<td>• Unit-root tests</td>
<td>• Predictive margins</td>
<td>• More density and distribution functions</td>
</tr>
<tr>
<td>• Standard errors for BLUPs</td>
<td>• Average marginal effects</td>
<td>• DFBETAs, likelihood displacement values, and LMAX statistics after Cox regression</td>
</tr>
<tr>
<td>(see page 5)</td>
<td>(see page 5)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>New interface</th>
<th>Fonts in graphics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Variables Manager</td>
<td>• Italics and bold</td>
<td>• Normal, GED, and Student’s $t$ errors in ARCH</td>
</tr>
<tr>
<td>• Live view on data</td>
<td>• Greek letters</td>
<td>• Multivariate tests</td>
</tr>
<tr>
<td>• Data and variable filters</td>
<td>• Mathematical symbols</td>
<td>• Numerical derivative functions</td>
</tr>
<tr>
<td>• Syntax highlighting</td>
<td>• Superscripts and subscripts</td>
<td>• Associative arrays</td>
</tr>
<tr>
<td>• Code folding</td>
<td>• Multiple fonts</td>
<td>• Soundex name-matching functions</td>
</tr>
<tr>
<td>• Bookmarks</td>
<td>(see page 3)</td>
<td>• Object-oriented programming</td>
</tr>
<tr>
<td>(see page 3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Multiple imputation for missing data

Stata’s new mi command provides a full suite of multiple-imputation methods for the analysis of incomplete data, data for which some values are missing. mi provides both the imputation and the estimation steps. mi’s estimation step encompasses both estimation on individual datasets and pooling in one easy-to-use procedure. Features are provided to enable examining the pattern of missing values in the data. Flexible imputation methods are provided, including five univariate imputation methods that can be used as building blocks for multivariate imputation, as well as MVN.

We want to study the linear relationship between \( y \) and predictors \( x_1 \) and \( x_2 \). Our data contain missing values, however, and standard casewise deletion would result in a 40% reduction in sample size! We will fit the model using multiple imputation. First, we impute missing values and arbitrarily create five imputation datasets:

\[ \text{mi impute mvi y x1 x2, add(5)} \]

That done, we can fit the model:

\[ \text{mi estimate: regress y x1 x2} \]

mi estimate fits the specified model (linear regression here) on each of the imputation datasets (five here) and then combines the results into one multiple-imputation inference.

mi can import already imputed data, including data from NHANES or ice, or you can form imputations yourself. Either way, dealing with multiple copies of the data is the bane of multiple-imputation analysis, and Stata solves that problem because mi automatically keeps all datasets in sync. You can create or drop variables or observations just as if you were working with one dataset! You can merge datasets, append datasets, and even reshape your data. The fact that the data management you perform might need to be carried out consistently over 5, 50, or even 500 datasets is irrelevant.

Multiple-Imputation Control Panel

The Multiple-Imputation Control Panel unifies mi’s capabilities and guides you from the beginning of your analysis to the end, from examining the pattern of missing values to performing estimation and inference with multiple imputation.

- Use Examine to check missing-value patterns so that you can determine the appropriate imputation method.
- Use Setup to set up your data when you need to impute values. If your data already contain imputations, skip Setup and use Import instead.
- Use Impute to create imputations. You can impute one variable or multiple variables simultaneously. Regression, predictive mean matching, logistic, ordered logistic, multinomial logistic, and MVN are all available. With monotone missingness, you can combine imputation methods.
- Use Manage to perform data management.
- Use Estimate to fit your model and combine results.
- Use Test to perform tests of hypotheses.
Graphs now support multiple fonts and symbols
You can now include multiple fonts in regular, bold, and italics; the full Greek alphabet in upper- and lowercase; and over 70 mathematical symbols anywhere that text appears in graphs.

Graphs now accept an extended version of Stata’s text markup and control language, SMCL. Greek letters and mathematical symbols have intuitive, easy-to-remember names (tags). To make the title of the x axis of your graph say “β-carotene”, you just specify xtitle(“{\beta}-carotene”). To include χ² in your text, specify {&chi}{superscript:2}, or even shorter, {&chi}{sup:2}.

Additional tags make it easy to select from among sans serif, serif, monospace, and symbol fonts.

