Jan/Feb/Mar
Vol 31 No 1

The Stata® News

Statistics Graphics Data Management & Analysis

How do you compare?

Examining effects in multilevel models using contrasts.

If your statistical background includes ANOVA modeling and analysis of designed experiments, you are likely already familiar with contrasts. If not, contrasts may be a new concept, or perhaps you use certain types of contrasts but call them by different names. So before we jump to contrasts with multilevel models, let's look at a few types of tests you can perform using Stata's **contrast** command.

Making comparisons using contrast

As with the example data in Keppel and Wickens (2004, chapter 13), let's suppose we have fifth-grade students who are asked to learn a set of vocabulary words and are tested on them a week later. We are interested in comparing three teaching methods and four types of words. When teaching, the teacher provides no verbal feedback, positive feedback, or negative feedback. The four word lists are categorized by the frequency with which the words are used, where **freq=1** corresponds to words least frequently used and **freq=4** corresponds to words most frequently used.

Let's start by regressing vocabulary test score on the frequency categories.

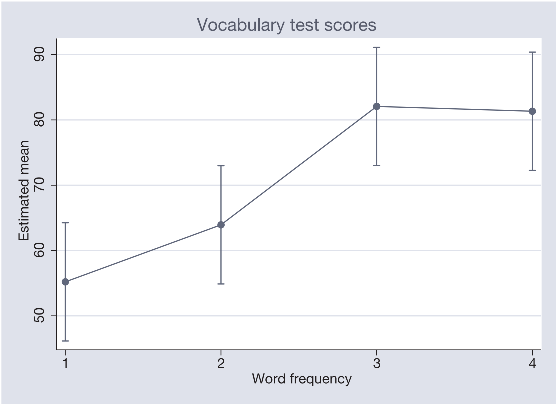
```
view contrasts1.smcl

. regress score i.freq

      Source |       SS           df       MS      Number of obs   =        60
-------------+----------------------------------   F(3, 56)        =      8.62
       Model |  7924.33333         3  2641.44444   Prob > F        =    0.0001
    Residual |     17151.6        56  306.278571   R-squared       =    0.3160
-------------+----------------------------------   Adj R-squared   =    0.2794
       Total |  25075.9333        59  425.015819   Root MSE        =    17.501

------------------------------------------------------------------------------
       score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   frequency |
          2  |   8.733333   6.390395     1.37   0.177    -4.068165    21.53483
          3  |   26.86667   6.390395     4.20   0.000     14.06517    39.66817
          4  |   26.13333   6.390395     4.09   0.000     13.33183    38.93483
             |
       _cons |       55.2   4.518691    12.22   0.000     46.14797    64.25203
------------------------------------------------------------------------------
```

The coefficients in the regression output are one type of contrast—comparisons with the base level of least frequently used words. Mean test scores for the third and fourth frequency levels are statistically greater than those for the first (base) level. However, it would be nice to know whether differences exist for each increase in word frequency level. Are there differences between the second and third levels? Between the third and fourth levels? We can use the **ar.** contrast operator to make these "reverse adjacent" comparisons.

```
view contrasts2.smcl

. contrast ar.freq, pveffects nowald

Contrasts of marginal linear predictions

Margins      : asbalanced

--------------------------------------------------------
             |   Contrast   Std. Err.       t    P>|t|
-------------+------------------------------------------
   frequency |
   (2 vs 1)  |   8.733333   6.390395     1.37    0.177
   (3 vs 2)  |   18.13333   6.390395     2.84    0.006
   (4 vs 3)  |  -.7333333   6.390395    -0.11    0.909
--------------------------------------------------------
```

Here we find that the change from level 2 to level 3 is the only one with a statistically significant difference in estimated mean scores.

We can see this large jump when we plot the estimated means using **margins freq** followed by the **marginsplot** command.



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

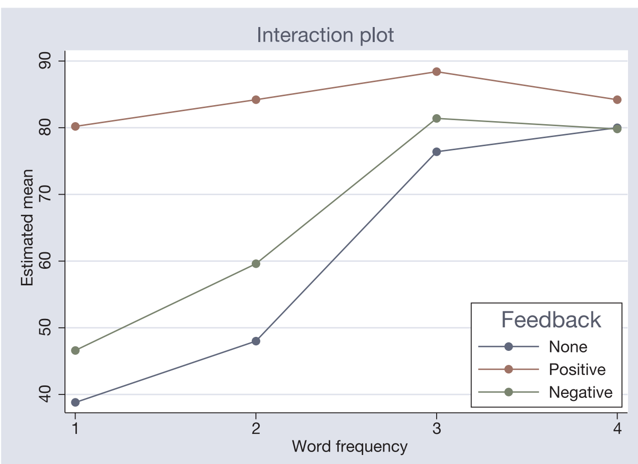
Now, let's fit a model allowing for an interaction between frequency and type of feedback by typing

```
. regress score freq##feedback
(output omitted)
```

