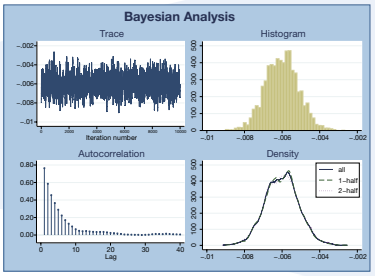
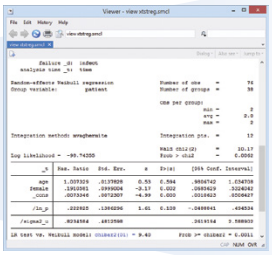
Meetings Courses Seminars Tips Conferences
Books Publications Training Software & more
Apr/May/Jun
Vol 30 No 2
The Stata® News
Statistics Graphics Data Management & Analysis

Announcing Stata 14

Bayesian analysis has come to Stata. Fit models using a Metropolis–Hastings algorithm, diagnose convergence, analyze posterior distributions, perform inference, and much, much more.

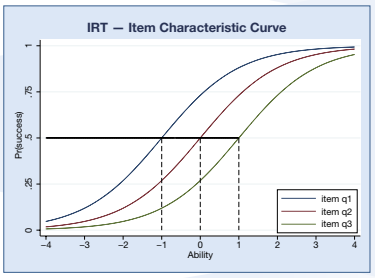


Unicode. Здравствуйте. こんにちは. Hello. Use Unicode for variable names, labels, data, and whatever else you wish.



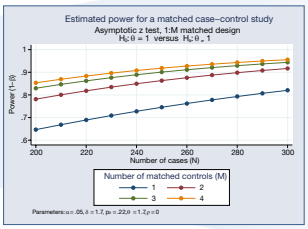
Panel and multilevel survival models let you introduce normally distributed heterogeneity into duration analyses. Random intercepts. Random coefficients. Crossed effects. Two-, three-, and higher-level models. And more.

Much more in treatment effects, including endogenous treatments, survival models, sampling (probability) weights, and balance analysis.



IRT (item response theory) lets you explore the relationship between a latent trait and the items that measure aspects of that trait.

Markov-switching regression models to analyze time series that transition over a set of unobserved regimes (states).

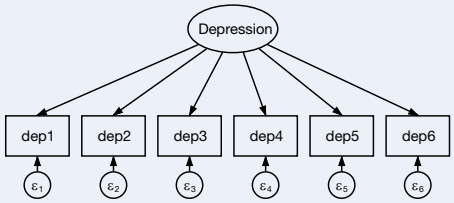


Power analysis for contingency tables and survival models lets you tabulate and graph sample size, power, and effect size for Cox models, Cochran–Mantel–Haenszel tests, and more.

Survey support for multilevel models. Just **svyset** your sampling design and put **svy:** in front of any multilevel model.

And much more:

Better integration with Excel • Fractional outcome models • Hurdle models • Censored Poisson models • Beta regression models • Denominator degrees of freedom for mixed models • Satorra–Bentler adjustments for SEM • Structural break tests • Marginal predictions for SEM and multilevel models • ICD-10 support • Interface in Spanish and Japanese • More than 2 billion observations • And even more



Have you upgraded? Visit stata.com/stata14.

The Stata News
Executive Editor.....Karen Strope
Production Supervisor .....Annette Fett

In the spotlight: Bayesian "random-effects" models

Stata 14 introduced **bayesmh** for fitting Bayesian models. You can choose from one of many built-in models or write your own. See **stata.com/stata14/bayesian-analysis** and **stata.com/stata14/bayesian-evaluators** for details.

In this article, we show you how to use **bayesmh** to fit a Bayesian "random-effects" model. We write "random effects" in quotes because all effects (parameters) are considered random within the Bayesian framework. These models are typically referred to as Bayesian multilevel or Bayesian hierarchical models.

We revisit, using the Bayesian approach, the random-effects meta-analysis model described in example 6 of [ME] **me**. The term "meta-analysis" refers to a statistical analysis that involves summarizing results from similar but independent studies.

We consider data from Turner et al. (2000) that contain estimates of treatment effects expressed as log odds-ratios (**lnOR**) and their respective variances (**var**) from nine clinical trials that examined the effect of taking diuretics during pregnancy on the risk of preeclampsia. Negative **lnOR** values indicate that taking diuretics lowers the risk of preeclampsia. The model can be written as

$$\begin{aligned} y_i &\sim N(\mu_i, \sigma_i^2) \\ \mu_i &\sim N(\theta, \tau^2) \end{aligned} \qquad (1)$$

where μ_i is the mean treatment effect of each trial, θ is an overall mean, σ_i^2 is the variance of the observed treatment effect and is considered fixed, and τ^2 is the between-trial variance. θ and τ^2 are parameters of interest—τ^2 close to 0 would suggest homogeneity across studies in log odds-ratio estimates.

