Meetings Courses Seminars Tips Conferences
Books Publications Training Software & More
Apr/May/Jun
Vol 31 No 2
The Stata® News
Statistics Graphics Data Management & Analysis

In the spotlight: eteffects and the challenges of making causal inferences

The challenge

Extracting causal relationships from data is one of the fundamental endeavors of researchers. Ideally, we could conduct a controlled experiment to extract causal relations. However, a controlled experiment is rarely feasible for researchers or individuals who need to make informed decisions based on their available observational data.

In the absence of experimental data, we construct models to capture the relevant features of the causal relationship we are interested in. This is the purview of everything in Stata's [TE] *Treatment Effects Reference Manual*.

The estimators in **teffects** help us obtain estimates of the effect of a treatment (for example, a job training program or an increase in out-of-pocket contributions for a health plan) on an outcome (for example, probability of employment or enrollment in a health plan). With these traditional treatment-effect models, assignment to a treatment must be independent of the outcome to interpret our results causally. What if that is not true? What if, for example, the individuals who participate in a job training program are highly motivated? Then the outcome of a higher probability of employment might be correlated to the person's inherent motivation rather than participation in the job training program.

Stata's endogenous treatment-effects command, **eteffects**, is designed for such cases. The key assumption behind the model is that treatment assignment is not independent of outcomes because the unobservables that affect treatment assignment and outcomes are correlated. If we incorporate this correlation into our model, we can obtain a causal effect. Using this model, we can also test whether the correlation is statistically significant; in other words, we can test for endogeneity.

The tool at work

Say we are interested in the effect of attending a private high school (**private**) on college grade point average (**gpa**). We conjecture that the quality of the available private schools affects the decision of parents to send their kids there and affects the college GPA. If this is the case, **eteffects** is a good alternative.

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

We model **private** as a function of parental income (**income**), whether the student lived in an urban area (**urban**), and, because of the prevalence of Catholic private schools, whether the student's parents are Catholic (**catholic**). We model **gpa** as a function of high school GPA (**hgpa**) and the parents' joint educational attainment (**pedu**).

```
view eteffects1.smcl

. eteffects (gpa hgpa pedu) (private i.catholic income i.urban)

Iteration 0:   EE criterion =  8.024e-23
Iteration 1:   EE criterion =  3.489e-31

Endogenous treatment-effects estimation         Number of obs     =     10,000
Outcome model   : linear
Treatment model: probit

------------------------------------------------------------------------------
             |               Robust
         gpa |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ATE          |
     private |
   (1 vs 0)  |  -.0419173   .0663428    -0.63   0.527    -.1719468    .0881121
-------------+----------------------------------------------------------------
POmean       |
     private |
          0  |   3.108173   .0111613   278.48   0.000     3.086297    3.130048
```

Our estimates show no evidence that the average treatment effect of attending private school is not zero. That is, in terms of college GPA, we have no evidence of differences between all kids attending a private school versus all kids attending a public school.

We can also test the assumption that the unobservables that affect the treatment assignment also affect the outcome.

```
view eteffects2.smcl

. estat endogenous

  Test of endogeneity
  Ho: treatment and outcome unobservables are uncorrelated

           chi2(  2) =   114.21
         Prob > chi2 =     0.0000
```

In this case, we reject the null hypothesis. We have strong evidence that the unobservables are correlated.

What if we had assumed that the unobservables that affect treatment assignment do not affect the outcome? We could have used one of the **teffects** estimators, say, the inverse-probability-weighted regression adjustment. We would get the following:

```
view eteffects3.smcl

. teffects ipwra (gpa hgpa pedu) (private i.catholic income i.urban)

Iteration 0:   EE criterion =  1.710e-24
Iteration 1:   EE criterion =  4.338e-34

Treatment-effects estimation                    Number of obs     =     10,000
Estimator      : IPW regression adjustment
Outcome model  : linear
Treatment model: logit

------------------------------------------------------------------------------
             |               Robust
         gpa |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ATE          |
     private |
   (1 vs 0)  |   .6292526   .0048564   129.57   0.000     .6197341    .6387711
-------------+----------------------------------------------------------------
POmean       |
     private |
          0  |   3.017993   .0029358  1027.99   0.000     3.012239    3.023747
```

This suggests that the average treatment effect of attending private school is an increase in GPA of 0.63, which is a very different conclusion than we reached when we accounted for the unobservables affecting both GPA and choice to attend private school.

Closing remarks

The example above uses artificial data. I know that the assumptions necessary for using **eteffects** are met and that the true average treatment effect should be exactly 0. Obviously, researchers face a much more daunting challenge to ascertain causality, but **eteffects** is a valuable tool in this endeavor.

—Enrique Pinzon
Senior Econometrician, StataCorp LP

In the spotlight: Intraclass correlations after multilevel survival models

As analysts, we like our data to be independent. Sometimes our data don't cooperate. Sometimes our data have groupings such that the observations within the groups are unlikely to be independent. For instance, you may have patients who were treated at the same hospital or students who attend the same school. One way to account for this lack of independence is to fit a multilevel survival model and include group-level random effects. Stata's **mestreg** command that was introduced in Stata 14 does just that.

Survival models are usually formulated using the proportional-hazards parameterization or using the accelerated failure-time (AFT) parameterization, both of which are available with **mestreg**. One of the advantages of the AFT metric, when fitting multilevel models, is that the variances of the random effects at different grouping levels can be interpreted as variance components of the log-time. This is possible because, in the AFT metric, we model the log-time as a linear combination of random effects (and covariates). We can use those variances and the variance of the residuals to compute intraclass correlations for the log-time.