Of course, these new features are also available via the dialog boxes and through the interactive Graph Editor.

**New Data Editor, Do-file Editor, and Variables Manager**

Stata’s new Data Editor, Do-file Editor, and Variables Manager make your data-management and programming tasks simpler than ever.

The new Data Editor under Windows includes syntax highlighting, code folding (so you can hide blocks of code while you focus on other parts), the ability to work on multiple files simultaneously, and no limit on file size. The new Data Editor allows you to sort variables, apply filters to view subsets of your data, take snapshots (so you can easily undo changes you make), and enter dates and times. The Data Editor is a live view onto your data: change your data by running a Stata command, and see the results reflected in the Data Editor instantly.

The new Variables Manager allows you to change variable names, storage types, formats, labels, and notes with the click of your mouse. You can apply a filter to view only variables that meet criteria you specify.
Factor variables

Stata now handles factor (categorical) variables elegantly. You can now prefix a variable with \texttt{i.} to specify indicators for each level (category) of the variable. You can put a \# between two variables to create an interaction—indicators for each combination of the categories of the variables. You can put \texttt{##} instead to specify a full factorial of the variables—main effects for each variable and an interaction. If you want to interact a continuous variable with a factor variable, just prefix the continuous variable with \texttt{c.}. You can specify up to eight-way interactions.

We run a linear regression of cholesterol level on a full factorial of age group and whether the person smokes along with a continuous body mass index (\texttt{bmi}) and its interaction with whether the person smokes.

\begin{verbatim}
. regress cholesterol smoker##agegrp bmi smoker#c.bmi
\end{verbatim}

We could have used parenthesis binding, to type the same model more briefly:

\begin{verbatim}
. regress cholesterol smoker##(agegrp c.bmi)
\end{verbatim}

Base levels can be changed on the fly: \texttt{i.agegrp} uses the default base level of 1, whereas \texttt{b3.agegrp} makes 3 the base level.

The level indicator variables are not created in your dataset, saving lots of space.

Factor variables are integrated deeply into Stata’s processing of variable lists, providing a consistent way of interacting with both estimation and postestimation commands.

Competing-risks regression

Competing-risks regression provides a useful alternative to Cox regression in the presence of one or more competing risks. For example, say that you are studying the time from initial treatment for cancer to recurrence of cancer in relation to the type of treatment administered and demographic factors. Death is a competing event: the person under treatment may die, impeding the occurrence of the event of interest, recurrence of cancer. Unlike censoring, which merely obstructs you from viewing the event, a competing event prevents the event of interest from occurring altogether, and your analysis should adjust accordingly.

Stata’s new \texttt{stcrreg} command implements competing-risks regression based on Fine and Gray’s proportional subhazards model. In Cox regression, you focus on the survivor function, which indicates the probability of surviving beyond a given time. In competing-risks regression, you instead focus on the cumulative incidence function, which indicates the probability of the event of interest happening before a given time. Competing-risks regression is semiparametric in that the baseline subhazard of the event of interest is left unspecified and the effects of covariates are assumed to be proportional. Time-varying covariates and coefficients are allowed.

Stata’s \texttt{stcurve} command has been fully updated to work with \texttt{stcrreg} and to produce cumulative incidence function graphs, and you can use \texttt{predict} to obtain an array of statistics.
**Marginal means, adjusted predictions, and marginal effects**

Stata 11 does margins. Margins are statistics calculated from predictions of a previously fit model at fixed values of some covariates and averaging or otherwise integrating over the remaining covariates. Included are estimated marginal means, least-squares means, average and conditional marginal/partial effects (as derivatives or elasticities), average and conditional adjusted predictions, predictive margins, and more.

Predictive margins are particularly apt for survey data and other samples that represent a population (either by weighting or random sampling). In such cases, the margins or marginal effects can be considered estimates of the same statistics in the population or data-generating process (DGP), and their standard errors and confidence intervals can be used to make inferences about the population or DGP.

You can compute the statistics for each subgroup or subpopulation designated by unique values of `varlist` by using the `by(varlist)` option.

Because of Stata 11’s new factor-variable features, `margins` understands squared variables and other polynomial forms, so you can get the overall marginal effect for age even when age enters as age and age squared.