In example 6 of [ME] **me**, we fit this random-effects model using **meglm** and obtain the estimates of θ and τ^2 of –0.52 and 0.24 with their respective 95% confidence intervals of [–0.92, –0.11] and [0.048, 1.19].

For our Bayesian analysis, we need to additionally specify priors for θ and τ^2 in model (1). Notice that the random-effects model (1) already assumed a normal prior for each individual trial effect μ_i. We consider fairly noninformative normal and inverse gamma prior distributions for the parameters θ and τ^2.

$$\begin{aligned} \theta &\sim \text{Normal}(0,10000) \\ \tau^2 &\sim \text{IG}(0.0001, 0.0001) \end{aligned} \qquad (2)$$

To fit model (1)–(2) using **bayesmh**, we type

```
. fvset base none trial
. bayesmh lnOR i.trial, noconstant likelihood(normal(var))
    prior({lnOR:i.trial}, normal({theta},{tau2}))
    prior({theta}, normal(0,10000))
    prior({tau2}, igamma(0.0001,0.0001))
    block({lnOR:i.trial}, split) block({theta}, gibbs)
    block({tau2}, gibbs) dots
```

The first four lines of the **bayesmh** specification are the straightforward model specification. The last two lines improve the efficiency of the sampler—a step that is crucial when fitting high-dimensional hierarchical models such as random-effects models. We sample all parameters separately by placing them in individual blocks and specifying option **split** for levels of **trial**, and we also request Gibbs sampling for θ and τ^2.

Because this is a random-effects example, we feel that a note of caution is in order. **bayesmh** uses an adaptive random-walk Metropolis–Hastings algorithm to fit all models (with potential Gibbs updates for some parameters). This algorithm provides an exceptionally flexible framework for fitting any Bayesian model. That flexibility comes with a price. The algorithm becomes inefficient for models with many parameters and may become prohibitively slow for some models. So, in theory, you can specify models with any number of levels of hierarchy and with any number of effects within a hierarchy. In practice, however, models with many random effects or many hierarchical levels may be infeasible. This poses no problem for our example with nine effects, but would likely be problematic for problems with thousands of random effects.

Here is the output:

Viewer - view bayes1.smcl

File Edit History Help

view bayes1.smcl

Dialog Also see Jump to

```
Model summary
------------------------------------------------------------------------------
Likelihood:
  lnOR ~ normal(xb_lnOR,var)

Prior:
  {lnOR:i.trial} ~ normal({theta},{tau2})                            (1)

Hyperpriors:
  {theta} ~ normal(0,10000)
  {tau2} ~ igamma(0.0001,0.0001)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_lnOR.

Bayesian normal regression                      MCMC iterations  =     12,500
Metropolis-Hastings and Gibbs sampling          Burn-in          =      2,500
                                                MCMC sample size =     10,000
                                                Number of obs    =          9
                                                Acceptance rate  =      .5267
                                                Efficiency:  min =     .05966
                                                             avg =      .1284
Log marginal likelihood = -10.005871                         max =      .2683
```

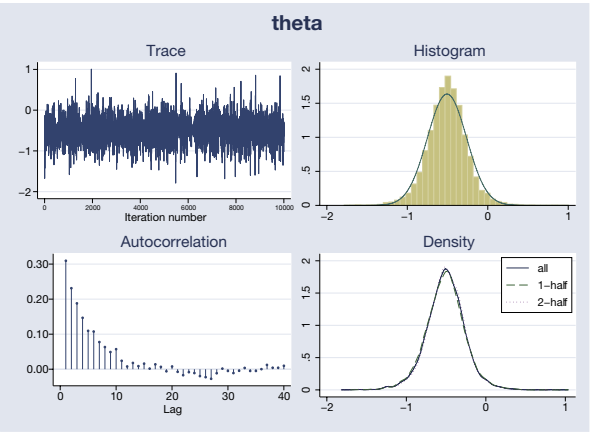
	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred. Interval]	
lnOR						
trial						
1	-.1725911	.3411875	.008538	-.1960632	-.830149	.52997
2	-.7802188	.3054875	.009207	-.7652055	-1.406827	-.2325144
3	-.848563	.354147	.010496	-.8291194	-1.593789	-.2051206
4	-.9614017	.4446485	.01509	-.9214434	-1.898545	-.2247392
5	-1.090177	.3241574	.013271	-1.091318	-1.749594	-.4724761
6	-.3107003	.0983935	.002373	-.3104993	-.5097513	-.1120432
7	-.3335097	.289512	.00702	-.3430311	-.8841905	.2680654
8	-.0148314	.5472792	.018604	-.086212	-.927254	1.233408
9	-.0232332	.2550613	.009728	-.0318037	-.5020652	.4952697
theta	-.5065265	.2438874	.004708	-.503503	-1.008817	-.0171035
tau2	.3841789	.4098976	.012025	.2703921	.0192085	1.412875