To demonstrate, I use an example based on the study of the survival probability of scrub-jay birds from Fox et al. (2006). The authors used a dataset on survival of scrub-jays collected over 35 years, where several territories were censused periodically, and data on birth cohort (year), territory, and parents were recorded, among other variables. Birds tend to stay in a territory, so mother and father have a constant territory throughout the years, although mating couples vary among years.

Fox et al. (2006) fit several two-level Weibull models to the data with fixed effects for the cohort and with random effects on territory, maternal family, or paternal family. They didn't fit multilevel models with more than two levels, perhaps because of the software limitations at that time. I use a fictional dataset with only five cohorts to demonstrate several multilevel models we could fit to the data. Observations are censored, assuming that many birds were alive at the end of this fictional five-year study.

We use **stset** to declare our survival-time data: **time** records time-to-death (or time-to-censoring) of birds in years, and **death** is a failure indicator of birds' death or censoring.

```
view me1.smcl

. stset time, fail(death)

     failure event:  death != 0 & death < .
obs. time interval:  (0, time]
 exit on or before:  failure

------------------------------------------------------------------------------
       5245  total observations
          0  exclusions
------------------------------------------------------------------------------
       5245  observations remaining, representing
       3490  failures in single-record/single-failure data
   36898.33  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =        13
```

First, let's fit a Weibull model with dummy variables for each of the five cohorts (years) and random effects on territory.

```
view me2.smcl

. mestreg  i.year || territory: ,distribution(weibull) time

        failure _d:  death
  analysis time _t:  time

Mixed-effects Weibull regression -- AFT         Number of obs     =      5,245
Group variable:        territory                Number of groups  =         20

                                                Obs per group:
                                                              min =        216
                                                              avg =      262.3
                                                              max =        296

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(4)      =    2250.67
Log likelihood = -6914.0566                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |
        1971 |   1.540283   .0490673    31.39   0.000     1.444113    1.636453
        1972 |   1.647524   .0496067    33.21   0.000     1.550297    1.744751
        1973 |   1.080513   .0463789    23.30   0.000     .9896121    1.171414
        1974 |   2.578857   .0617178    41.78   0.000     2.457893    2.699822
             |
       _cons |   .8818091   .1725921     5.11   0.000     .5435348    1.220084
-------------+----------------------------------------------------------------
       /ln_p |   .0342417   .0140015     2.45   0.014     .0067993     .061684
-------------+----------------------------------------------------------------
territory    |
   var(_cons)|   .5767268     .184568                      .3080087    1.079885
------------------------------------------------------------------------------
LR test vs. Weibull model: chibar2(01) = 1675.05      Prob >= chibar2 = 0.0000
```

The main part of the table shows the coefficients for the fixed part of the model. The baseline cohort (reference category) is 1970. The mean survival time, assuming zero random effects, can be computed with **predict, mean conditional(fixedonly)**.

```
view me3.smcl

. predict m, mean conditional(fixedonly)

. summ m if year ==1970

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
           m |      1,039    2.382012           0   2.382012   2.382012
```

The mean survival for birds in the baseline cohort (assuming zero random effects) is 2.38 years.

Based on our estimate of the coefficient for the dummy for year 1971, the mean survival of birds born in 1971 is exp(1.54)=4.67 times longer than those born in 1970. Similarly for other cohorts, the mean survival times are 5.19, 2.95, and 13.18 times longer than for the baseline cohort. The best year (that is, the cohort with the largest mean survival time) is 1974. The likelihood-ratio test, reported at the bottom of the **mestreg** output, indicates that the model with random effects for territory fits better than the model without random effects. Random effects on territory have a variance equal to 0.58.

Now, let's fit a model with random effects on mother.

```
view me4.smcl

. mestreg  i.year  || mother: , distribution(weibull) time

        failure _d:  death
  analysis time _t:  time

Mixed-effects Weibull regression -- AFT         Number of obs     =      5,245
Group variable:           mother                Number of groups  =        266

                                                Obs per group:
                                                              min =         16
                                                              avg =       19.7
                                                              max =         25

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(4)      =    2759.49
Log likelihood =  -6685.807                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |
        1971 |   1.569194   .0449148    34.94   0.000     1.481162    1.657225
        1972 |    1.63761   .0455232    35.97   0.000     1.548386    1.726834
        1973 |   1.100873    .042576    25.86   0.000     1.017425     1.18432
        1974 |   2.606122   .0555546    46.91   0.000     2.497237    2.715007
             |
       _cons |    .867686   .0630198    13.77   0.000     .7441694    .9912026
-------------+----------------------------------------------------------------
       /ln_p |   .1533112   .0141703    10.82   0.000      .125538    .1810845
-------------+----------------------------------------------------------------
mother       |
   var(_cons)|   .8501798    .0818362                      .7040058    1.026704
------------------------------------------------------------------------------
LR test vs. Weibull model: chibar2(01) = 2131.55      Prob >= chibar2 = 0.0000
```

The variance of random effects on mother is estimated to be 0.85. Fox et al. (2006) mentioned that the mother effect is confounded with the territory effect. Because mothers are nested within territories, we can fit a three-level nested model:

```
view me5.smcl

. mestreg  i.year || territory: || mother: , distribution(weibull) time

        failure _d:  death
  analysis time _t:  time

Mixed-effects Weibull regression -- AFT         Number of obs     =      5,245

-------------------------------------------------------------
                |     No. of       Observations per Group
 Group Variable |     Groups    Minimum    Average    Maximum
----------------+--------------------------------------------
      territory |         20        216      262.3        296
         mother |        266         16       19.7         25
-------------------------------------------------------------

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(4)      =    2768.07
Log likelihood =    -6591.4                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |
        1971 |   1.566131   .0447887    34.97   0.000     1.478347    1.653915
        1972 |   1.639032   .0453955    36.11   0.000     1.550058    1.728005
        1973 |   1.097432   .0424516    25.85   0.000     1.014228    1.180635
        1974 |   2.603477   .0554919    46.92   0.000     2.494715     2.71224
             |
       _cons |   .8277945   .1686613     4.91   0.000     .4972245    1.158365
-------------+----------------------------------------------------------------
       /ln_p |    .152976     .014169    10.80   0.000     .1252053    .1807467
-------------+----------------------------------------------------------------
territory    |
   var(_cons)|   .5523799    .1817404                      .2898552    1.052676
-------------+----------------------------------------------------------------
territory>   |
mother       |
   var(_cons)|   .2879098    .0325725                      .2306515    .3593823
------------------------------------------------------------------------------
LR test vs. Weibull model: chi2(2) = 2320.37          Prob > chi2 = 0.0000
```