We can use **contrast** to obtain tests of the main and interaction effects.

```
view contrasts3.smcl

. contrast freq##feedback

Contrasts of marginal linear predictions

Margins      : asbalanced

------------------------------------------------
                   |         df           F        P>F
-------------------+----------------------------
         frequency |          3       14.97     0.0000
                   |
          feedback |          2       16.80     0.0000
                   |
frequency#feedback |          6        2.60     0.0291
                   |
       Denominator |         48
------------------------------------------------
```

We have statistical evidence for an interaction effect, but at this point, we can't say much about where differences exist. A plot allows us to visualize the interaction.



We can test specific hypotheses about the relationship between word frequency and teaching method. For example, when positive feedback is given, **feedback**=2, does word frequency have an effect?

```
view contrasts3a.smcl

. contrast freq@2.feedback

Contrasts of marginal linear predictions

Margins      : asbalanced

------------------------------------------------
                   |         df           F        P>F
-------------------+----------------------------
frequency@feedback |
          Positive |          3        0.32     0.8125
                   |
       Denominator |         48
------------------------------------------------
```

There is no statistically significant difference in the four means for this teaching method.

We can also test whether there is an effect of teaching method at individual frequency levels.

```
view contrasts4.smcl

. contrast feedback@frequency

Contrasts of marginal linear predictions

Margins      : asbalanced

------------------------------------------------
                   |         df           F        P>F
-------------------+----------------------------
feedback@frequency |
                 1 |          2       13.71     0.0000
                 2 |          2        9.68     0.0003
                 3 |          2        1.03     0.3649
                 4 |          2        0.17     0.8401
             Joint |          8        6.15     0.0000
                   |
       Denominator |         48
------------------------------------------------
```

For the first and second frequency levels (but not the third and fourth), there is a statistical difference in the estimated means for the different teaching methods.

Is there really any interaction if we consider only the change from the first to second frequency level?

```
view contrasts5.smcl

. contrast ar(1 2).frequency#feedback

Contrasts of marginal linear predictions

Margins      : asbalanced

------------------------------------------------
                   |         df           F        P>F
-------------------+----------------------------
frequency#feedback |          2        0.29     0.7502
                   |
       Denominator |         48
------------------------------------------------
```

We do not find evidence of an interaction effect at these frequency levels.

Multilevel models

We can easily apply the same types of contrasts when fitting multilevel models. For example, suppose that each student was tested multiple times. We can fit a random-effects model allowing for student-level variation in the intercepts and perform the same types of contrasts. To do this, we could type

```
. mixed score feedback##freq || id:
. contrast feedback##freq
. contrast feedback@frequency
```

However, acknowledging that we have repeated measurements on students and a fairly small sample size, we request that both **mixed** and **contrast** report small-sample tests using a repeated-measures ANOVA method for computing denominator degrees of freedom.

```
view contrasts6.smcl

. mixed score freq##feedback|| id:, dfmethod(repeated)

Mixed-effects ML regression                     Number of obs     =        180
Group variable: id                              Number of groups  =         60

                                                Obs per group:
                                                              min =          3
                                                              avg =        3.0
                                                              max =          3
DF method: Repeated                             DF:           min =      48.00
                                                              avg =      48.00
                                                              max =      48.00

                                                F(11,     48.00)  =       9.37
Log likelihood = -738.55761                     Prob > F          =     0.0000

------------------------------------------------------------------------------------
             score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
         frequency |
                 2 |   11.46667   7.577433     1.51   0.137    -3.768784    26.70212
                 3 |         37   7.577433     4.88   0.000     21.76455    52.23545
                 4 |   39.93333   7.577433     5.27   0.000     24.69788    55.16878
                   |
          feedback |
          Positive |         41   7.577433     5.41   0.000     25.76455    56.23545
          Negative |   10.73333   7.577433     1.42   0.163    -4.502117    25.96878
                   |
frequency#feedback |
        2#Positive |  -5.866667   10.71611    -0.55   0.587    -27.41285    15.67951
        2#Negative |   4.97e-14   10.71611     0.00   1.000    -21.54618    21.54618
        3#Positive |      -30.8   10.71611    -2.87   0.006    -52.34618   -9.253819
        3#Negative |       -4.8   10.71611    -0.45   0.656    -26.34618    16.74618
        4#Positive |  -36.93333   10.71611    -3.45   0.001    -58.47951   -15.38715
        4#Negative |  -9.266667   10.71611    -0.86   0.391    -30.81285    12.27951
                   |
             _cons |         38   5.358054     7.09   0.000     27.22691    48.77309
------------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                   var(_cons) |   93.08077   27.00498      52.71156    164.3668
-----------------------------+------------------------------------------------
                var(Residual) |   151.3889   19.54422       117.545    194.9771
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 23.54         Prob >= chibar2 = 0.0000

. contrast freq##feedback, small

Contrasts of marginal linear predictions

Margins      : asbalanced

------------------------------------------------------------
                   |         df        ddf          F        P>F
-------------------+----------------------------------------
score              |
         frequency |          3      48.00      15.62     0.0000
                   |
          feedback |          2      48.00      18.57     0.0000
                   |
frequency#feedback |          6      48.00       3.18     0.0105
------------------------------------------------------------
```