CAP NUM OVR

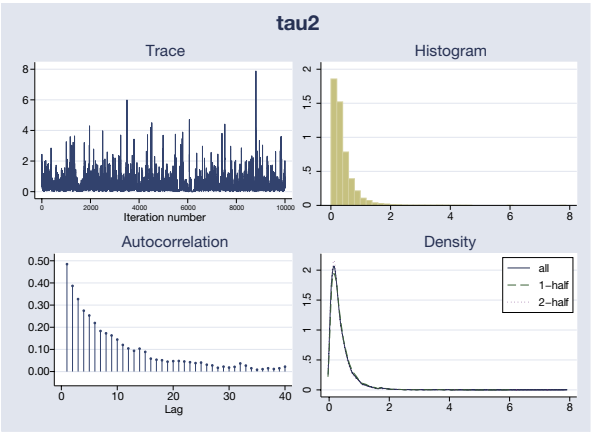
Because we used noninformative priors, our Bayesian results are similar to the frequentist results from example 6. For example, the posterior mean estimate of the overall mean θ is –0.51 with a 95% credible interval of [–1.01, –0.017] that represents the ranges to which θ belongs with a probability of 0.95.

We use **bayesgraph** to evaluate MCMC convergence visually. For example, here are diagnostics for θ and τ^2:

```
. bayesgraph diagnostics {theta}, histopts(normal)
```




```
. bayesgraph diagnostics {tau2}
```

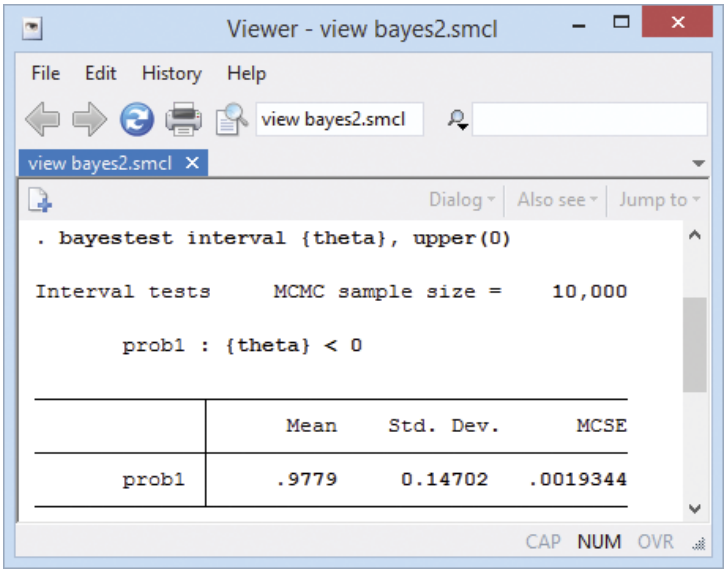


Our visual diagnostics raise no concern for these parameters. We can also check MCMC convergence for trial-specific μ_i's by typing

```
. bayesgraph diagnostics {lnOR:i.trial}
```

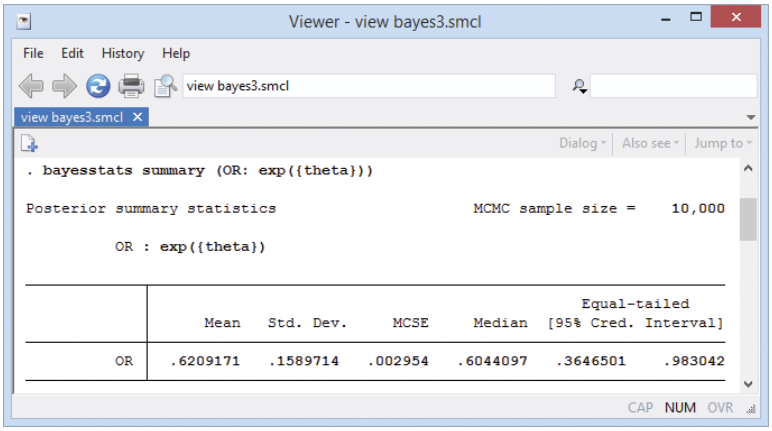
(output omitted)

We can test whether taking diuretics reduces the risk of preeclampsia overall by computing the probability that θ is negative.



