# The STATA News ®

Statistics Graphics Data Management & Analysis

STATA® 13

Have you upgraded yet?

Feature-packed Stata 13 started shipping in June. Here are just some of the highlights:

- Treatment effects
- Power and sample size
- Easier data import

- Generalized multilevel/mixed models
- Long strings—*really* long strings
- Random-effects ordered probit/logit

- Generalized SEM
- Forecast models
- Effect sizes
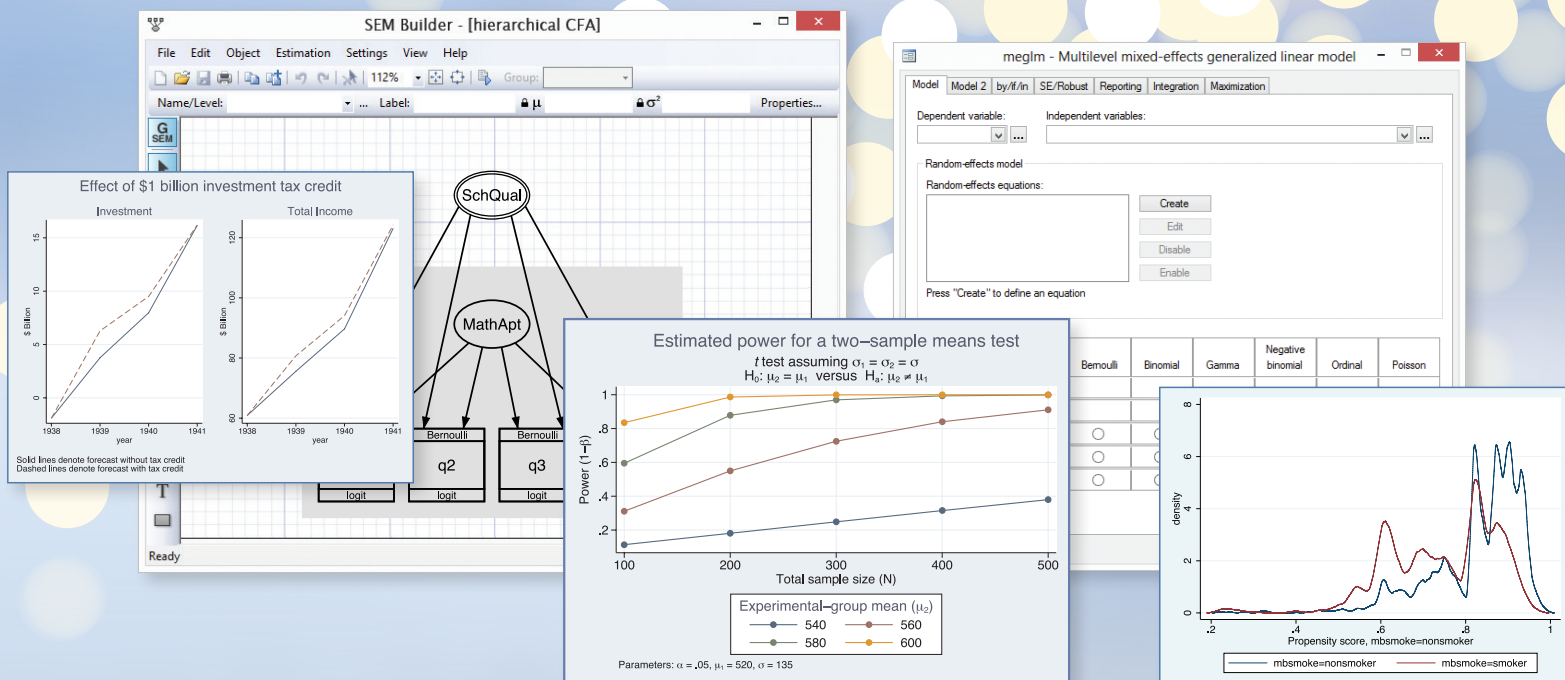
The all-new Project Manager lets you organize all of your datasets, do-files, and ado-files for easier access and collaboration. New APIs let you call Java programs from within Stata and access your datasets, matrices, and other objects. New tools make importing and exporting data easier than ever.