Now that we have accounted for the variability due to territory, the variability due to mother is lower.

The model above can be written as follows in the AFT parameterization,

$$ln(T) = xb + u_i + u_{ij} + \epsilon_{ijk}$$

where level-3 (territory) and level-2 (mother nested within territory) random effects, u_i and u_{ij}, are normally distributed with zero means and with variances σ_3^2 and σ_2^2. ϵ_{ijk} are the error terms; their variance, σ_1^2, is the variance component at the individual level. For AFT models in general, ϵ_{ijk} are assumed to be independent and identically distributed, and their distribution depends on the model we are fitting. See Rodríguez (2010) for a discussion of the distribution of the errors for different AFT survival models. In the case of the Weibull distribution, ϵ_{ijk} follow the Gumbell distribution, and their variance is $\pi^2/(6 \times p^2)$, where p is the ancillary parameter in the Weibull distribution. We can compute the residual variance using the estimate of the log of p stored in **_b[ln_p:_cons]** by **mestreg**:

```
. display _pi^2/(6 *exp(2*_b[ln_p:_cons]))
1.2113656
```

From the values obtained above, we can construct the following table of variance components:

		Variance component	
Level	Grouping variable	Symbol	Estimate
1	individual (_n)	σ_1^2	1.2113656
2	mother<territory	σ_2^2	0.2879098
3	territory	σ_3^2	0.5523799

We can use these values to compute intraclass correlations (ICCs). For example, we can obtain the formulas for ICCs from the *Methods and formulas* section in the [ME] **mixed postestimation** entry.

Level-3 (territory) ICC: $\rho^{(3)} = \sigma_3^2/(\sigma_1^2 + \sigma_2^2 + \sigma_3^2)$

Level-2 (mother) ICC: $\rho^{(2)} = (\sigma_2^2 + \sigma_3^2)/(\sigma_1^2 + \sigma_2^2 + \sigma_3^2)$

We can compute ICCs manually using the values of variance components from the table, or we can use **nlcom**:

```
view me7.smcl
. local sigmasq1  _pi^2/(6 *exp(2*_b[ln_p:_cons]))

. local sigmasq2 _b[var(_cons[territory>mother]):_cons]

. local sigmasq3  _b[var(_cons[territory]):_cons]


.
. nlcom (rho_3: (`sigmasq3')/(`sigmasq1'+`sigmasq2'+`sigmasq3')) ///
>  (rho_2: (`sigmasq2'+`sigmasq3')/(`sigmasq1'+`sigmasq2'+`sigmasq3'))

       rho_3:  (_b[var(_cons[territory]):_cons])/(_pi^2/(6 *exp(2*_b[ln_p:_cons
> ]))+_b[var(_cons[territory>mother]):_cons]+_b[var(_cons[territory]):_cons])
       rho_2:  (_b[var(_cons[territory>mother]):_cons]+_b[var(_cons[territory])
> :_cons])/(_pi^2/(6 *exp(2*_b[ln_p:_cons]))+_b[var(_cons[territory>mother]):_c
> ons]+_b[var(_cons[territory]):_cons])

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       rho_3 |   .2692362   .0648888     4.15   0.000     .1420566    .3964158
       rho_2 |   .4095667    .053254     7.69   0.000     .3051908    .5139425
------------------------------------------------------------------------------
```

The correlation between log survival-times for birds in the same territory is 0.27 with a 95% CI of [0.14, 0.40], whereas the correlation between log survival-times for birds with the same mother (and therefore also in the same territory) is 0.41 with a 95% CI of [0.31, 0.51]. We can also apply transformations as described in *Methods and formulas* in [ME] **mixed postestimation** to ensure that the confidence limits are always between 0 and 1.

References

Fox, G. A., B. E. Kendall, J. W. Fitzpatrick, and G. E. Woolfenden. 2006. Consequences of heterogeneity in survival probability in a population of Florida scrub-jays. *Journal of Animal Ecology* 75: 921–927.

Rodríguez, G. 2010. Parametric Survival Models. **http://data.princeton.edu/pop509/ParametricSurvival.pdf**.

—Isabel Canette
Principal Mathematician and Statistician, StataCorp LP

New blog series: Programming an estimation command in Stata®

David Drukker, StataCorp's Executive Director of Econometrics, wants to teach you how to write an estimation command in Stata. He wants your command to act and work just like official Stata estimation commands. He wants your command to support robust standard errors, to perform predictions, and to work with Stata's postestimation commands like **test** and **margins**. He wants to show you how easily you can accomplish all this and more.

David is writing a blog series to take you from a Stata programming novice to an accomplished Stata estimation command programmer. He wants you to produce commands you will be confident to share with colleagues, but he won't be disappointed if you just use them yourself.