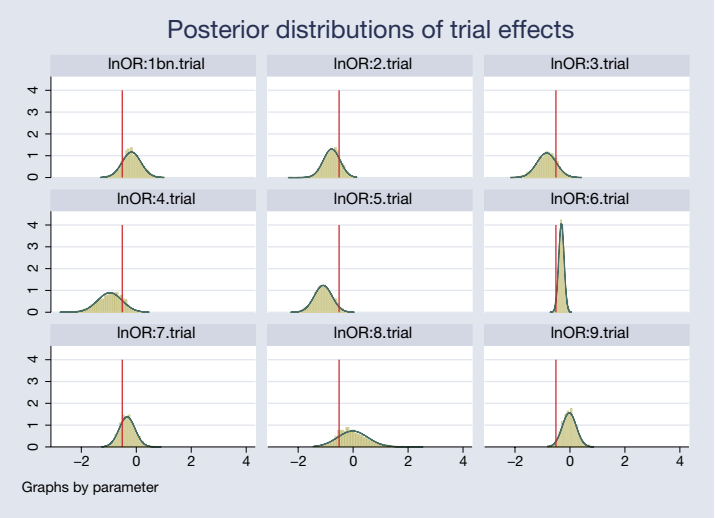
That probability is 0.98.

If desired, we can also compute an estimate of the overall odds ratio.



A neat feature of our Bayesian analysis is that we can explore the distributions of the estimated parameters. For example, we can look at distributions of effects from individual trials.

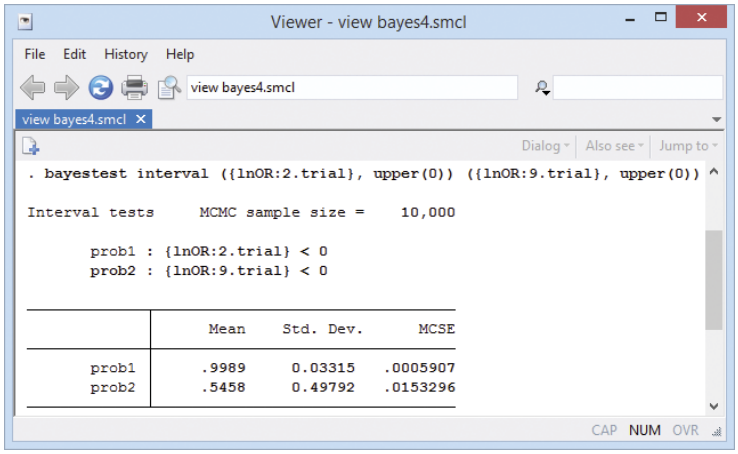
```
. bayesgraph histogram {lnOR:i.trial},
    byparm(legend(off) noxrescale noyrescale
    title(Posterior distributions of trial effects))
    normal xline(-0.51)
```



The posterior distributions are fairly normal for most trials. Posteriors for some trials are more closely centered on the overall mean of –0.51, particularly for trials 2 and 7.

There is a noticeable variability in the estimated treatment effects between trials. Trials 6 and 9 are more precise compared with other trials.

We can also test for an effect in each trial. We can estimate a probability that an effect is negative (meaning diuretics work) for each trial. For example,



The probability of a negative treatment effect is almost 1 for trial 2 and only about 0.55 for trial 9.

Reference

Turner, R. M., R. Z. Omar, M.Yang, H. Goldstein, and S. G.Thompson. 2000. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* 19: 3417–3432.

— Yulia Marchenko
Director of Biostatistics, StataCorp

— Nikolay Balov
Senior Statistician and Software Developer, StataCorp

2015 International Stata Users Group meetings

Germany

Nuremberg June 26, 2015
stata.com/meeting/germany15

Portugal

Lisbon September 18, 2015
stata.com/meeting/portugal15

Japan

Tokyo August 28, 2015
stata.com/meeting/japan15

Australia

Canberra September 24–25, 2015
stata.com/meeting/australia15

Sweden

Stockholm September 4, 2015
stata.com/meeting/nordic-and-baltic15

UK

London September 10–11, 2015
stata.com/meeting/uk15

Italy

Florence November 12–13, 2015
stata.com/meeting/italy15

Keep up with future Stata Users Group meetings. We post our schedule at **stata.com/meeting**. Want to be notified when new meeting information is posted? Go to **stata.com/alerts** and sign up for an email alert today.

In the spotlight: Formatting Excel® tables from within Stata

New capabilities of **putexcel** in Stata 14 let you format cells and fonts in Excel. This means that you can now easily export matrices, expressions, and stored results from Stata to Excel and control how they appear. You can also insert Stata graphs and add cell formulas. Now it's easy to format the cells to create custom tables or reports. Let's see how it's done.