As before, we find an interaction between the type of feedback and the frequency of word use. Using the **contrast** command below, we also find that the simple effects of feedback exist only for the lower-frequency words.

```
view contrasts7.smcl

. contrast feedback@frequency, small

Contrasts of marginal linear predictions

Margins      : asbalanced

----------------------------------------------------------------
                   |         df        ddf          F        P>F
-------------------+--------------------------------------------
score              |
feedback@frequency |
                 1 |          2      48.00      15.75     0.0000
                 2 |          2      48.00      11.29     0.0001
                 3 |          2      48.00       0.91     0.4078
                 4 |          2      48.00       0.15     0.8630
             Joint |          8      48.00       7.02     0.0000
----------------------------------------------------------------
```

And beyond

We have considered fairly traditional experimental design applications. However, the use of **contrast** is not so limited. **contrast** can be used after fitting most models in Stata. Options are available for working with unequal group sizes and for adjusting results for multiple comparisons. Find details on these and many other extensions in [R] **contrast**.

—Kristin MacDonald
Asst. Director of Statistical Services

Reference

Keppel, G., and T. D. Wickens. 2004. *Design and Analysis: A Researcher's Handbook. 4th ed.* Upper Saddle River, NJ: Pearson.

Bookmarks: Series 8 now available

Collect the latest set of five commemorative bookmarks documenting the accomplishments of Thomas Bayes, Vera Nikolaevna Kublanovskaya, Max Otto Lorenz, Andrey Markov, and John Snow.

Available March 7. $2.50 in North America, $3.75 internationally. Prices include shipping.

Read the bios for these and other statisticians featured in our bookmarks at **stata.com/giftshop/bookmarks**.

In the spotlight: Bayesian IRT—4PL model

Item response theory (IRT) is often used for modeling the relationship between the latent abilities of a group of subjects and the examination items used for measuring their abilities. Stata 14 introduced a suite of commands for fitting IRT models using maximum likelihood; see, for example, "In the spotlight: **irt**" by Rafal Raciborski in *Stata News*, Volume 30 Number 3 and the [IRT] *Item Response Theory Reference Manual* for more details. One-parameter logistic (1PL), 2PL, and 3PL IRT models are commonly used to model binary responses. Models beyond 3PL, such as 4PL and 5PL models, have not been as widely used. One of the reasons is the difficulty in estimating the additional parameters introduced by these models using maximum likelihood. In recent years, these models have been reconsidered within the Bayesian framework (Loken and Rulison 2010; Fox 2010; Kim and Bolt 2007). In this article, we demonstrate how to fit a Bayesian 4PL model using **bayesmh**.

Data

We will use the abridged version of the mathematics and science data from DeBoeck and Wilson (2004), which contains 800 student responses, **y**, to 9 test questions (items) intended to measure mathematical ability. To fit IRT models using **bayesmh**, the data must be in long form with items, **item**, recorded as multiple observations per subject, **id**.

Model

We consider the following 4PL model,

$$P(Y_{ij} = 1) = c + (d - c)\text{InvLogit}\{a_i(\theta_j - b_i)\},\ c < d < 1$$

where $i = 1, 2, \ldots, 9$ and $j = 1, 2, \ldots, 800$. The 4PL model extends the 3PL model by adding an upper asymptote parameter $d \neq 1$. The d parameter can be viewed as an upper limit on the probability of correct response to the ith item. The probability of giving correct answers by subjects with very high ability can thus be no greater than d. a_i and b_i are item-specific discrimination and difficulties. Here we consider a common guessing parameter c and a common upper asymptote parameter d, but they can also be item specific. InvLogit( ) is an inverse-logit function. The latent abilities θ_j are assumed to be normally distributed:

$$\theta_j \sim N(0,1)$$

A Bayesian formulation also requires prior specifications for all other model parameters. This is an important step in Bayesian modeling and must be considered carefully. For illustration, we consider the following priors.

