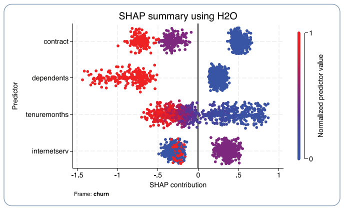
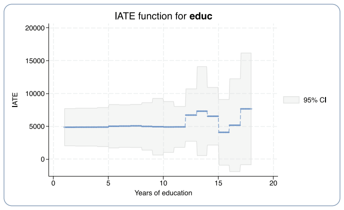
New in STATA 19



Machine learning via H2O: Ensemble decision trees

Employ machine learning to uncover insights from the data when traditional statistical models fall short. Use gradient boosting machine (GBM) and random forest to perform regression or classification.



Conditional average treatment effects

Go beyond estimating overall treatment effects in your causal analysis to estimating individualized and group-specific ones. Compare different interventions and policies. Explore treatment-effects heterogeneity.

```
regress, absorb(z1 z2 z3)
xtreg, absorb(z1 z2 z3) fe
ivregress 2sls, absorb(z1 z2 z3)
```

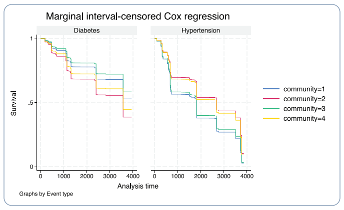
High-dimensional fixed effects

Absorb not just one but multiple high-dimensional categorical variables in your linear, fixed-effects linear, and instrumental-variables linear models, and enjoy remarkable speed gains!

```
Bayesian variable selection                         MCMC iterations  =     12,500
Metropolis-Hastings and Gibbs sampling              Burn-in          =      2,500
                                                    MCMC sample size =     10,000
Global-local shrinkage coefficient prior:           Number of obs    =        441
  Horseshoe(1)                                      Acceptance rate  =      .8633
                                                    Efficiency:  min =      .1516
                                                                 avg =        .44
Log marginal-likelihood = -475.56099                             max =          1

                                                     Equal-tailed       Inclusion
    diabetes       Mean   Std. dev.      MCSE   [95% cred. interval]       coef.

      serum5   .3366779   .0671548   .0013409   .2182886   .4861579        0.73
         bmi   .3284608   .0414561   .0004504    .247525   .4087023        0.72
          bp   .1900507   .0411339   .0004625   .1096927   .2718276        0.59
         sex  -.1279497   .0398828   .0005348  -.2048249  -.0483072        0.50
      serum1   -.125876    .127556   .0032762  -.4417064   .0527829        0.44
      serum3  -.0929577   .0768004    .001911  -.2375419    .046901        0.42
      serum4    .048891   .0733476   .0017154  -.0749584   .2131595        0.32
      serum2   .0188206   .1037191   .0023902  -.1479433   .2834479        0.32
      serum6   .0300323    .036144   .0006945  -.0308045   .1095077        0.26
         age   -.001932   .0282981    .000283  -.0620468   .0574765        0.20
```

Bayesian variable selection

Select variables in a linear regression and account for variable-selection uncertainty with Bayesian variable selection. Choose between global–local shrinkage and spike-and-slab priors for regression coefficients, and perform Bayesian inference.

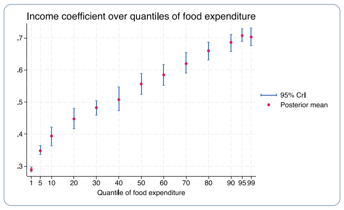


Interval-censored multiple-event Cox model

Need to analyze event times from multiple events, such as the onsets of diabetes and hypertension?

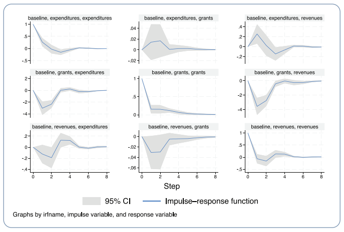
Don't know the exact event times?

Use the marginal interval-censored multiple-event Cox model.



Bayesian quantile regression

Use Bayesian quantile regression to obtain full posterior distributions of quantile regression coefficients for comprehensive inference, including model-based "standard errors".