Upgrade to Stata 13 today to see what you've been missing!

stata.com/stata13
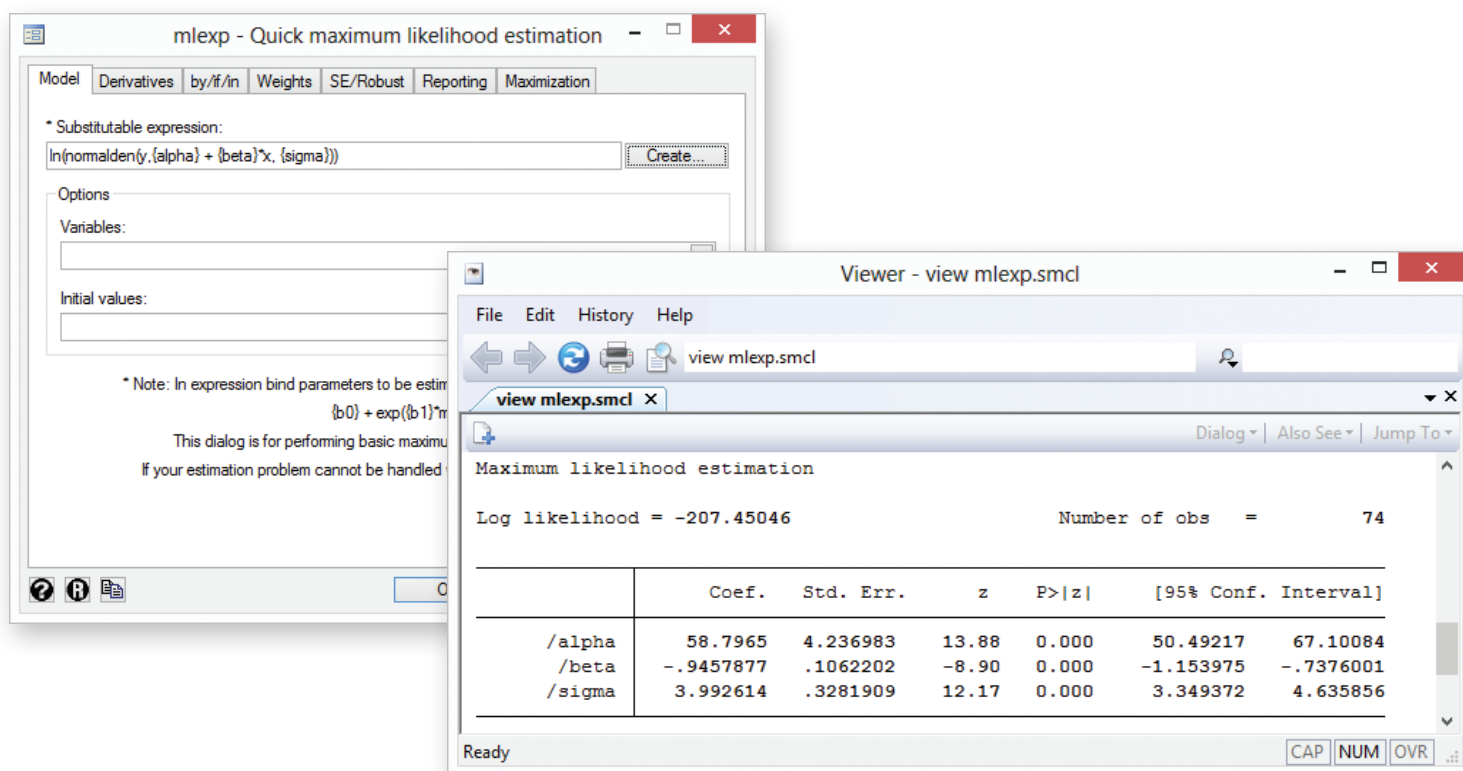
In the spotlight: mlexp

We are really excited about the new features added in Stata 13, and because there are so many, you may have missed a small but extremely useful one. I want to tell you about the new **mlexp**.

Stata has long had **ml**, which provides a set of tools to implement maximum likelihood estimators. Indeed, many of Stata's most popular commands, such as **probit**, **mlogit**, and **logistic**, are implemented using those tools. Many users have implemented their own commands with **ml** as well. The downside is that to use **ml**, you must write your own program, and even uttering the word "program" is often enough to scare off some users. Sometimes, a "quick–and–dirty" method is all you need to get the job done.

That's where **mlexp** comes in.