Discrimination parameters a_i's are assumed to be positive and are often modeled in the log scale. Because we have no prior knowledge about the discrimination and difficulty parameters, we assume that the prior distributions of $ln(a_i)$ and b_i have support on the whole real line and are symmetric. A normal prior distribution is thus a natural choice. To control the impact of the prior on these parameters, we consider a hierarchical Bayesian model specification and introduce hyperparameters to model means and variances of the normal prior distribution.

$$ln(a_i) \sim N(\mu_a, \sigma_a^2)$$

$$b_i \sim N(\mu_b, \sigma_b^2)$$

We use informative priors for the guessing parameter c and the upper asymptote parameter d. We assume that the prior mean of c is about 0.1 and use an inverse-gamma prior with shape 10 and scale 1 for c. We restrict d to the (0.8,1) range and assign it a Uniform(0.8,1) prior.

$$c \sim \text{InvGamma}(10, 1)$$

$$d \sim \text{Uniform}(0.8, 1)$$

The mean hyperparameters, μ_a and μ_b, and variance hyperparameters, σ_a^2 and σ_b^2, require informative prior specifications. We assume that the means are centered at 0 with a variation of 0.1. To lower the variability of the $ln(a_i)$ and b_i parameters, we use an inverse-gamma prior with shape 10 and scale 1 for the variance parameters so that their prior means are about 0.1.

$$\mu_a, \mu_b \sim N(0, 0.1)$$

$$\sigma_a^2, \sigma_b^2 \sim \text{InvGamma}(10, 1)$$

Using bayesmh

We specify the model above using **bayesmh** as follows:

```
. bayesmh y,                                                         ///
    likelihood(dbernoulli(({gues}+({d}-{gues})*invlogit({discrim:}*({subj:}-{diff:})))*
                          cond({gues}<{d},1,.)))                      ///
    redefine(discrim:i.item) redefine(diff:i.item)                    ///
    redefine(subj:i.id)                                               ///
    prior({subj:i.id},      normal(0, 1))                             ///
    prior({discrim:i.item}, lognormal({mu_a}, {var_a}))               ///
    prior({diff:i.item},    normal({mu_b}, {var_b}))                  ///
    prior({gues},           igamma(10, 1))                            ///
    prior({d},              uniform(0.8, 1))                          ///
    prior({mu_a} {mu_b},    normal(0, 0.1))                           ///
    prior({var_a} {var_b},  igamma(10, 1))                            ///
    block({mu_a mu_b var_a var_b gues d}, split)                      ///
    init({discrim:i.item} 1 {gues} 0.1 {d} 0.9)                       ///
    exclude({subj:i.id}) burnin(5000) saving(sim4pl)

. estimates store est4pl
```

The first two lines model the probability of success of a Bernoulli outcome as a nonlinear function of model parameters. Subject-specific parameters **{subj:}** and item-specific parameters **{discrim:}** and **{diff:}** are included as "random effects" in the model by using the corresponding **redefine()** options (available in Stata 14.1) for computational efficiency. The priors for model parameters are specified in the corresponding **prior()** options. We place model parameters in separate blocks to improve the simulation efficiency and provide more sensible initial values for some of the parameters. Here we treat the abilities **{subj:i.id}** as nuisance parameters and exclude them from the final results. We use a longer burn-in period, **burnin(5000)**, to allow for longer adaptation of the MCMC sampler, which is needed given the large number of parameters in the model.