You will learn how to

- use macros effectively
- program ado-files
- process options
- allow for **if** and **in** qualifiers
- support time-series and factor-variable operators
- program in Mata
- validate and certify your estimator
- and much more

Stata has many tools that simplify all aspects of programming estimation commands. David carefully leads you through all of these tools so you can understand why they exist and how you can leverage them to make hard things easy.

If you've ever wanted to program your own estimator, but were unsure where to start, David removes that hurdle. Get started with the road map to the blog entries: **stata.com/blog-programming**.

STATA CONFERENCE

REACHING NEW HEIGHTS

July 28–29, 2016

Looking to increase your Stata IQ, or perhaps just want an excuse to enjoy a few days in the city that is second to none? Join us for the 2016 Stata Conference, which will be held July 28–29 in Chicago, Illinois (immediately before the 2016 Joint Statistical Meetings, also in Chicago).

The Stata Conference provides a unique users-group experience, bringing together top researchers from around the world and Stata developers in an intimate atmosphere where everyone is welcome. If you haven't attended the Stata Conference before, come see what you've been missing!

- Meet with expert users and Stata developers.
- Gain new perspectives and new ways to use Stata.
- Share your insights and build connections.

Don't miss this opportunity to connect with fellow researchers and Stata developers.

Registration now open. Seats are limited.

When	July 28–29, 2016
Where	Gleacher Center The University of Chicago Booth School of Business 450 North Cityfront Plaza Drive Chicago, IL 60611
Who	Stata developers You and Stata users from around the world
Cost	$195 two days, $75 student $125 one day, $50 student

stata.com/chicago16
#stata2016

PROGRAM

Regression discontinuity designs and models

Multidimensional regression discontinuity and regression kink designs with difference-in-differences
Rafael Ribas
University of Amsterdam

TED: Stata module for testing stability of regression discontinuity models
Giovanni Cerulli
CNR - IRCrES

rdlocrand: A Stata package for inference in regression discontinuity designs under local randomization
Gonzalo Vazquez-Bare
Department of Economics, University of Michigan

Integrating Stata with web technologies

Mining Twitter data for fun and profit
Joseph Canner
Johns Hopkins University School of Medicine

Stata tweets and other API libraries
William Matsuoka
San Diego Gas & Electric

Static to live: Combining Stata with Google Charts API
Belen Chavez
San Diego Gas & Electric

Interactive data visualization for the web: Using jsonio, libd3, and libhtml to create reusable D3js
Billy Buchanan
Office of Research, Evaluation, and Assessment, Minneapolis Public Schools

Topics in Stata programming

Setting up loops and macros in Stata to estimate value-added scores for teacher evaluations
Chen An
Department of Accountability, Research and Evaluation, Orange County Public Schools

Code and algorithm visualization for statistical education and programming in Stata
E. F. Haghish
Department of Mathematics and Computer Science
University of Southern Denmark

Data analysis: Tools and techniques

What does your model say? It may depend on who is asking
David Drukker
StataCorp

Polychoric, by any other "namelist"
Stas Kolenikov
Abt SRBI

Modeling rating transition matrices for wholesale loan portfolios
Kit Baum
Boston College and DIW Berlin

Combining difference-in-difference and matching for panel-data analysis
Weihua An
Departments of Sociology and Statistics, Indiana University

A practitioners guide to implementing the two-stage residual inclusion method in Stata
Joseph Terza
Department of Economics
Indiana University Purdue University Indianapolis

Fitting new statistical models with Stata

ardl: Stata module to estimate autoregressive distributed lag models
Sebastian Kripfganz
University of Exeter Business School, Department of Economics

reghdfe: Estimating linear models with multiway fixed effects
Sergio Correia
Duke University, Fuqua School of Business

runmixregls: A mixed-effects location scale model run within Stata
Donald Hedeker
University of Chicago

Reproducible research

Combine Stata with Python using the Jupyter Notebook
Ties de Kok
Tilburg University

Analysis Manager: A reproducible research tool for generating dynamic documents using Microsoft Word
Leah Welty
Division of Biostatistics, Department of Preventive Medicine, Northwestern University

Visualizing data

Inchworm plots: Visual representation of inferential uncertainty
Dale Rhoda
Biostat Global Consulting

kwstat: A Stata module to compute kernel-weighted statistics
Florian Chávez Juárez
Centro de Investigación y Docencia Económicas (CIDE)

Vote for quantile plots! New planks in an old campaign
Nicholas J. Cox
Durham University

Hypothesis testing: Multiple comparisons and power

Comparing multiple comparisons
Philip Ender
UCLA (Ret)

Versatile tests for comparing survival curves based on weighted log-rank statistics
Theodore Karrison
University of Chicago

Bayesian analysis in Stata

Bayesian hierarchical models in Stata
Nikolay Balov
StataCorp

SCIENTIFIC COMMITTEE

Phil Schumm (Chair)
The University of Chicago
Department of Public Health Sciences

Richard Williams
University of Notre Dame
Department of Sociology

Scott Long
Indiana University
Department of Sociology

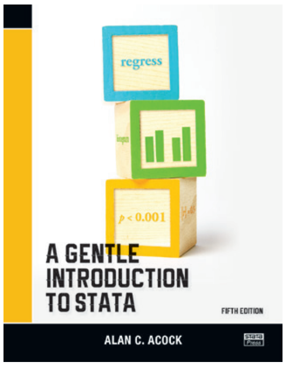
Matias Cattaneo
University of Michigan
Department of Economics

Register now:
stata.com/chicago16

New from

Stata® Press

A Gentle Introduction to Stata, Fifth Edition



Author: Alan C. Acock
Copyright: 2016
Pages: 546; paperback
Price: $58.00

Alan C. Acock's *A Gentle Introduction to Stata, Fifth Edition* is aimed at new Stata users who want to become proficient in Stata. After reading this introductory text, new users will not only be able to use Stata well but also be able to learn new aspects of Stata.

