New in **Stata 16**

**Lasso—variable selection, prediction, inference**

All the tools you expect for lasso machine learning

- Lasso, square-root lasso, and elastic net
- Cross-validation
- Adaptive lasso
- Knot analysis
- Coefficient paths

Alongside cutting-edge inferential methods

- Robust to mistakes in variable selection
- Proper inference for coefficients of interest
- Double selection
- Partialing out
- Cross-fit partialing out
- Double machine learning

Select predictors for continuous, binary, and count outcomes

**Lasso with selection via cross-validation**

- `. lasso linear  y x1-x1000`
- `. lasso logit   y x1-x1000`
- `. lasso probit  y x1-x1000`
- `. lasso poisson y x1-x1000`

**Adaptive lasso**

- `. lasso linear y x1-x1000, selection(adaptive)`

Selection via plugin method

- `. lasso linear y x1-x1000, selection(plugin)`

**Elastic net with selection via cross-validation**

- `. elasticnet linear  y x1-x1000`
- `. elasticnet logit   y x1-x1000`
- `. elasticnet probit  y x1-x1000`
- `. elasticnet poisson y x1-x1000`

**Square-root lasso**

- `. sqrtlasso y x1-x1000`

Examine the results

- View selected variables
  - `. lassoknots`
  - `. lassoinfo`
  - `. lassocoef`

- Plot cross-validation function
  - `. cvplot`

- Plot coefficient path
  - `. coefpath`

- Obtain predictions
  - `. use newdata`
  - `. predict yhat`

- Evaluate fit
  - `. lassogof`
Lasso for inference

With lasso inferential methods, you can estimate coefficients, standard errors, test statistics, and confidence intervals for variables of interest while using lassos to select from a potentially large number of control variables.

- Double-selection method; estimate coefficients for $x_1$ and categorical $x_2$; selection of controls via plugin
  . dsregress y x1 i.x2, controls(cl-cl1000)

- Logit model for binary outcome; estimate odds ratios for $x_1$ and $x_2$
  . dologit y x1 i.x2, controls(cl-cl1000)

- Poisson model for count outcome; estimate incidence-rate ratios for $x_1$ and $x_2$
  . dspoisson y x1 i.x2, controls(cl-cl1000)

- Selection of controls via cross-validation
  . dsregress y x1 i.x2, controls(cl-cl1000) selection(cv)

- Partialing-out method
  . poregress y x1 i.x2, controls(cl-cl1000)

- Cross-fit partialing-out method (double machine learning)
  . xporegress poregress y x1 i.x2, controls(cl-cl1000)

Evaluate results using Stata’s standard tools

- Perform tests on coefficients
  . test x1=1

- Estimate contrasts such as differences across levels
  . contrast ar.x2

- View selected controls in the lasso for $y$
  . lassocoef (., for(y))

- Plot coefficient path in the lasso for $y$
  . coefpath, for(y)

Stata is a registered trademark of StataCorp LLC, 4905 Lakeway Drive, College Station, TX 77845, USA.