bayesmh produces the following results:

```
view irt_masc1_4pl_b.smcl

Model summary
------------------------------------------------------------------------------
Likelihood:
  y ~ bernoulli(<expr1>)

Priors:
  {discrim:i.item} ~ lognormal({mu_a},{var_a})                             (1)
     {diff:i.item} ~ normal({mu_b},{var_b})                                (2)
       {subj:i.id} ~ normal(0,1)                                           (3)
            {gues} ~ igamma(10,1)
               {d} ~ uniform(0.8,1)

Hyperpriors:
    {mu_a mu_b} ~ normal(0,0.1)
  {var_a var_b} ~ igamma(10,1)

Expression:
  expr1 : ({gues}+({d}-{gues})*invlogit(xb_discrim*(xb_subj-xb_diff)))* cond({
          gues}<{d},1,.)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_discrim.
(2) Parameters are elements of the linear form xb_diff.
(3) Parameters are elements of the linear form xb_subj.

Bayesian Bernoulli model                         MCMC iterations  =     15,000
Random-walk Metropolis-Hastings sampling         Burn-in          =      5,000
                                                 MCMC sample size =     10,000
                                                 Number of obs    =      7,200
                                                 Acceptance rate  =      .3847
                                                 Efficiency:  min =    .006254
                                                              avg =     .02634
Log marginal likelihood =          .                          max =      .1088

------------------------------------------------------------------------------
             |                                                Equal-tailed
             |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
discrim      |
        item |
          1  |  2.131242    .579461   .062313   2.025073   1.397764   3.500651
          2  |  .9194758   .1565105   .010559   .9090188   .6381238   1.261683
          3  |  1.281349   .2375522   .025492   1.255722    .905015   1.812689
          4  |  1.112241   .2060753   .014597   1.092666   .7583655   1.580387
          5  |  1.782875   .4446668   .032616   1.711008   1.102032    2.81183
          6  |  1.705848   .4304981   .035988   1.628216   1.100351   2.725116
          7  |  .9538824     .22431     .0194    .925272   .6077872   1.514065
          8  |  2.366938   .6741331   .058931   2.266726   1.374332    3.92576
          9  |  .8958829   .1491062   .008835   .8874469   .6267439   1.233588
-------------+----------------------------------------------------------------
diff         |
        item |
          1  | -.3278638   .0891546    .00731     -.3276  -.5027045  -.1585413
          2  |  .1251416   .1403733   .008707   .1273502    -.14859    .419007
          3  | -1.396515   .1856709   .013997  -1.378771  -1.807856  -1.073561
          4  |  .5588576   .1299679   .008046   .5591652   .3151135   .8273157
          5  |  1.573063   .1866931   .012489   1.555101   1.256049   1.990383
          6  |  .7530183   .1090224   .007194   .7516552   .5422617   .9769538
          7  |  1.902306   .2683965   .015209   1.874026   1.447272   2.505409
          8  | -1.464989   .1427442   .010788  -1.453497  -1.767098  -1.215461
          9  | -1.071631    .188026   .011624  -1.061141  -1.487697  -.7293511
-------------+----------------------------------------------------------------
        gues |  .1190381   .0246742    .00312   .1191699   .0717466   .1662539
           d |  .9642994   .0147669   .001249   .9643465   .9354639   .9920165
        mu_a |  .2550047     .14719   .011946    .251386  -.0287508   .5540051
       var_a |  .1324061   .0453594   .002314   .1239014   .0686404   .2446596
        mu_b |  .0484546   .1969978   .006302   .0465187  -.3450639   .4273441
       var_b |  .5500093   .1683348   .005102   .5186126   .3140066   .9534982
------------------------------------------------------------------------------
```

The upper asymptote parameter **d** is estimated to be 0.96 with a 95% credible interval of [0.94, 0.99]. The estimate is fairly close to one, so a simpler 3PL model may be sufficient for these data.

Comparing models

More formally, we can compare deviance information criteria (DIC) of the 4PL and the 3PL (with $d = 1$) models.

```
view irt_masc1_ic.smcl

. bayesstats ic est4pl, diconly

Deviance information criterion

------------------------
             |       DIC
-------------+----------
      est4pl |  8032.699
------------------------
```

The DIC of the 3PL model (not shown here) is 8049.4. The 4PL model has a lower DIC value, 8032.7, which suggests that the 4PL model provides a better fit. However, we should not rely solely on the DIC values to make our final model selection. A practitioner may still prefer the simpler 3PL model given that the upper asymptote estimate is close to one.

For more examples of Bayesian binary IRT models and details about model specifications, see our blog entry: "Bayesian binary item response theory models using **bayesmh**" (**stata.com/blog/bayes-irt**).

—Nikolay Balov
Senior Statistician and Software Developer, StataCorp

—Yulia Marchenko
Executive Director of Statistics, StataCorp

References

De Boeck, P., and M. Wilson, ed. 2004. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer.

Fox, J.-P. 2010. *Bayesian Item Response Modeling: Theory and Applications*. New York: Springer.

Kim, J.-S., and D. M. Bolt. 2007. Estimating item response theory models using Markov chain Monte Carlo methods. *Educational Measurement: Issues and Practice* 26: 38–51.

Loken, E., and K. L. Rulison. 2010. Estimation of a four-parameter item response theory model. *British Journal of Mathematical and Statistical Psychology* 63: 509–525.