Suppose I am estimating a fractional-response model for 401(k) participation and want to create a table of my results that I could include in a presentation or publication. A short description of the variables is shown to the right.

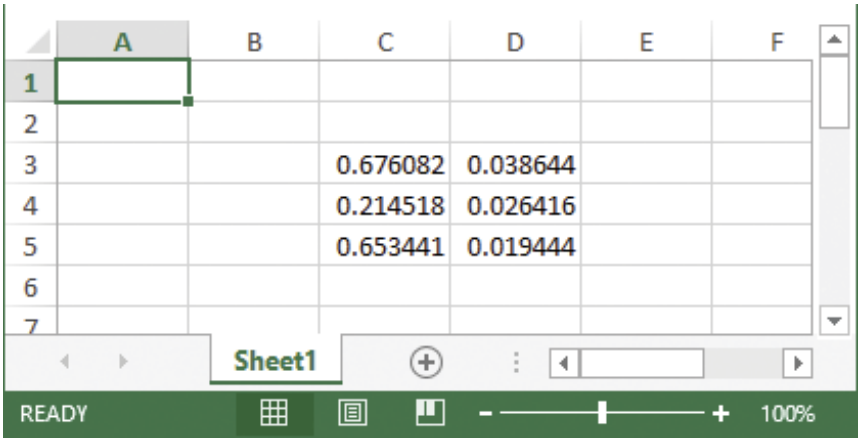
Variable	Description
prate	participation rate
mrate	plan match rate, per $
sole	=1 if only retirement plan

```
. webuse 401k
. fracreg probit prate mrate sole
```
(output omitted)

After I fit my model, I want to write out the estimated coefficients and standard errors to Excel. Most estimation commands in Stata store results in the **r(table)** matrix, giving you an easy way to access estimated coefficients, standard errors, test statistics, and the like from a single stored result. All you need to do is extract the appropriate rows and columns from the matrix.

I do this by typing

```
. matrix a = r(table)
. matrix b = a[1 ..2,1 ...]'
. putexcel C3=matrix(b) using 401k_report.xlsx replace
```



But my table is not quite publication-ready yet.

I can add bold column titles, "Estimate" and "S.E.", and a cell border between the title and results to the worksheet by typing

```
. putexcel C2=("Estimate") D2=("S.E.") (C2:D2)=bold("on")
          (B2:D2)=border("bottom", "thin") using 401k_report.xlsx, modify
```

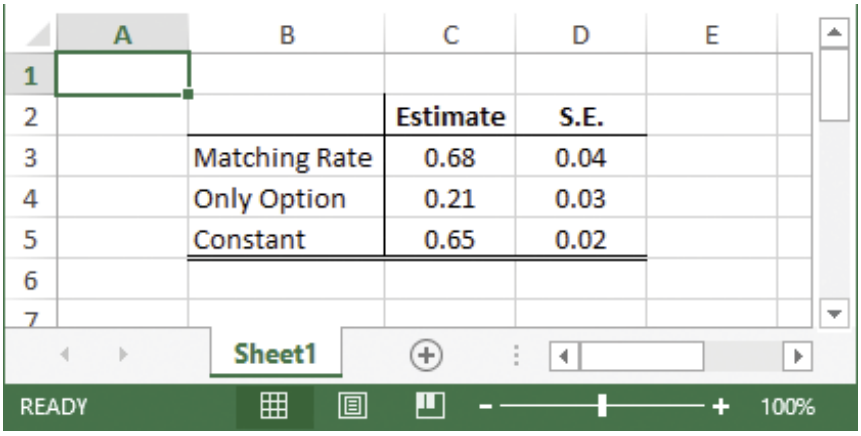
To add row labels and a cell border between the label and results, I type

```
. putexcel B3=("Matching Rate") B4=("Only Option") B5=("Constant")
          (B2:B5)=border("right", "thin") using 401k_report.xlsx, modify
. putexcel (B5:D5)=border("bottom", "double") using 401k_report.xlsx, modify
```

Now, I want to center the column text and results and format the numeric output to two decimals:

```
. putexcel (C3:D5)=nformat("number_sep_d2") (C2:D5)=halign("center")
           using 401k_report.xlsx, modify
```

I now have a nice-looking table.