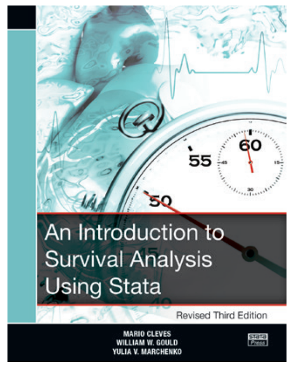
Acock assumes that the reader is not familiar with any statistical software. This assumption of a blank slate is central to the structure and contents of the book. Acock starts with the basics; for example, the part of the book that deals with data management begins with a careful and detailed example of turning survey data on paper into a Stata-ready dataset on the computer. When explaining how to go about basic exploratory statistical procedures, Acock includes notes that will help the reader develop good work habits. This mixture of explaining good Stata habits and good statistical habits continues throughout the book.

Acock is quite careful to teach the reader all aspects of using Stata. He covers data management, good work habits (including the use of basic do-files), basic exploratory statistics (including graphical displays), and analyses using the standard array of basic statistical tools (correlation, linear and logistic regression, and parametric and nonparametric tests of location and dispersion). He also successfully introduces some more advanced topics such as multiple imputation and structural equation modeling in a very approachable manner. Acock teaches Stata commands by using the menus and dialog boxes while still stressing the value of do-files. In this way, he ensures that all types of users can build good work habits. Each chapter has exercises that the motivated reader can use to reinforce the material.

The fifth edition of the book includes two new chapters that cover multilevel modeling and item response theory (IRT) models.

Read more and order online:
stata-press.com/books/gentle-introduction-to-stata

An Introduction to Survival Analysis Using Stata, Revised Third Edition



Authors: Mario Cleves, William W. Gould, and Yulia V. Marchenko
Copyright: 2016
Pages: 428; paperback
Price: $59.00

An Introduction to Survival Analysis Using Stata, Revised Third Edition is the ideal tutorial for professional data analysts who want to learn survival analysis for the first time or who are well versed in survival analysis but are not as dexterous in using Stata to analyze survival data. This text also serves as a valuable reference to those readers who already have experience using Stata's survival analysis routines.

The revised third edition has been updated for Stata 14, and it includes a new section on predictive margins and marginal effects, which demonstrates how to obtain and visualize marginal predictions and marginal effects using the **margins** and **marginsplot** commands after survival regression models.

Survival analysis is a field of its own that requires specialized data management and analysis procedures. To meet this requirement, Stata provides the **st** family of commands for organizing and summarizing survival data.

This book provides statistical theory, step-by-step procedures for analyzing survival data, an in-depth usage guide for Stata's most widely used **st** commands, and a collection of tips for using Stata to analyze survival data and to present the results. This book develops from first principles the statistical concepts unique to survival data and assumes a knowledge of only basic probability and statistics and a working knowledge of Stata.

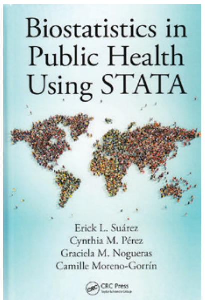
Read more and order online:
stata-press.com/books/survival-analysis-stata-introduction

Want to know what's coming next from Stata Press? Sign up to be alerted when new titles are published:

stata-press.com/alerts

New from the Stata Bookstore

Biostatistics in Public Health Using Stata

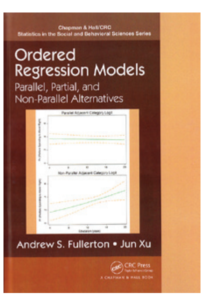


Authors: Erick L. Suárez, Cynthia M. Pérez, Graciela M. Nogueras, and Camille Moreno-Gorrín
Publisher: CRC Press
Copyright: 2016
Pages: 190; hardcover
Price: $79.50

This book demonstrates the use of Stata for statistical analyses common in public health research and in many other disciplines. For those new to Stata, the authors first provide an introduction to Stata's interface and to commands for descriptive statistics and graphics. In the discussions of statistical procedures that follow, readers will find extensive coverage of Stata's menus, dialog boxes, commands, and output. Topics include linear regression, analysis of variance, logistic regression, Poisson regression, survival analysis, multilevel mixed-effects models, and power and sample-size analysis.

Order online:
stata.com/bookstore/biostatistics-in-public-health-using-stata

Ordered Regression Models: Parallel, Partial, and Non-Parallel Alternatives



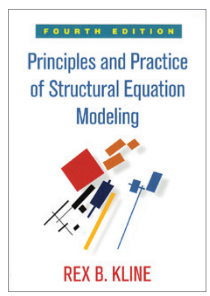
Authors: Andrew S. Fullerton and Jun Xu
Publisher: CRC Press
Copyright: 2016
Pages: 171; hardcover
Price: $69.75

In *Ordered Regression Models: Parallel, Partial, and Non-Parallel Alternatives*, Fullerton and Xu provide a thorough treatment of models for ordinal data. This book will appeal to researchers from any discipline who wish to build on their knowledge of linear, logistic, and probit regression and learn both theoretical and practical concepts related to a variety of models for ordinal outcomes.

As the title indicates, the models presented are partitioned into three groups based on whether a parallel regression assumption is made for all covariates, for a subset of the covariates, or for none of the covariates. Under each of these assumptions, the authors describe three models—cumulative, continuation ratio, and adjacent category—from which a researcher can choose, depending on the probability of interest. They also include worked examples with real data and provide advice regarding interpretation, presentation of results, choice of model, and common problems that arise. Example Stata commands for fitting these models are shown at the end of each chapter.