With **mlexp**, you just specify your log-likelihood function and click on OK. To specify the log–likelihood function, you use the same substitutable expressions that **nl**, **nlsur**, and **gmm** use. A substitutable expression is very similar to an expression you would use with **generate** or **replace**, with one additional feature: parameters that you want to estimate are placed in braces. **mlexp** looks at your substitutable expression, figures out what the parameters are, and does the rest.

Simple classical linear regression assumes that the dependent variable y is related to the independent variable x as $y_i = \alpha + \beta\, x_i + e_i$, where e_i is a normally distributed error term with a mean of 0 and standard deviation σ. Together with Stata's **normalden()** normal probability density function, this model is a snap to estimate with **mlexp**:

Of course, in this case, we could have just as well used **regress**, but think about how you learned maximum likelihood estimation. If your courses were like mine, all you saw were lots of mathematics and a promise to see applications later. **mlexp** provides an easy way for instructors to illustrate maximum likelihood and let students "get their feet wet" early on.

You can also use **mlexp** to explore the properties of maximum likelihood estimators, such as how well they work with finite samples. For example, say we are investigating the beta distribution and want to know how precisely we can estimate the parameters. Let's see what happens if we have 20 observations from a beta(1, 2) distribution:

```
. set seed 1
. set obs 20
. generate y = rbeta(1, 2)
. mlexp (ln(betaden({alpha}, {beta}, y)))
Maximum likelihood estimation
Log likelihood =  2.0427892                        Number of obs   =         20
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
/alpha	.7792227	.2178846	3.58	0.000	.3521766	1.206269
/beta	1.284411	.39757	3.23	0.001	.505188	2.063634

With these confidence intervals, we need more than 20 observations if we hope to estimate the parameters of a beta distribution with any degree of certainty.

What if we have 2,000 observations?

```
. clear
. set seed 1
. set obs 2000
. generate y = rbeta(1, 2)
. mlexp (ln(betaden({alpha}, {beta}, y)))

Maximum likelihood estimation

Log likelihood =  383.04713                        Number of obs   =       2000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
/alpha	1.02032	.028882	35.33	0.000	.9637123	1.076928
/beta	2.024953	.0639167	31.68	0.000	1.899678	2.150227

I've shown the basic features of **mlexp**, but it can do much more. You can perform weighted estimation with it, and you can even use the **svy:** prefix with it to perform survey maximum-pseudolikelihood estimation. You can also specify the derivatives of the log-likelihood function, and you can obtain bootstrap, jackknife, robust, or cluster–robust standard errors.

Give **mlexp** a try. It's a simple command with a lot of uses.

Find out more about the new features added in Stata 13 at **stata.com/stata13**.

—**Brian Poi**
Senior Economist

In the spotlight: meglm

Have you ever had data where the units of analysis were grouped at multiple levels? For example, patients who were treated by the same physician or students who were in the same class? If you answered yes, then Stata 13's **meglm** may be of interest to you.

meglm fits mixed-effects generalized linear models, also known as multilevel models or hierarchical models. These models contain fixed effects and random effects. Fixed effects are observed characteristics used to model the outcomes for the primary units of analysis; random effects model the unobservable characteristics of the nesting groups, for example, hospital quality or teacher productivity.

meglm fits mixed-effects models for a variety of generalized linear response functions, including binary outcomes (probit, logit, cloglog), count outcomes (Poisson, negative binomial), ordered outcomes (ordered probit, ordered logit, ordered cloglog), and more.
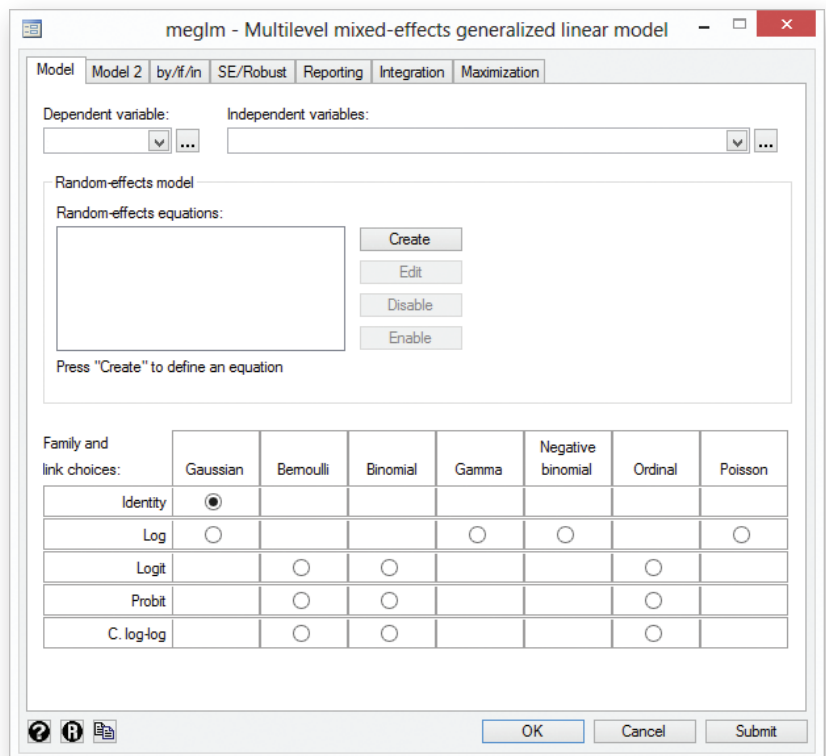
Example