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? You can't miss with the 2016 Stata Conference, which will be held July 28–29 in Chicago, Illinois (immediately before the 2016 Joint Statistical Meetings, also in Chicago).

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

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 opens in April.

Call for presentations

Guidelines

All users are encouraged to submit abstracts, from which a subset will be chosen for either a short (15 minutes) or a long (25 minutes) presentation (both to be followed by 5 minutes for questions). Any topic related to Stata or relevant to Stata users is appropriate, including the following:

- New user-written commands for model estimation, graphing, data management, results reporting, or other purposes
- New approaches for using Stata to manage or analyze data, for programming Stata or Mata, or for using Stata together with other software or frameworks
- Innovative use or evaluations of existing Stata commands
- New analytic methods of particular relevance to Stata users
- Case studies of using Stata for specific applications or in specific disciplines or settings
- Methods and resources for teaching statistics with Stata or for teaching the use of Stata
- Comparisons of Stata to other software

Submissions

Please submit an abstract of no more than 200 words (ASCII text, no math symbols) using the web submission form at **stata.com/chicago16**. All abstracts must be received by **March 31, 2016**. Please include a short, informative title, and indicate whether you wish to be considered for a short (15-minute) or long (25-minute) presentation.

For complete guidelines, visit **stata.com/chicago16**.

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

#stata2016
stata.com/chicago16

2016 Stata Users Group meetings

EUSMEX 2016
Aguascalientes, Mexico | May 18

stata.com/meeting/mexico16

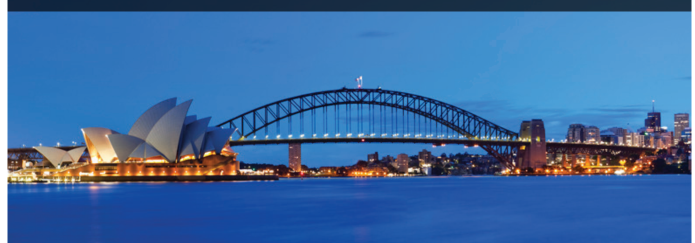
2016 Nordic and Baltic Stata Users Group meeting
Oslo, Norway | September 13

stata.com/meeting/nordic-and-baltic16

2016 German Stata Users Group meeting
Cologne, Germany | June 10

stata.com/meeting/germany16

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



stata.com/meeting/oceania16

2016 Belgium Stata Users Group meeting
Brussels, Belgium | September 6

stata.com/meeting/belgium16

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

stata.com/meeting/italy16

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

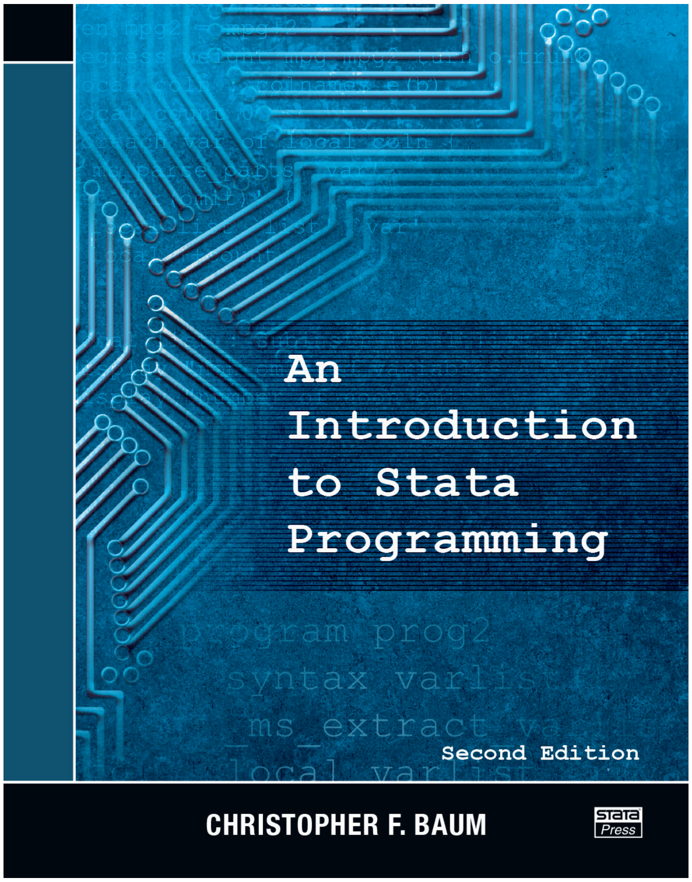
stata.com/meeting/uk16

Coming soon

More dates and locations are forthcoming. Check our site frequently for updates at **stata.com/meeting**, or sign up for an email alert at **stata.com/alerts**.