Panel-data VAR model

Fit a panel-data vector autoregressive (VAR) model to analyze the trajectories of related variables when you observe multiple units or panels over time.

```
xtreg, re + xtreg, fe = xtreg, cre
```

Correlated random-effects model

Want coefficient estimates of time-invariant covariates in your panel-data model? Fit an RE model. Want to allow for correlation between covariates and unobserved panel-level effects? Fit an FE model. Want both? With **xtreg, cre**, you can now fit a CRE model.

```
Bayesian bootstrap
Observation prior: Improper

Linear regression                                   Number of obs =      74
                                                    Replications  =      50
                                                    Wald chi2(1)  =  109.21
                                                    Prob > chi2   =  0.0000
                                                    R-squared     =  0.6515
                                                    Adj R-squared =  0.6467
                                                    Root MSE      =  3.4389

                          Bayesian
                Observed  bootstrap                          Normal-based
         mpg coefficient  std. err.      z    P>|z|     [95% conf. interval]

      weight   -.0060087    .000575   -10.45   0.000    -.0071356   -.0048817
       _cons    39.44028   1.866948    21.13   0.000     35.78113    43.09943
```

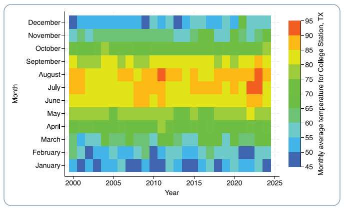
Bayesian bootrstrap

Perform Bayesian bootstrap to obtain more precise parameter estimates in small samples and incorporate prior information when sampling observations. Use it with official commands or community-contributed commands.

```
. cfregress rent pcturban (hsngval = faminc i.region, interact(faminc))

Control-function linear regression                  Number of obs =      50
                                                    Wald chi2(2)  =   95.16
                                                    Prob > chi2   =  0.0000
                                                    R-squared     =  0.5945
                                                    Root MSE      = 22.2851
Endogenous variable model:
    Linear: hsngval

        rent  Coefficient  Std. err.      z    P>|z|     [95% conf. interval]

rent
     hsngval    2.155381   .3437284     6.27   0.000     1.481686    2.829076
    pcturban    .4794597   .2362242     2.03   0.042     .0164688    .9424506
       _cons    98.15909   13.86958     7.08   0.000     70.97521     125.343
e.rent
  cf(hsngval)   10.66765   3.619442     2.95   0.003     3.573673    17.76163

  cf(hsngval)
      faminc   -.5610651   .1743049    -3.22   0.001    -.9026965   -.2194338
Instruments for hsngval: faminc 2.region 3.region 4.region
```

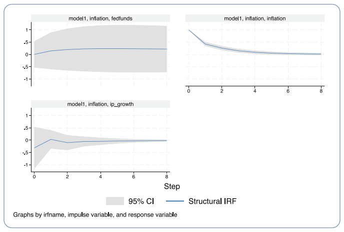
Control-function linear and probit models

Fit control-function linear and probit models, with continuous, binary, fractional, and count endogenous variables. Easily test for endogeneity.



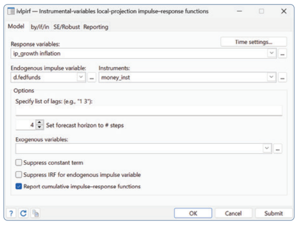
Graphics: Bar graph CIs, heat maps, ...

Enjoy exciting new graphical features such as heat maps, spike and capped-spike range plots, bar graphs with CIs, and more!



SVAR models via instrumental variables

Use instruments instead of short-run constraints to estimate dynamic causal effects.



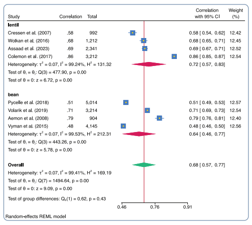
IV local-projection IRFs

Account for endogeneity when using local projections to estimate dynamic causal effects.

Classes	BIC	LMR (P>LMR)	Class marginal probabilities (SE)		
1	1,108.80		1.00 (0.00)		
2	1,057.31	75.55 (<0.001)	0.72 (0.06)	0.28 (0.06)	
3	1,081.86	2.25 (0.687)	0.16 (15.31)	0.63 (11.94)	0.21 (3.37)

Latent class model-comparison statistics

Easily compare latent class models with varying numbers of latent classes. Construct and export publication-quality tables comparing models.