We have data on 134 employees of 27 large multinational corporations, and we have repeated observations on these employees over the years. On average, we have 12 years of data. For some employees, we have 28 years. Our data include professional status (1,2,3,4), age, education, and years of job experience. To fit an ordinal logistic mixed-effects regression, we use **meglm** and specify an ordinal family and a logit link. We can use Stata's parenthetical form for multiple equations,

```
meglm (status educ c.age##c.age exp) (firm:) (id:), family(ordinal) link(logit)
```

or we can use the shorthand form,

```
meglm status educ c.age##c.age exp || firm: || id:, family(ordinal) link(logit)
```

You may be familiar with the latter: it was used by Stata's **xtmixed** command (renamed **mixed** in Stata 13). In each case, the order of the random-effects equations does matter. In our case, individuals are nested within firms; thus we specify the firm level before the ID level. By convention, the fixed-effects equation is specified first, but it can be specified anywhere. Here we put it at the very end to emphasize the nesting structure:

```
meglm (firm:) (id:) (status educ c.age##c.age exp), family(ordinal) link(logit)
```

Each of the three syntaxes results in the following output.

```
Mixed-effects GLM                                    Number of obs      =        1574
Family:                    ordinal
Link:                        logit


                      No. of        Observations per Group
 Group Variable       Groups    Minimum    Average    Maximum

          firm            27         18       58.3        137
            id           134          1       11.7         28


Integration method: mvaghermite                 Integration points =           7

                                                Wald chi2(4)        =      126.12
Log likelihood = -2076.5262                     Prob > chi2         =      0.0000

      status         Coef.    Std. Err.     z     P>|z|    [95% Conf. Interval]

        educ      .4172262    .0400348   10.42    0.000     .3387595     .495693
         age      .0808823    .0502453    1.61    0.107    -.0175967    .1793613

 c.age#c.age     -.0395296     .025602   -1.54    0.123    -.0897087    .0106495

         exp      .6977023    .1501874    4.65    0.000     .4033403    .9920642

       /cut1     -.2134323    .1428344   -1.49    0.135    -.4933827    .0665181
       /cut2      1.058385    .1445196    7.32    0.000     .7751313    1.341638
       /cut3      2.267386    .1538823   14.73    0.000     1.965783     2.56899

firm
   var(_cons)     .0465933    .0461825                      .0066777    .3251019

firm>id
   var(_cons)     .1602062    .0667473                      .0708011    .3625087

LR test vs. ologit regression:         chi2(2) =     23.77   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

The estimation table reports the fixed effects, the estimated cutpoints, and the estimated variance components of the random effects. The fixed effects can be interpreted just as you would interpret the output from **ologit**. We find that professional status increases with education and experience. We find that age is not a significant predictor of professional status.