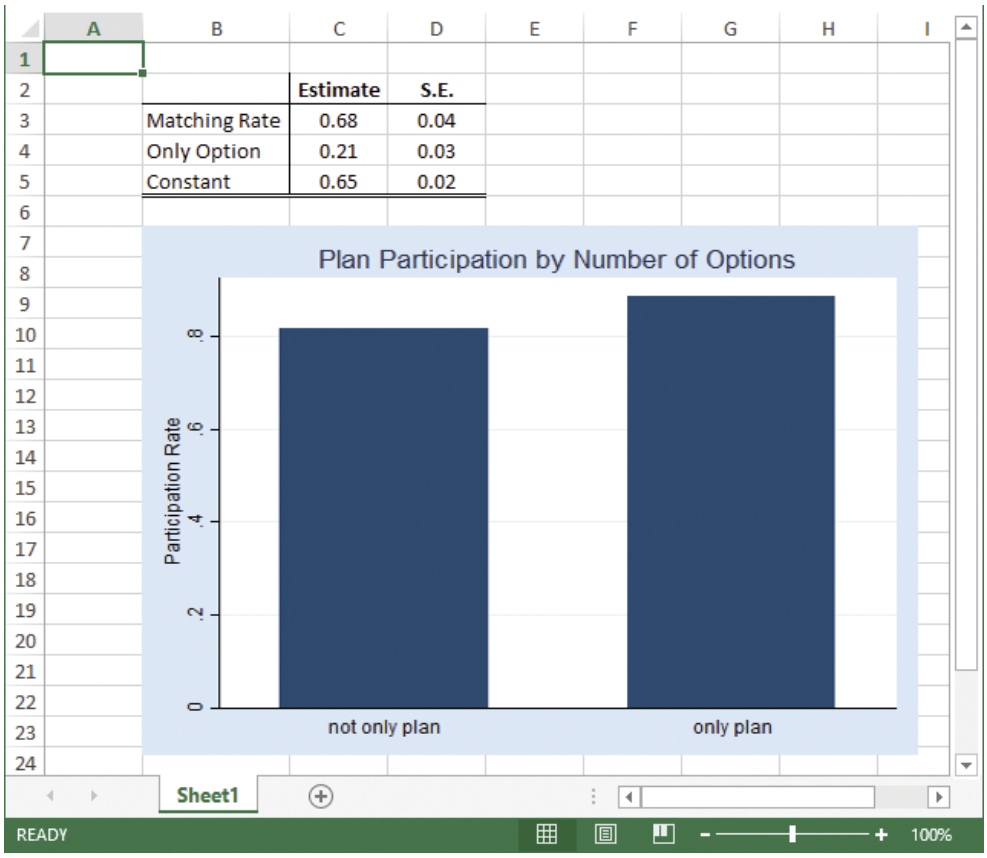
	Estimate	S.E.
Matching Rate	0.68	0.04
Only Option	0.21	0.03
Constant	0.65	0.02

But, **putexcel** can do even more. You can also write a wide range of graph file formats, including PNG, JPEG, WMF, and TIFF.

For example, I can add a bar graph of the average participation rate for companies over whether the 401(k) is the only retirement option offered by typing

```
. graph bar (mean) prate, over(sole) ytitle("Participation Rate")
        title("Plan Participation by Number of Options")
. graph export bar1.png
. putexcel (C7)=picture("bar1.png") using 401k_report.xlsx, modify
```

My worksheet now looks like this:



	Estimate	S.E.
Matching Rate	0.68	0.04
Only Option	0.21	0.03
Constant	0.65	0.02

There are many more formats and settings that you can change in Excel by using **putexcel**. To view a list, see [P] **putexcel**.

— Kevin Crow
Senior Software Developer, StataCorp

STATA® CONFERENCE COLUMBUS

Come join us in Columbus, home to the state capital of Ohio and Ohio State University, for two days of networking and Stata exploration. The conference includes, in addition to user contributions, presentations by StataCorp developers on new Stata 14 features.

Don't miss this opportunity to connect with fellow researchers as well as developers who wrote Stata 14!

When	July 30–31, 2015
Where	Hyatt Regency Columbus 350 North High Street Columbus, Ohio
Who	Stata developers You and Stata users from around the world
Cost	$195 two days, $75 student $125 one day, $50 student Optional dinner TBA

Preliminary program

Estimating treatment effects for ordered outcomes using maximum simulated likelihood
Christian Gregory
Economic Research Service, USDA

Linear dynamic panel-data estimation using maximum likelihood and structural equation modeling
Richard Williams
Department of Sociology, University of Notre Dame

Paul Allison
Department of Sociology, University of Pennsylvania

Enrique Moral Benito
Banco de España Madrid

15 years a consultant
Phil Ender
UCLA Statistical Consulting Group (Ret)

Robust inference in regression-discontinuity designs
Matias Cattaneo
University of Michigan