Order online:
stata.com/bookstore/ordered-regression-models

Principles and Practice of Structural Equation Modeling, Fourth Edition



Author: Rex B. Kline
Publisher: Guilford Press
Copyright: 2016
Pages: 534; paperback
Price: $54.50

The fourth edition of *Principles and Practice of Structural Equation Modeling* by Rex Kline, like previous editions, is an ideal text for both students and researchers who want to learn the fundamental concepts of structural equation modeling (SEM) and then apply it to their own data. Along with introducing different types of structural equation models, Kline carefully discusses practical issues, such as data preparation, assumptions, identification, and interpretation. Easy-to-follow examples use real data, and the book's website provides files demonstrating how to reproduce results using a variety of software packages, including Stata.

In the fourth edition, Kline adds new coverage of Judea Pearl's structural causal modeling, confirmatory factor analysis with ordinal indicators, bootstrapping, significance testing, and item response theory.

Order online:
stata.com/bookstore/principles-and-practice-of-structural-equation-modeling

Browse these and many more titles at **stata.com/bookstore**.

Learn how to exploit the full power of Stata with our affordable, convenient, web-based courses. We offer NetCourses for Stata users of all experience levels, from beginning to advanced.

Introduction to Stata

Learn how to use all of Stata's tools and become a sophisticated Stata user. You will understand the Stata environment, how to import and export data from different formats, how Stata's intuitive syntax works, data management in Stata, and more.

July 8–August 19, 2016.................................... $95.00

Statistical Graphics Using Stata

Learn how to communicate your data with Stata's powerful graphics features. Topics include using graphs to check model assumptions; formatting, saving, and exporting your graphs for publication; using the Graph Editor; creating custom graph schemes; creating complex graphs by layering and combining multiple graphs; using **margins** and **marginsplot**; and more.

July 8–August 19, 2016................................. $150.00

Introduction to Stata Programming

Become an expert in organizing your work in Stata. Make the most of Stata's scripting language to improve your workflow and to create concretely reproducible analyses. Learn how to speed up your work and do more complete analyses.

July 8–August 19, 2016................................. $150.00

Writing Your Own Stata Commands

Learn how to create and debug your own commands that are indistinguishable from the commands that ship with Stata.

July 15–September 2, 2016 .... $175.00

Univariate Time Series with Stata

Learn univariate time-series analysis with an emphasis on the practical aspects most needed by practitioners and applied researchers.

July 15–September 2, 2016 .... $295.00

Introduction to Panel Data Using Stata

Become an expert in the analysis and implementation of linear, nonlinear, and dynamic panel-data estimators using Stata. Geared for researchers and practitioners in all fields, this course focuses on the interpretation of panel-data estimates and the assumptions underlying the models that give rise to them.

July 15–August 26, 2016............................... $295.00

Introduction to Survival Analysis Using Stata

Learn how to effectively analyze survival data using Stata. We cover censoring, truncation, hazard rates, and survival functions. Discover how to set the survival-time characteristics of your dataset just once and apply any of Stata's many estimators and statistics to those data.

July 15–September 2, 2016 .... $295.00

stata.com/netcourse

NetCourseNow™

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 instructor to guide you through the course.

stata.com/netcourse/ncnow

Public training

Learn Stata from StataCorp's experts. These courses take place in Washington, DC, and are ideal for researchers and individuals who want to learn more or gain a deeper understanding of Stata.

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

- May 24–25, 2016 ....$950
- June 20–21, 2016 ....$950

Aimed at both new Stata users and those who wish to learn techniques for efficient day-to-day use of Stata, this course teaches you to use Stata in a reproducible manner, making collaborative changes and follow-up analyses much simpler.

Regression Modeling Using Stata

- June 22, 2016 .... $695

Regression modeling is a fundamental tool for researchers who want to establish causal quantitative relationships from observational data. Learn the theoretical concepts necessary to understand regression models and how to implement them using Stata, and learn to reinforce those concepts with exercises and examples you will solve with the assistance of the instructor.

Panel-Data Analysis Using Stata

- June 23–24, 2016.. $1,295

Learn both the theory and the practice of panel-data analysis. 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 nonlinear models. Exercises and Stata examples supplement the lessons.

Save $195 when you combine Regression Modeling Using Stata and Panel-Data Analysis Using Stata.

Find out how at stata.com/public-training.

stata.com/public-training

May

1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	**24**	**25**	26	27	28
29	30	31				

June

			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	**20**	**21**	**22**	**23**	**24**	25
26	27	28	29	30		

EUSMEX 2016

Make your plans now to attend EUSMEX 2016 on May 18 at CIDE Región Centro in Aguascalientes.

Presentations: May 18

A contingent valuation application using Stata
Arturo Robles Valencia, Universidad de Sonora

DJA command to perform the decomposition of inequalities
Linda Llamas, Universidad Estatal de Sonora and CIAD

Abdelkrim Araar, Université Laval & CIRPÉE

Luis Huesca, Centro de Investigación en Alimentación y Desarrollo, A.C. (CIAD)

Introduction to fractional outcome regression models using the fracreg and betareg commands
Miguel Dorta, StataCorp LP

Microdatos de la Encuesta Intercensal 2015 con Stata
Juan Francisco Islas Aguirre, FAO México

Programming financial models with Stata and Excel®
Carlos Alberto Dorantes Dosamantes, ITESM

Endotoxin associated to particulate matter (PM10) of a landfill facility in Cuautla, Morelos, Mexico
María Alejandra Terrazas-Meraz, Universidad Autónoma del Estado de Morelo

Introduction to Markov-switching regression models using the mswitch command
Gustavo Sanchez, StataCorp LP

New methods of interpretation using marginal effects for nonlinear models
J. Scott Long, Indiana University at Bloomington