```
. testparm c.age c.age#c.age

 ( 1)  [status]age = 0
 ( 2)  [status]c.age#c.age = 0

        chi2(  2) =      3.36
      Prob > chi2 =    0.1867
```

Underneath the fixed effects and the cutpoints, the table shows the estimated variance components. The variance-component estimates are organized and labeled according to level. The variance component for ID is labeled **firm>id** to emphasize that employees are nested within firms. There is more variation in the random intercept at the employee level than at the firm level.

Having estimated the model, we can now predict the random intercepts, which account for unobserved heterogeneity at each level. Here we obtain the posterior means of the random effects. Stata automatically labels the new variables with an appropriate label so that you know which random effect corresponds to which level.

variable name	storage type	display format	value label	variable label
re_firm	float	%9.0g		empirical Bayes' means for _cons[firm]
re_id	float	%9.0g		empirical Bayes' means for _cons[firm>id]
se_firm	float	%9.0g		S.E. of empirical Bayes' means for _cons[firm]
se_id	float	%9.0g		S.E. of empirical Bayes' means for _cons[firm>id]

Here we plot the 10 lowest and 10 highest random effects at the employee level. It is interesting to see how wide the confidence intervals are for the effects, even though the confidence interval for the variance of the effects suggests their variance is not zero.
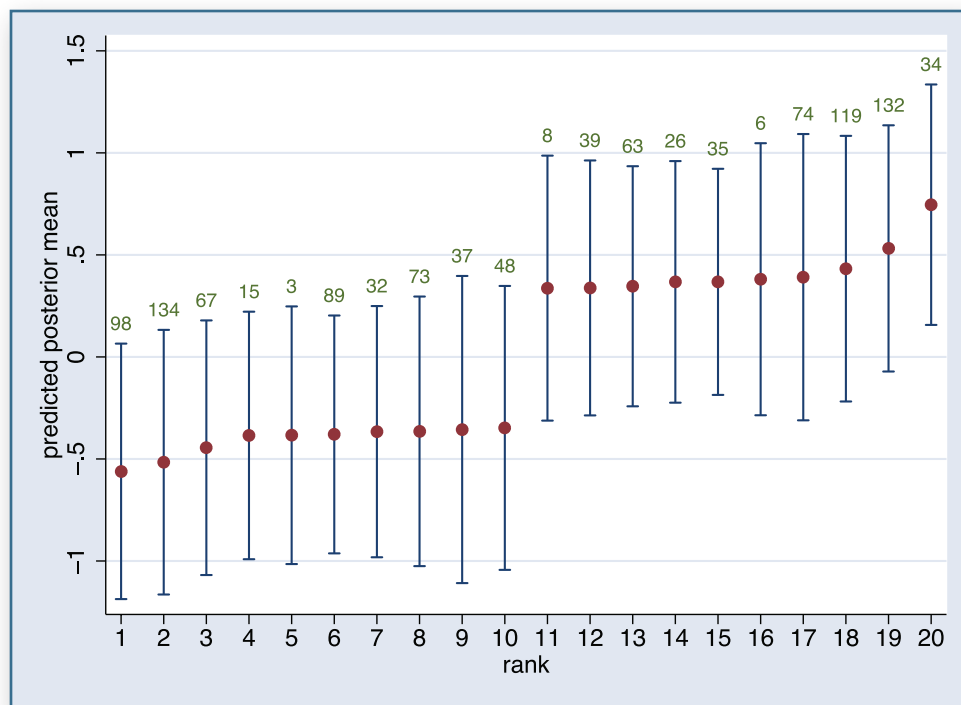
Summary

We used Stata 13's **meglm** command to illustrate a three–level ordered logistic model. Make sure to check out the new *Multilevel Mixed-Effects Reference Manual*, where we present more advanced models with random slopes, block–diagonal covariance structures, crossed random effects, and more. You can read the PDF online at **stata.com/manuals13/me**.

Read more about the features of Stata's new multilevel models at **stata.com/stata13**.

—Rafal Raciborski
Senior Statistical Developer

New from the Stata Bookstore

Applied Panel Data Analysis for Economic and Social Surveys

Authors:	Hans-Jürgen Andreß, Katrin Golsch, and Alexander W. Schmidt
Publisher:	Springer
Copyright:	2013
ISBN-13:	978-3-642-32913-5
Pages:	327; hardcover
Price:	$49.75

Applied Panel Data Analysis for Economic and Social Surveys, by Hans-Jürgen Andreß, Katrin Golsch, and Alexander W. Schmidt, provides a rigorous yet intuitive discussion of panel-data methods as they are used in the social sciences. Because the book does not use matrix algebra or advanced econometric methods, it will appeal to master's-level students and practitioners as well as doctoral students and researchers. The book covers all commonly used estimation methods, including fixed- and random-effects estimators, first-difference estimators, and maximum likelihood estimators. Separate chapters consider continuous and categorical dependent variables. In contrast to many panel-data texts, *Applied Panel Data Analysis for Economic and Social Surveys* also considers discrete-time event-study (hazard) models and their relationships to other panel-data models. All the datasets used in the book are available in Stata format as well as Stata do-files to replicate all the examples in the text.
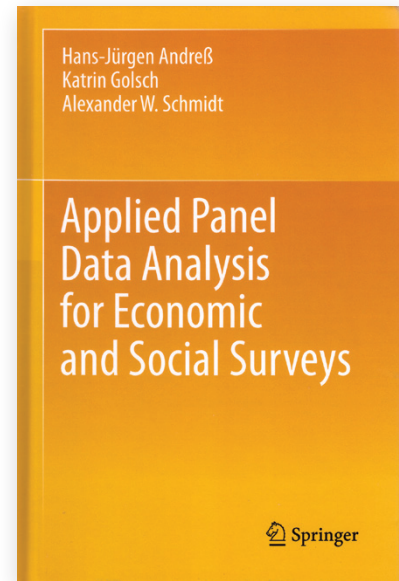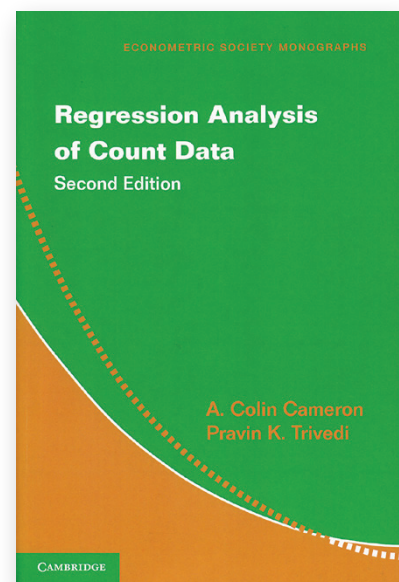
Read the table of contents or order online at **stata.com/bookstore/applied-panel-data-analysis**.

Regression Analysis of Count Data, Second Edition

Authors:	A. Colin Cameron and Pravin K. Trivedi
Publisher:	Cambridge University Press
Copyright:	2013
ISBN-13:	978-1-107-66727-3
Pages:	566; paperback
Price:	$49.75