Sebastian Calonico
University of Michigan

Rocio Titiunik
University of Michigan

Estimation in panel data with individual effects and AR(p) remainder disturbances
Long Liu
Department of Economics, The University of Texas at San Antonio

Item response theory models in Stata
Rebecca Pope
Health Econometrician, StataCorp

Meta-analysis on the effects of interviewer supportiveness on the accuracy of children's reports
Christine Wells
Statistical Consulting Group, UCLA

Karen Saywitz, PhD
UCLA

Rakel Larson, MA
University of California, Riverside

Sue Hobbs, PhD
University of California, Davis

tetrad: A program for confirmatory tetrad analysis
Shawn Bauldry
University of Alabama at Birmingham

Kenneth Bollen
University of North Carolina at Chapel Hill

Postestimation parameter recentering and rescaling
Douglas Hemken
Social Science Computing Cooperative, University of Wisconsin–Madison

Statistical process control charts
Barbara Williams
Virginia Mason Medical Center

Data workflows with Stata and Python
Stephen Childs
Education Policy Research Initiative, University of Ottawa

Dejan Pavlic
Education Policy Research Initiative

Distribution-free estimation of heteroskedastic binary-response models in Stata
Jason Blevins
Department of Economics, The Ohio State University

Shakeeb Khan
Duke University

A comparison of modeling scales in flexible parametric models
Noori Akhtar-Danesh
McMaster University

Estimating Markov-switching regression models in Stata
Ashish Rajbhandari
Senior Econometrician, StataCorp

midasinla: midas goes Bayesian via R-INLA
Ben Adarkwa Dwamena
University of Michigan Medical School

Between and beyond: Irregular series, interpolation, variograms, and smoothing
Nicholas Cox
Department of Geography, Durham University

Public program sensitivity: Using ROC curves to characterize classification efficiency of state Medicaid systems
Lisa Frazier
John Glenn College of Public Affairs, The Ohio State University

Small-sample inference for linear mixed-effects models
Xiao Yang
Senior Statistician and Software Developer, StataCorp

Development of a project-based statistics course for applied biostatistics using Stata
Frank Snyder
Purdue University

Brewing color schemes in Stata: Making it easier for end users to customize Stata graphs
William Buchanan
Mississippi Department of Education

Colombian industrial structure behavior and its regions between 1974 and 2005
Luis Fernando López Pineda
Chamber of Commerce of Cartagena

Accommodations

The Hyatt Regency Columbus is offering a special rate of $169 per night for Stata Conference attendees staying between July 28 and August 1, 2015. Book by July 7 to receive the special rate.

Hyatt Regency Columbus
350 North High Street
Columbus, Ohio 43215
(614) 463-1234
resweb.passkey.com/go/statacorp15

See you in Columbus!

Scientific committee

- Stanley Lemeshow (chair of review team)
  The Ohio State University
  Biostatistics
- Timothy R. Sahr (coordinator)
  The Ohio Colleges of Medicine Government Resource Center
  Applied Research
- Marcus Berzofsky
  RTI, International
  Survey Research
- Christopher Browning
  The Ohio State University
  Sociology
- Anand Desai
  The Ohio State University
  Public Policy
- Christopher Holloman
  The Ohio State University
  Statistics
- Bo Lu
  The Ohio State University
  Biostatistics
- Eric Seiber
  The Ohio State University
  Health Economics

NetCourses®

NetCourses are convenient web-based courses that teach you how to exploit the full power of Stata. Learn Stata from the comfort of your own home or office!

Introduction to Stata

Dates: July 10–August 21, 2015
Cost: $95

Introduction to Stata Programming

Dates: July 10–August 21, 2015
Cost: $150

Advanced Stata Programming

Dates: June 12–July 31, 2015
Cost: $175

Introduction to Survival Analysis Using Stata

Dates: June 12–July 31, 2015
Cost: $295

Introduction to Univariate Time Series with Stata

Dates: June 12–July 31, 2015
Cost: $295

Introduction to Panel Data Using Stata

Dates: June 19–July 31, 2015
Cost: $295

For course details or to enroll, visit **stata.com/netcourse**.

The dates above don't work for you? No problem! NetCourseNow allows you to set the time and work at your own pace as well. It also gives you a personal NetCourse instructor to guide you through the course. Visit **stata.com/netcourse/ncnow**.