GMM and maximum likelihood estimators with Mata and moptimize
Alfonso Miranda, División de Economía, Centro de Investigación y Docencia Económicas (CIDE)

Stationary and multiple structural change with Stata
Alfonso Mendoza-Velázquez, Universidad Popular Autónoma del Estado de Puebla

Omar Stabridis-Arana, Centro de Investigación e Inteligencia Económica (CIIE)

Consumption of tobacco in high school students
Paola Adanari Ortega-Ceballos, Facultad de Enfermería, Universidad Autónoma del Estado de Morelos

Edith Ruth Arizmendi-Jaime, Facultad de Enfermería, Universidad Autónoma del Estado de Morelos

Miriam Tapia-Domínguez, Facultad de Enfermería, Universidad Autónoma del Estado de Morelos

María Alejandra Terrazas-Meraz, Facultad de Enfermería, Universidad Autónoma del Estado de Morelos

Report to users and Wishes and grumbles

Keynote speaker: J. Scott Long

New methods of interpretation using marginal effects for nonlinear models

Marginal effects are commonly used to interpret linear and nonlinear regression models. Most simply, a marginal effect (ME) computes the change in the outcome for a fixed amount of change in one predictor while holding other predictors constant. This presentation considers a variety of nonstandard applications of MEs in a single model and compares effects across models.

Read more online: **stata.com/meeting/mexico16**

Registration

Meeting fees	Price
Professionals	MEX $1,700.00 + IVA (16%)
Students	MEX $850.00 + IVA (16%)

Scientific committee

Alfonso Miranda
Centro de Investigación y Docencia Económica (CIDE)

Luis Huesca Reynoso
Centro de Investigación en Alimentación y Desarrollo, CIAD—Hermosillo

Benjamín Sexto
MultiON Consulting—Estadístico y especialista en Stata

View the program schedule and register:
stata.com/meeting/mexico16

2016 International Stata Users Group meetings

2016 German Stata Users Group meeting

Make your plans now to attend the 2016 German Stata Users Group meeting at GESIS Cologne on June 10.

Presentations: June 10

Social network analysis using Stata
Thomas Grund, University College Dublin

Nonparametric frontier analysis using Stata
Oleg Badunenko, University of Cologne

Bayesian data analysis using Stata
Yulia Marchenko, StataCorp LP

Dynamic Stata help files using MarkDoc
E. F. Haghish, Institute for Medical Biometry and Statistics (IMBI), University of Freiburg

texdoc 2.0: An update on creating LaTeX documents from within Stata
Ben Jann, University of Bern

Marginal effects in multiply imputed datasets
Daniel Klein, University of Kassel

The assessment of fit in the class of logistic regression models: A pathway out of the jungle of pseudo-R^2s using Stata
Wolfgang Langer, University of Halle-Wittenberg

Influence functions at work
Phillipe van Kerm, Luxembourg Institute of Socio-Economic Research

Analysis of sequences using Stata, 2.0
Ulrich Kohler, University of Potsdam

Report to users and Wishes and grumbles

View the program schedule and register:
stata.com/meeting/germany16

Workshop: Introduction to Mata

by Ulrich Kohler, University of Potsdam
June 9, 9:30 a.m. to 4:30 p.m.

Mata is a full-fledged programming language that operates in the Stata environment. It is designed to make programming functions for matrices really easy. The workshop provides an introduction to basic Mata concepts and a step-by-step example of implementing an estimator for Stata with Mata.

Get the details online:
stata.com/meeting/germany16

Registration

Meeting fees	Price
Meeting only: Professionals	€45
Meeting only: Students	€35
Workshop only	€65
Workshop + meeting	€85

There will also be an optional informal meal at a restaurant in Cologne on Friday evening at additional cost.

To enroll, contact Christiane Senczek.

Christiane Senczek
Dittrich & Partner Consulting GmbH
Prinzenstr. 2
42697 Solingen
Germany
Tel: +49 (0) 212 2 60 66-0
Email: **christiane.senczek@dpc.de**

Scientific committee

Johannes Giesecke
Humboldt University of Berlin

Ulrich Kohler
University of Potsdam

Alexander Schmidt-Catran
University of Cologne

Alexander Jedinger
GESIS Cologne

Alexia Katsanidou
GESIS Cologne

View the program schedule and register:
stata.com/meeting/germany16

Belgian Stata Users Group meeting
Brussels | September 6, 2016

stata.com/meeting/belgium16

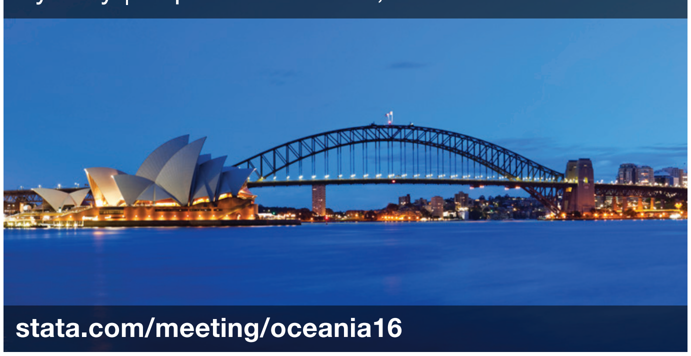
London Stata Users Group meeting
London | September 8–9, 2016

stata.com/meeting/uk16

Nordic and Baltic Stata Users Group meeting
Oslo | September 13, 2016

stata.com/meeting/nordic-and-baltic16

Oceania Stata Users Group meeting
Sydney | September 29–30, 2016



stata.com/meeting/oceania16

Spanish Stata Users Group meeting
Barcelona | October 20, 2016

stata.com/meeting/spain16

Swiss Stata Users Group meeting
Bern | November 17, 2016

stata.com/meeting/switzerland16

Italian Stata Users Group meeting
Rome | November 17–18, 2016

stata.com/meeting/italy16

Check our site frequently for updates:

stata.com/meeting

Get the latest meeting details direct to your inbox. Sign up for an email alert today:

stata.com/alerts

ICPSR Summer Program in Quantitative Methods of Social Research

June–August 2016

Since 1963, the Inter-university Consortium for Political and Social Research (ICPSR) has offered the ICPSR Summer Program in Quantitative Methods of Social Research as a complement to its data services. The Summer Program provides a comprehensive program of studies in research design, statistics, data analysis, and social science methodology. The Summer Program has become internationally recognized as a preeminent learning environment for basic and advanced training in the methodologies and technologies of social science research.