Cameron and Trivedi's *Regression Analysis of Count Data, Second Edition* has been completely revised to reflect the latest developments in the analysis of count data. A new chapter approaches count-data modeling from a Bayesian perspective, and simulation and bootstrap methods have been incorporated into most of the chapters. Material from the first edition has also been reorganized to improve exposition and present the most popular methods earlier within each chapter. The book is written in a technical yet lucid manner so that it will be valuable both to those who are new to count-data models and to seasoned professionals looking for a concise reference.

Read the table of contents or order online at **stata.com/bookstore/regression-analysis-count-data**.

Course	Dates	Location	Cost
Using Stata Effectively	September 9–10	Washington, DC	$950
Structural Equation Modeling Using Stata	September 11–12	Washington, DC	$1,295
Programming an Estimation Command in Stata	October 22–23	Washington, DC	$1,295
Estimating Average Treatment Effects Using Stata NEW	October 24–25	Washington, DC	$1,295
Regression Modeling Using Stata NEW	November 5	Washington, DC	$695
Panel-Data Analysis Using Stata	November 6–7	Washington, DC	$1,295

Sign up for both **Regression Modeling Using Stata** and
Panel-Data Analysis Using Stata
for only $1,795 (save $195)!

Estimating Average Treatment Effects Using Stata NEW

Learn how to use observational data to estimate average treatment effects and average treatment effects on the treated in Stata.

Course topics

- The potential-outcome framework, the average treatment effect, and the average treatment effect on the treated
- Observational data differ from experimental data
- Six estimators:
  › Regression–adjustment
  › Inverse-probability-weighted (IPW)
  › Augmented IPW
  › IPW regression–adjustment
  › Nearest–neighbor matching
  › Propensity–score matching
- The double-robustness property of the augmented IPW and IPW regression-adjustment estimators
- Using different functional forms for outcome model and treatment model
- Multivalued treatments
- Some estimators when the treatment is endogenous

Regression Modeling Using Stata NEW

Regression modeling is a fundamental tool for researchers who want to establish causal quantitative relationships from observational data. This course is intended as an overview of the theoretical concepts necessary to understand regression models and how to implement them using Stata. The concepts will be reinforced with exercises that the participants will work on with the assistance of the instructor.

Course topics

- Types of data: Cross-sectional, time-series, and panel-data
- Statistical concepts: Conditional expectation, conditional variance, and conditional covariance
- Mathematical concepts: Sum of squared errors, maximum likelihood, and conditional moments
- Properties of estimators: Bias, consistency, efficiency, and normality
- Regression model and conditional expectation
- Standard errors
- Parameter interpretation
- Probit and logit models
- Endogeneity

Using Stata Effectively: Data Management, Analysis, and Graphics Fundamentals

Become intimately familiar with all three components of Stata: data management, data analysis, and graphics. Aimed at both new Stata users and those who wish to learn techniques for efficient day-to-day use of Stata, this course enables you to use Stata in a reproducible manner, making collaborative changes and follow-up analyses much simpler. The course will be interactive and use real data.

Programming an Estimation Command in Stata

Learn how to write an estimation command for Stata. No Stata or Mata programming experience is required. This course will provide an introduction to basic Stata do-file programming, basic and advanced ado-file programming, and an introduction to Mata, the byte-compiled matrix language that is part of Stata. Exercises will supplement the lectures and Stata examples.

Structural Equation Modeling Using Stata