Stata 11’s `margins` includes options so that the reported standard errors account for complex survey sampling including weights, sampling units, pre- and poststratification, and subpopulations.

After `margins`, you can test equality or any linear or nonlinear combination of estimated results.

```stata
.test 0.sex = 1.sex
( 1) 0b.sex - 1.sex = 0
   chi2(  1) =    8.53
   Prob > chi2 =    0.0035
```

You can even form linear and nonlinear combinations, so you can examine risk ratios, risk differences, etc., and perform tests with them.

`margins` works after almost every Stata estimation command.

---

**State-space and dynamic-factor models**

Stata’s new `sspace` and `dfactor` commands make it easy to fit a wide variety of multivariate time-series models by casting them as linear state-space models, including vector autoregressive moving-average (VARMA) models, structural time-series (STS) models, and dynamic-factor models.

We have data on industrial production (`ipman`), real disposable income (`dsp`), weekly hours worked (`awhi`), and the unemployment rate (`unrate`). We suspect there exists a latent factor that can explain all four of these series, and we conjecture that the latent factor follows an AR(2) process.

First, we fit our model by typing

```stata
.dfactor (ipman dsp awhi unrate = , noconstant)
  (f = , ar(1/2))
```

With our model fit, let’s obtain dynamic forecasts for disposable income beginning in December 2008:

```stata
.predict dsp_f, dynamic(tm(2008m12))
.tsline dsp dsp_f if month >= tm(2005m1)
```

---

Because of Stata 11’s new factor-variable features, `margins` understands squared variables and other polynomial forms, so you can get the overall marginal effect for age even when age enters as age and age squared.

Stata 11’s `margins` includes options so that the reported standard errors account for complex survey sampling including weights, sampling units, pre- and poststratification, and subpopulations.

After `margins`, you can test equality or any linear or nonlinear combination of estimated results.

```stata
.test 0.sex = 1.sex
( 1) 0b.sex - 1.sex = 0
   chi2(  1) =    8.53
   Prob > chi2 =    0.0035
```

You can even form linear and nonlinear combinations, so you can examine risk ratios, risk differences, etc., and perform tests with them.

`margins` works after almost every Stata estimation command.
Generalized method of moments (GMM)

Stata's new *gmm* command makes generalized method of moments estimation as simple as nonlinear least-squares regression and nonlinear seemingly unrelated regression. Just specify your residual equations by using substitutable expressions, list your instruments, select a weight matrix, and obtain your results.

Here we fit a Poisson model of the number of doctor visits as a function of gender, income, and whether a person has a chronic disease or private health insurance. We have reason to believe that income is endogenous, so we use age and race as instruments.

By default, *gmm* uses the two-step estimator and a weight matrix that assumes the errors are independent but not necessarily identically distributed.

Whether our instruments are valid is certainly open for debate—age likely influences the number of doctor visits—and we can test their validity by using *estat overid*.

```
. estat overid
```

Test of overidentifying restriction:

```
  Hansen’s J chi2(2) = 9.52598  (p = 0.0085)
```

Hansen’s J statistic indicates that one or more of our instruments are not valid.

*gmm* allows for weight matrices that assume observations are independent and identically distributed, independent but not identically distributed, clustered, or autocorrelated. Automatic bandwidth selection is available for heteroskedasticity- and autocorrelation-consistent weight matrices.

In addition to standard instruments, *gmm* allows you to create panel-style instruments used in dynamic and other panel models with endogenous regressors.

For more complicated analyses, *gmm* allows you to write a program to evaluate your residual equations instead of using substitutable expressions. These programs are structured like those that *ml*, *nl*, and *nlsur* use. Both the interactive and programmable versions of *gmm* allow you to specify analytic derivatives.

PDF documentation

Stata now ships with complete PDF documentation! Stata 11’s new PDF manuals are fully integrated with Stata’s interactive help system. Click and jump to the right manual, the right page! PDF manuals ship with every copy of Stata. Of course, printed documentation is still available.

Order now

To place your order, visit www.stata.com
Mark your calendar now for the 2009 Stata Conference in Washington, DC, on July 30 and 31. The conference will be your best chance to meet with other Stata users and your only chance this year to talk with Stata developers at a meeting in the United States.