Three of this year's ICPSR courses are taught by StataCorp statisticians and will be of particular interest to Stata users.

Handling Missing Data Using Multiple Imputation in Stata

Rose Medeiros, Senior Statistician
July 6–8, 2016

This course will cover the use of Stata to perform multiple-imputation analysis. Multiple imputation (MI) is a simulation-based technique for handling missing data. The course will provide a brief introduction to multiple imputation and will focus on how to perform MI in Stata using the **mi** command. The three stages of MI (imputation, complete-data analysis, and pooling) will be discussed in detail with accompanying Stata examples. Various imputation techniques will be discussed, including multivariate normal imputation (MVN) and multiple imputation using chained equations (MICE). Also, several examples demonstrating how to efficiently manage multiply imputed data within Stata will be provided. Linear and logistic regression analysis of multiply imputed data as well as several postestimation features will be presented.

Structural Equation Modeling with Stata

Kristin MacDonald, Asst. Director of Statistical Services
July 18–20, 2016

This workshop covers the use of Stata for structural equation modeling (SEM). SEM is a class of statistical techniques for modeling relationships among variables, both observed and unobserved. SEM encompasses some familiar models such as linear regression, multivariate regression, and factor analysis and extends to a variety of more complicated models. The workshop will give an introduction to structural equation modeling. In addition, a number of models that fall within the SEM framework will be discussed with an emphasis on using Stata to fit each one. Stata allows for fitting structural equation models in two ways—by using the **sem** command syntax or by using the graphical user interface to draw path diagrams. Examples will demonstrate both approaches. Knowledge of basic statistical techniques such as correlation and linear regression is recommended.

Multilevel and Mixed Models Using Stata

Rose Medeiros, Senior Statistician
July 27–29, 2016

This three-day workshop is an introduction to using Stata to fit multilevel mixed models.

Mixed models contain both fixed effects analogous to the coefficients in standard regression models and random effects not directly estimated but instead summarized through the unique elements of their variance-covariance matrix. Mixed models may contain more than one level of nested random effects, and hence these models are also referred to as "multilevel" or "hierarchical models," particularly in the social sciences. Stata's approach to linear mixed models is to assign random effects to independent panels where a hierarchy of nested panels can be defined for handling nested random effects.

We will start by learning how random-intercept models are related to classical linear models and will become familiar with the terminology for both approaches. Next, we will make the jump from random intercepts to random coefficients and the various covariance structures that can be imposed with multiple random-effects terms. We will then finish our estimation for linear mixed models by seeing that Stata has niceties that allow fitting more complex models, including crossed-effects models, growth curve models, and models with complex and grouped constraints on covariance structures. After all the model fitting, we will turn to common postestimation tasks such as predictions, model diagnostics, and model comparisons. To finish up, we will apply what we have learned about linear mixed models to models for other types of responses, in particular binary and count responses.

The workshop will be interactive in nature. We will consider concrete examples using Stata as we learn each of the concepts.

For more information, visit **stata.com/news/icpsr2016**.

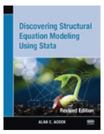
Stata Press

Stata Press titles now available as eBooks

You asked for them. We listened. Your favorite Stata Press titles are now available as eBooks, or will be soon.

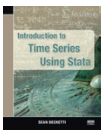
You can now read and refer to Stata Press books on your tablet, your computer, your Kindle®, or even on your phone (in a pinch).

The following titles are available right now.



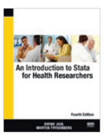
Discovering Structural Equation Modeling Using Stata, Revised Edition

By Alan C. Acock



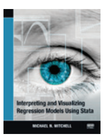
Introduction to Time Series Using Stata

By Sean Becketti



An Introduction to Stata for Health Researchers, Fourth Edition

By Svend Juul and Morten Frydenberg

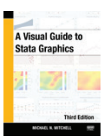


Interpreting and Visualizing Regression Models Using Stata

By Michael N. Mitchell

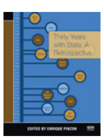
Stata for the Behavioral Sciences

By Michael N. Mitchell



A Visual Guide to Stata Graphics, Third Edition

By Michael N. Mitchell



Thirty Years with Stata: A Retrospective

Edited by Enrique Pinzon

All seven titles are available now using VitalSource Bookshelf®. Becketti and Acock are also available in Kindle format from Amazon. The remaining five titles will be available in Kindle format soon—between Monday, May 9th and Tuesday, May 17th.

All other Stata Press titles are being converted to eBooks. Be patient—we want our titles to take full advantage of the electronic format. As with the first seven titles, we always want our Stata output to be readable, our graphs to be pretty, and our mathematical formulas to behave as you would wish—to look as good as the text and scale with the text.

You can see all of our electronic titles at **stata-press.com/ebooks**. You can also sign up there to be alerted when other titles become available.

stata-press.com/ebooks

Contact us

service@stata.com **stata.com**

Find a Stata distributor near you:
stata.com/worldwide