MEAFA Professional Development Workshop on Quantitative Analysis Using Stata

Venue: University of Sydney
Sydney, Australia
Dates: June 29–July 3, 2015

MEAFA is offering four workshops on using Stata. Workshop topics are as follows:

- **Working efficiently with Stata**
  Instructor: Demetris Christodoulou, MEAFA General Convenor
  June 29, 2015
- **Introduction to programming**
  Instructor: Demetris Christodoulou, MEAFA General Convenor
  June 30, 2015
- **Treatment effects**
  Instructor: David Drukker, Director of Econometrics StataCorp LP
  July 1-2, 2015
- **Programming estimation commands**
  Instructor: David Drukker, Director of Econometrics StataCorp LP
  July 3, 2015

You can attend any single day or combination of days. Spaces are limited. For more information, visit **stata.com/news/meafa2015**.

International short courses

Looking for a Stata course in your area? Short courses are held by a variety of institutions to help people learn more about statistics and Stata. The short courses on our website are offered by institutions other than StataCorp and provide a multitude of ways to become proficient at Stata quickly. For more information, visit **stata.com/meeting/short-courses**.

Public training courses

Sign up for our summer courses in Washington, DC! Enrollment is limited, so register today.

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

May 20–21, 2015
June 9–10, 2015

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. Exercises supplement the lessons and Stata examples.

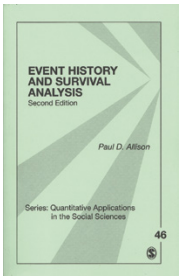
Structural Equation Modeling Using Stata

June 11–12, 2015

Learn how to illustrate, specify, and estimate structural equation models in Stata using 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. Exercises supplement the lessons and Stata examples.

Enroll online at **stata.com/training/public**.

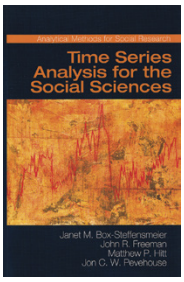
New from the Stata Bookstore



Event History and Survival Analysis, Second Edition

Author: Paul D. Allison
Publisher: Sage
Copyright: 2014
Price: $17.75

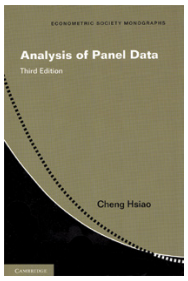
stata.com/bookstore/event-history-and-survival-analysis



Time Series Analysis for the Social Sciences

Authors: Janet M. Box-Steffensmeier, John R. Freeman, Matthew P. Hitt, and Jon C. W. Pevehouse
Publisher: Cambridge University Press
Copyright: 2014
Price: $29.75

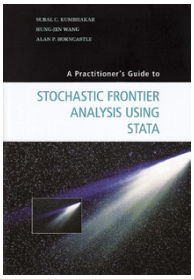
stata.com/bookstore/time-series-analysis-social-sciences



Analysis of Panel Data, Third Edition

Author: Cheng Hsiao
Publisher: Cambridge University Press
Copyright: 2014
Price: $39.75

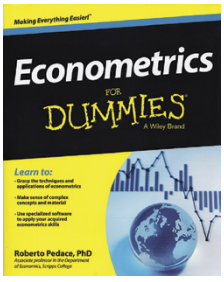
stata.com/bookstore/analysis-panel-data



A Practitioner's Guide to Stochastic Frontier Analysis Using Stata

Authors: Subal C. Kumbhakar, Hung-Jen Wang, and Alan P. Horncastle
Publisher: Cambridge University Press
Copyright: 2015
Price: $42.50

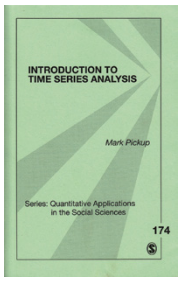
stata.com/bookstore/practitioners-guide-stochastic-frontier-analysis



Econometrics for Dummies

Author: Roberto Pedace
Publisher: Wiley
Copyright: 2013
Price: $21.50

stata.com/bookstore/econometrics-for-dummies



Introduction to Time Series Analysis

Author: Mark Pickup
Publisher: Sage
Copyright: 2015
Price: $18.75

stata.com/bookstore/introduction-time-series-analysis

Testing for endogeneity: New feature for eteffects in Stata 14

There has been great interest in Stata 14's **eteffects**, which obtains treatment effects when unobserved variables affect both treatment assignment and outcomes.

If an unobserved variable affects treatment and outcome, we have an endogeneity problem and cannot obtain accurate estimates of effects using conventional treatment-effects estimators.

We have enhanced **eteffects**. We can now test if the unobserved variable affects both outcome and treatment.

How?

eteffects estimates the correlation between the unobservables that affect treatment and outcome. If these correlations are 0, we have no endogeneity. We test this hypothesis by typing

```
. estat endogenous
```

It is that simple.

Learn more at **stata.com/stata14/endogenous-treatment-effects**, or see [TE] **eteffects**.

New from the Stata Gift Shop

Mug: $6.50

T-shirt: $12.00

stata.com/giftshop

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