New from Stata Press

An Introduction to Stata Programming, Second Edition



Author:	Christopher F. Baum
Copyright:	2016
Pages:	412; paperback
Price:	$58.00

Christopher F. Baum's *An Introduction to Stata Programming, Second Edition*, is a great reference for anyone who wants to learn Stata programming.

Baum assumes readers have some familiarity with Stata, but readers who are new to programming will find the book accessible. He begins by introducing programming concepts and basic tools. More advanced programming tools such as structures and pointers and likelihood-function evaluators using Mata are gradually introduced throughout the book alongside examples.

This new edition reflects some of the most important statistical tools added since Stata 10. Of note are factor variables and operators, the computation of marginal effects, marginal means, and predictive margins using **margins**, the use of **gmm** to implement generalized method of moments estimation, and the use of **suest** for seemingly unrelated estimation.

As in the previous edition of the book, Baum steps the reader through the three levels of Stata programming. He starts with do-files. Do-files are powerful batch files that support loops and conditional statements and are ideal to automate your workflow as well as to guarantee reproducibility of your work.

He then delves into ado-files, which are used to extend Stata by creating new commands that share the syntax and behavior of official commands. Baum gives an example of how to write a command to calculate percentiles and the range of a variable, complete with documentation and certification.

After introducing the fundamentals of command development, Baum shows users how these concepts can be applied to help them write their own custom estimation commands by using Stata's built-in numerical maximum-likelihood estimation routine, **ml**, its built-in nonlinear least-squares routines, **nl** and **nlsur**, and its built-in generalized method of moments estimation routine.

Finally, he introduces Mata, Stata's matrix programming language. Mata programs are integrated into ado-files to build a custom estimation routine that is optimized for speed and numerical stability. Baum briefly discusses how ado-file programming concepts relate to Mata functions and objects. He also explains some of the advantages of using Mata for certain programming tasks.

Baum introduces concepts by providing the background and importance of the topic, presents common uses and examples, and then concludes with larger, more applied examples he refers to as "cookbook recipes".

Many of the examples are of particular interest because they arose from frequently asked questions from Stata users. If you want to understand basic Stata programming or want to write your own routines and commands using advanced Stata tools, Baum's book is a great reference.

Read the table of contents or order online:
stata-press.com/books/introduction-stata-programming.

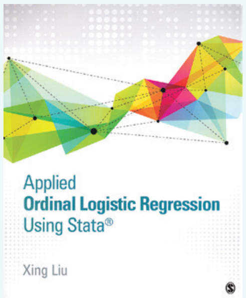
Want to know what's coming next from Stata Press? Sign up to be alerted when new titles are published:

stata-press.com/forthcoming

Stata® Press

New from the Stata Bookstore

Applied Ordinal Logistic Regression Using Stata

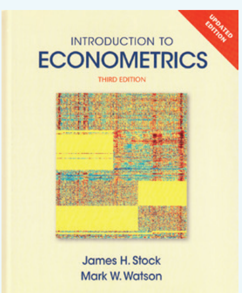


Author:	Xing Liu
Publisher:	Sage Publications
Copyright:	2016
Pages:	523; paperback
Price:	$58.75

Applied Ordinal Logistic Regression Using Stata by Xing Liu is an approachable introduction to ordinal logistic regression for students and applied researchers in education, the behavioral sciences, the social sciences, and related fields. Liu provides worked examples and the corresponding Stata commands. This book is a practical guide to understanding and implementing a variety of models for ordinal data.

Read more or order online:
stata.com/bookstore/applied-ordinal-logistic-regression-using-stata.

Introduction to Econometrics, Update, Third Edition

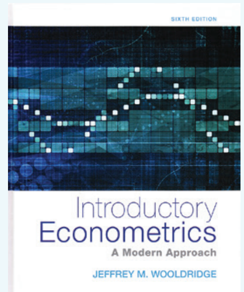


Authors:	James H. Stock and Mark W. Watson
Publisher:	Pearson Education
Copyright:	2015
Pages:	790; hardcover
Price:	$198.75

Introduction to Econometrics, Update, Third Edition, by James H. Stock and Mark W. Watson, is a real page-turner. By ingeniously introducing statistical methods as a means of answering four interesting empirical questions, the authors have written a rigorous text that makes you want to keep reading to find out how the story ends. The authors use the excitement generated by the questions as a springboard for an excellent introduction to estimation, inference, and interpretation in econometrics.

Read more or order online:
stata.com/bookstore/introduction-econometrics.