Learn how to illustrate, specify, and estimate structural equation models in Stata, using both Stata's SEM builder and the **sem** command. The course introduces several types of models, including path analysis, confirmatory factor analysis, full structural equation models, and latent growth curves.

Panel-Data Analysis Using Stata

After introducing the fixed-effects and random-effects approaches to unobserved individual-level heterogeneity, the course covers linear models with exogenous covariates, linear models with endogenous variables, dynamic linear models, and some nonlinear models. The generalized method of moments estimation technique is also introduced.

We offer a 15% discount for group enrollments of three or more participants. Please contact us at **training@stata.com** for details.

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**.

New from the Stata Gift Shop

stata.com/giftshop

United Kingdom

Dates: September 12–13, 2013

Venue: Cass Business School
London, UK

Cost: Both days £96.00; £66.00 student
Single day £66.00; £48.00 student
Dinner £36.00

All prices include VAT.

Details: **stata.com/meeting/uk13**

Preliminary program

Thursday, September 12

Introduction and Welcome
Stephen P. Jenkins, London School of Economics

Roger B. Newson, Imperial College London

Creating factor variables in resultssets and other datasets
Roger B. Newson, Imperial College London

treatrew: A user-written Stata routine for estimating average treatment effects by reweighting on propensity score
Giovanni Cerulli, Institute for Economic Research on Firms and Growth, Rome

Multiple imputation of covariates in the presence of interactions and nonlinearities
Jonathan Bartlett, London School of Hygiene and Tropical Medicine

A Monte-Carlo analysis of multilevel binary logit model estimator performance
Stephen Jenkins, London School of Economics

Multilevel mixed effects parametric survival analysis
Michael J. Crowther, University of Leicester

Multiple imputation of missing data in longitudinal health records
Irene Petersen, UCL Department of Primary Care and Population Health

Catherine Welch, UCL Department of Primary Care and Population Health

A multiple imputation and coarse data approach to residually confounded regression models
Robert Grant, St. George's, University of London, and Kingston University

The stiqsp command: A nonparametric approach for the simultaneous analysis of quality-of-life and survival data using the integrated quality survival product
Piers Gaunt, University of Birmingham

Michael J. Crowther, University of Leicester

Lucinda Billingham, University of Birmingham

Estimating spatial panel models using unbalanced panels
Gordon Hughes, University of Edinburgh

Repeated half-sample bootstrap resampling
Phillipe Van Kerm, CEPS/INSTEAD, Luxembourg

Sample size by simulation for clinical trials with survival outcomes: The simsam package in action
Richard Hooper, Barts and The London School of Medicine & Dentistry, QMUL

Strategy and tactics for graphic multiples in Stata
Nicholas J. Cox, Durham University

A general approach to testing for autocorrelation
Christopher F. Baum, Boston College

Mark Schaffer, Heriot-Watt University

Semiparametric regression in Stata
Vincenzo Verardi, Free University of Brussels

Friday, September 13

Power of the power command
Yulia Marchenko, StataCorp

Adaptive dose-finding designs to identify multiple doses that achieve multiple response targets
Adrian Mander, MRC Unit, Cambridge

Simon Bond, Cambridge University Hospitals NHS Foundation

usecspro: Importing CSPro hierarchical datasets to Stata
Sergiy Radykin, The World Bank

From Stata to aML
Sara Ayllón, University of Girona

Two-stage individual participant data meta-analysis and forest plots
David Fisher, MRC Clinical Trials Unit

A suite of Stata programs for network meta-analysis
Ian White, MRC Biostatistics Unit, Cambridge

Path diagrams as a notational formalism—noodling around with pictures
Vince Wiggins, StataCorp