Meta-analysis for correlations

Perform meta-analysis for correlations just like you already do for two-sample means and proportions. Use all standard features such as forest plots and subgroup analysis.

```
. estat weakrobust

Weak-instrument-robust test
Model VCE: Robust

 ( 1)  hsngval = 0

Cond. likelihood-ratio (CLR) test =   5.48
                       Prob > CLR = 0.0253

Notes: CLR test reported by default because
       model is overidentified.
       p-value computed by simulation
       (25,000 replications).
```

Inference robust to weak instruments

Do you have weak instruments in your instrumental-variables regression? Use tests robust to weak instruments to perform reliable inference on endogenous regressors.

```
RE? FE? CRE?

. xtreg ...
. mundlak
```

Mundlak specification test

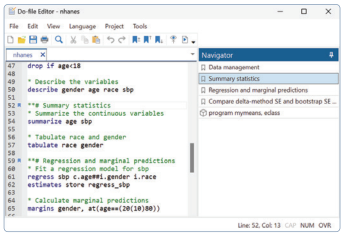
Use the Mundlak specification test to choose between random-effects and fixed-effects or correlated random-effects models, even with cluster–robust, bootstrap, or jackknife standard errors.

```
Bayesian asymm. Laplace (quantile) regression    MCMC iterations  =     15,000
Random-walk Metropolis–Hastings sampling         Burn-in          =      5,000
                                                 MCMC sample size =     10,000
                                                 Number of obs    =        887
                                                 Acceptance rate  =       .339
                                                 Efficiency:  min =     .01429
                                                              avg =     .06357
Log marginal-likelihood                                       max =      .2034
```

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
math5_q25						
math3	.781409	.0331454	.001575	.7818769	.7174074	.844201
U1	1	0	0	1	1	1
_cons	27.11279	.4114259	.03442	27.09728	26.33384	27.93378
math5_q75						
math3	.4021812	.0285305	.0013	.4029019	.344931	.4566547
U2	1	0	0	1	1	1
_cons	34.21954	.2754823	.018921	34.22476	33.66447	34.76017
scale	1.557948	.0373122	.000827	1.557052	1.486446	1.632176
var_U1	7.793937	2.041174	.076138	7.560785	4.519766	12.38803
var_U2	2.256973	.7545756	.036895	2.153597	1.128305	4.057043

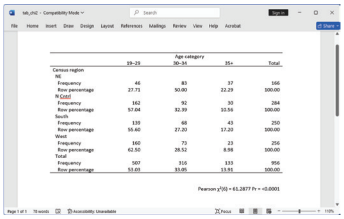
Bayesian asymmetric Laplace model

Go beyond classic quantile regression by fitting Bayesian simultaneous, multilevel, and nonlinear quantile regression models.



Do-file Editor: Autocompletion, templates, ...

Navigator panel, file templates, code folding improvements, word and selection highlighting, more autocompletion, temporary bookmarks, and much more.



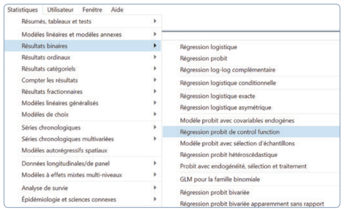
Tables: Easier tabulations, exporting, ...

Create a table, customize it with a title and notes, and export it in one command. Easily collect and customize tabulations with measures of association, tabulations of survey data, and ANOVA tables.

```
frames modify, add(myframe)
frames modify, drop(myframe)
```

Multiple datasets: Modify a set of frames

If you work with multiple datasets in memory or frames, you can now modify a frameset file without loading it into memory: add frames in memory to it or drop frames from it.



Stata in French

All of Stata's interface—all menus and all dialogs—is now available in French.

More

- Alternative at-risk table for survival graphs
- PyStata enhancements
- Log-scale analysis of bioequivalence
- Robust standard errors for VAR models
- Bayesian predictions in user-defined evaluators
- Mata functions for least-squares solvers, labels, and more

stata.com/new