Introductory Econometrics: A Modern Approach, Sixth Edition



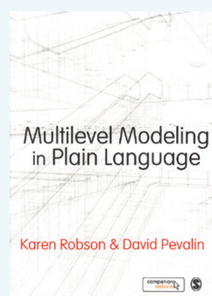
Author:	Jeffrey M. Wooldridge
Publisher:	Cengage Learning
Copyright:	2016
Pages:	789; hardcover
Price:	$209.75

The sixth edition of Jeffrey Wooldridge's textbook, *Introductory Econometrics: A Modern Approach*, lives up to its subtitle in its choice of topics and its treatment of standard material.

Wooldridge recognizes that modern econometrics involves much more than ordinary least squares (OLS) with extensions for autocorrelation, heteroskedasticity, and such. He does cover OLS in detail, but does so within the larger context of current techniques of estimation and inference for time-series data, panel data, limited dependent variables, and sample selection.

Read more or order online:
stata.com/bookstore/introductory-econometrics.

Multilevel Modeling in Plain Language



Authors:	Karen Robson and David Pevalin
Publisher:	Sage Publications
Copyright:	2016
Pages:	146; paperback
Price:	$32.00

Multilevel Modeling in Plain Language by Karen Robson and David Pevalin is a unique book on multilevel modeling. The book provides a conceptual understanding of multilevel models and the motivation for using them. The book includes many examples, all worked using Stata.

Read more or order online:
stata.com/bookstore/multilevel-modeling-in-plain-language.

NetCourses®

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.

March 4–April 15, 2016
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.

March 4–April 15, 2016
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 create concretely reproducible analyses. Learn how to speed up your work and do more complete analyses.

March 4–April 15, 2016
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 two-day 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

March 8–9, 2016
April 26–27, 2016

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.

Structural Equation Modeling Using Stata

March 10–11, 2016

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.

Handling Missing Data Using Multiple Imputation

April 21–22, 2016

Learn all aspects of multiple-imputation (MI) analysis, including creation of MI data using the multivariate normal and chained-equations (or fully conditional specification) imputation methods, manipulation of MI data, and analysis of MI data.

March

		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		

April

					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

stata.com/public-training

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.

The ICPSR Summer Program offers two four-week sessions that include a variety of statistical workshops and lectures, as well as workshops that examine the impact of various methodologies on specific substantive issues. Additionally, the four-week curriculum features nightly informal lectures presented by research scholars who have made important contributions to the development of social science methodology. The sessions take place in Ann Arbor, Michigan, on the University of Michigan campus.

The ICPSR Summer Program also offers numerous three- to five-day workshops on a variety of statistical topics and substantive issues. Many of these workshops take place in Ann Arbor, Michigan, as well as other locations across the United States and Canada.

Participants in each year's Summer Program generally represent about 25 different departments and disciplines from over 350 colleges, universities, and organizations around the world.

Three of this year's ICPSR courses are taught by StataCorp staff 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

Structural Equation Modeling with Stata

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

Multilevel and Mixed Models Using Stata

Rose Medeiros, Senior Statistician
July 27–29, 2016

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

THE STATA JOURNAL
Editors' Prize

The editors of the *Stata Journal*, H. Joseph Newton (Texas A&M University) and Nicholas J. Cox (Durham University), are delighted to announce the award of the Editors' Prize 2015 to Richard Williams.

The aim of the *Stata Journal* Editors' Prize is to reward contributions to the Stata community in respect of one or more outstanding papers published in the *Journal* in the previous three calendar years.

Richard Williams is being awarded based on exemplary clarity and care in exposition and on excellent programs in a key sector of statistical science greatly extending the functionality available to users.

Read more about Williams's outstanding contributions to the Stata community at **stata-journal.com/editors-prize/2015**.

Accepting nominations for 2016

Nominate your favorite *Stata Journal* authors for the *Stata Journal* Editors' Prize 2016.

Nominations should name the authors and one or more papers published in the *Stata Journal* in the previous three years and explain why the work is worthy of the prize. The rationale might include originality, depth, elegance, or unifying power of work; usefulness in cracking key problems or allowing important new methodologies to be widely implemented; and clarity or expository excellence of the work.

The prize will consist of a framed certificate and an honorarium of U.S. $1,000, courtesy of the publisher of the *Stata Journal*.

Nominations should be sent as a private email to **editors@stata-journal.com** by July 31, 2016. Nominations will be considered confidential both before and after the awarding of the prize. Neither anonymous nor public nominations will be accepted.

For a complete list of rules and to view all previous winners, visit

stata-journal.com/editors-prize