Estimating and modeling cumulative incidence functions using time-dependent weights
Paul C. Lambert, University of Leicester

Mixed logit modeling in Stata—an overview
Arne Rise Hole, University of Sheffield

Report to users, wishes and grumbles
William Gould, StataCorp

Registration

Find out more and register for the meeting by contacting Timberlake sales and training either by email at **training@timberlake.co.uk** or by phone at +44 (0) 20 8697 3377.

Scientific organizers

Stephen P. Jenkins
London School of Economics and Political Science

Roger B. Newson
Imperial College London

Logistics organizer

Timberlake Consultants

Japan

Date:	September 20, 2013
Venue:	Graduate School of Finance, Accounting, and Law, Waseda University Nihonbashi, Tokyo
Cost:	¥5,000
Details:	**stata.com/meeting/japan13**

The first Japanese Stata Users Group meeting will be held at the Graduate School of Finance, Accounting, and Law at Waseda University on Friday, September 20, 2013. The meeting is open to everyone, and StataCorp will be represented. This meeting will provide Stata users from all disciplines the opportunity to exchange ideas, experiences, and information on new applications of Stata.

The meeting will be conducted mostly in Japanese.

Program

A whirlwind tour of multilevel and mixed models in Stata *(English)*
  Bill Rising, StataCorp

Q&A

Statistical analysis of human growth pattern
  Yukinobu Kitamura, Hitotsubashi University

Estimating the term structure of default probabilities considering macroeconomic factors
  Soichiro Moridaira, Waseda University

Quantile regression of personal network size
  Itaru Ishiguro, Japan Women's University

Data management using Stata—quality of care measurement based on hospital-based cancer registry and insurance claims data
  Takahiro Higashi, National Cancer Center

Propensity-score analysis using Stata
  Osamu Takahashi, Center for Clinical Epidemiology

Report to users, wishes and grumbles *(English)*
  Bill Rising, StataCorp

Registration

Register online at **https://www.lightstone.co.jp/stata/stata_users**.

Organizer

  LightStone Corporation

Nordic and Baltic States

Date:	September 27, 2013
Venue:	Nobel Forum at Karolinska Institutet Nobelkansliet Box 270, STOCKHOLM Sweden
Cost:	Free
Details:	**stata.com/meeting/nordic-and-baltic13**

Spain

Date:	October 10, 2013
Venue:	Universidad Carlos III de Madrid Madrid, Spain 28903 Getafe
Cost:	€60 regular; €30 student
Details:	**stata.com/meeting/spain13**

Italy

Submissions for presentations are currently being accepted. The deadline is September 5.

Date:	November 14–15, 2013		
Venue:	Hotel Brunelleschi Piazza S. Elisabetta, 3 Florence, Italy		
Cost:	Meeting	€95 regular;	€71 student
	+ training	€400 regular;	€300 student
	Training only €400 regular;		€300 student
	Prices do not include local sales tax, which is 21%.		
Details:	**stata.com/meeting/italy13**		

Want to keep up with the latest developments in Stata Users Group meetings? Sign up for an email alert at **stata.com/alerts.**

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 /StataCorp   **t** /Stata   **You Tube** /StataCorp   **g+** /+Stata   **b** blog.stata.com

Free webinar: An overview of new treatment-effects estimators in Stata 13

October 3, 2013, at 4:00 p.m. EST

Join us for a demonstration of Stata 13's new treatment-effects estimators. Learn how to use the new **teffects** command to estimate average treatment effects (ATEs), ATEs on the treated, and potential-outcome means using regression adjustment, inverse-probability weighting, covariate and propensity-score matching, and doubly robust methods. We will discuss continuous, binary, and count outcomes, as well as, both binary and multivalued treatments. StataCorp developers will be available for questions.

All you need is a web browser and an Internet connection. To register, visit stata.com/webinar/treatment-effects.