In addition to user contributions, the conference includes overviews and tutorials on new Stata 11 features presented by StataCorp developers.

The tentative schedule for the conference is as follows. It looks to be busy and exciting. For abstracts, please visit www.stata.com/meeting/dcconf09/abstracts.html

### Schedule

**Thursday, July 30**
- 8:30 AM  |  **Easy approaches to GMM**  
  David Drukker, StataCorp
- 9:20 AM  |  **Mixed-process models with cmp**  
  David Roodman, Center for Global Development
- 10:10 AM |  Coffee break
- 10:25 AM |  **Multivariate time series**  
  David Drukker, StataCorp
- 11:15 AM |  **Survey statistics in Stata**  
  Jeff Pitblado, StataCorp
- 12:05 PM |  Lunch break
- 1:00 PM  |  **Regression diagnostics for survey data**  
  Rick Valliant, University of Maryland
- 1:45 PM  |  **Using Stata for subpopulation analysis of complex sample survey data**  
  Brady West, University of Michigan
- 2:30 PM  |  Coffee break
- 2:45 PM  |  **Implementing econometric estimators with Mata**  
  Christopher F. Baum, Boston College
- 3:10 PM  |  **Estimating high-dimensional fixed-effects models**  
  Paulo Guimaraes, University of South Carolina
- 3:45 PM  |  **Data envelopment analysis in Stata**  
  Choonjoo Lee and Ji Yong-bae, Korea National Defense University
- 4:15 PM  |  Coffee break
- 4:30 PM  |  **Estimating the fractional response model with an endogenous count variable**  
  Hoa Nguyen, Michigan State University
- 4:50 PM  |  **Threshold regression with threg**  
  Mei-Ling Ting Lee and Tao Xiao, University of Maryland
- 5:20 PM  |  Causal inference  
  Austin Nichols, Urban Institute
- 6:00 PM  |  Adjourn
- 7:00 PM  |  Optional dinner at Oyamel (401 7th St. NW)

**Friday, July 31**
- 8:30 AM  |  **Handling categorical covariates gracefully**  
  Jeff Pitblado, StataCorp
- 9:20 AM  |  **Between tables and graphs**  
  Nicholas J. Cox, Durham University (UK)
- 10:10 AM |  Coffee break
- 10:25 AM |  **Integrating Stata into your workflow**  
  Bill Rising, StataCorp
- 11:15 AM |  **Stata in large-scale development**  
  Michael Lokshin, The World Bank
- 12:05 PM |  Lunch break
- 1:00 PM  |  **Stata for microtargeting using C++ and ODBC**  
  Masahiko Aida, Greenberg Quinlan Rosner
- 1:20 PM  |  **A Stata regression-space database server module**  
  Mario Alberto Barabino, Bergamo University (Italy)
- 1:45 PM  |  **Meta-analytic depiction of ordered categorical diagnostic test accuracy in ROC space**  
  Ben Dwamena, University of Michigan
- 2:30 PM  |  Coffee break
- 2:45 PM  |  **Stata commands for moving data between PHASE and HaploView**  
  Chuck Huber, Texas A&M Health Science Center School of Rural Public Health
- 3:05 PM  |  **Automated individualized student assessment**  
  Stas Kolenikov, University of Missouri
- 3:30 PM  |  **Altruism squared: The economics of Statalist exchanges**  
  Martin Weiss, University of Tuebingen (Germany)
- 3:50 PM  |  **Implementing custom graphics in Stata**  
  Sergiy Radyakin, The World Bank
- 4:15 PM  |  Coffee break
- 4:30 PM  |  **Report to users**  
  Bill Gould, StataCorp
- 5:15 PM  |  **Wishes and grumbles: User feedback and Q&A**

Register: [www.stata.com/meeting/dcconf09/](http://www.stata.com/meeting/dcconf09/)

### Fees

<table>
<thead>
<tr>
<th>Price</th>
<th>Student price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single day</td>
<td>$125</td>
</tr>
<tr>
<td>Both days</td>
<td>$195</td>
</tr>
<tr>
<td>Optional dinner</td>
<td>$38</td>
</tr>
</tbody>
</table>
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.

Copyright 2009 by StataCorp LP.