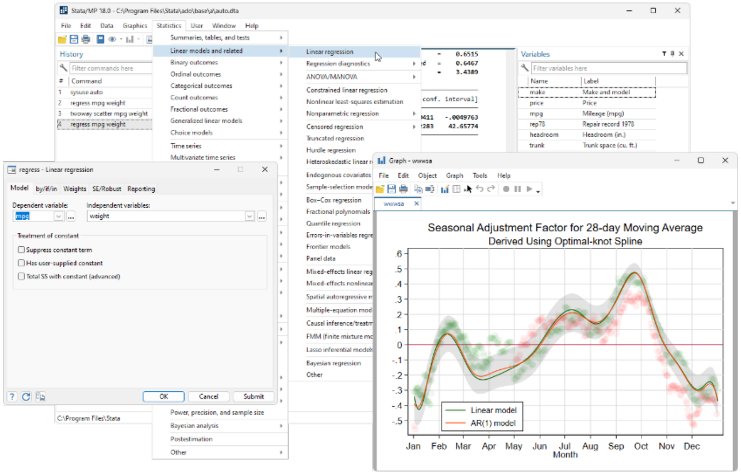
WHAT'S NEW

Stata 18 contains an exciting array of new features, amongst which the following may be of interest:

- BAYESIAN MODEL AVERAGING
- CAUSAL MEDIATION ANALYSIS
- TABLES OF DESCRIPTIVE STATISTICS
- HETEROGENEOUS DID
- GROUP SEQUENTIAL DESIGNS
- MULTILEVEL META-ANALYSIS
- META-ANALYSIS FOR PREVALENCE
- ROBUST INFERENCE FOR LINEAR MODELS
- WILD CLUSTER BOOTSTRAP
- LOCAL PROJECTIONS FOR IRFS
- FLEXIBLE DEMAND SYSTEM MODELS
- TVCS WITH INTERVAL-CENSORED COX MODEL
- LASSO FOR COX MODEL
- RERI
- IV QUANTILE REGRESSION
- IV FRACTIONAL PROBIT MODEL
- ALIAS VARIABLES ACROSS FRAMES
- DATA EDITOR ENHANCEMENTS
- DO-FILE EDITOR ENHANCEMENTS
- ALL-NEW GRAPH STYLE

Other new features include:

- Corrected and consistent AIC,
- Model selection for ARIMA and ARFIMA
- GOF plots for survival models
- New spline functions
- Graph colors by variable
- Create, load, and save sets of frames
- Boost-based regular expressions,
- Vectorized numerical integration, and
- New reporting features in putdocx, putexcel, and putpdf.



Stata/MP 18.0 - C:\Program Files\Stata\ado\base\a\auto.dta

History

```
#  Command
1  sysuse auto
2  regress mpg weight
3  twoway scatter mpg weight
4  regress mpg weight
```

```
     =    0.6515
ed   =    0.6467
     =    3.4389

conf. interval]
411    -.0049763
283     42.65774
```

Variables

Name	Label
make	Make and model
price	Price
mpg	Mileage (mpg)
rep78	Repair record 1978
headroom	Headroom (in.)
trunk	Trunk space (cu. ft.)

regress - Linear regression

Dependent variable: mpg

Independent variables: weight

Treatment of constant

☐ Suppress constant term

☐ Has user-supplied constant

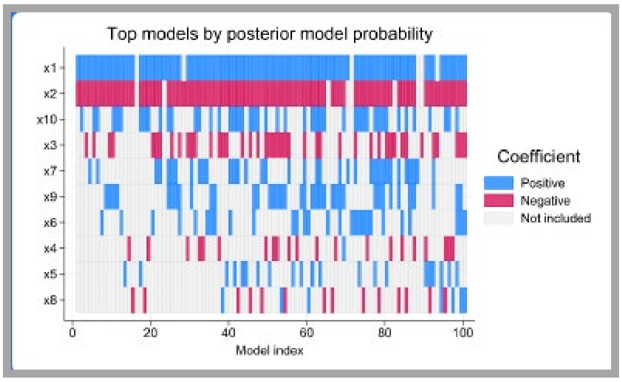
☐ Total SS with constant (advanced)

C:\Program Files\Stata

Graph - wwwsa

WWW.TSTAT.IT | WWW.TSTATTRAINING.EU

BAYESIAN MODEL AVERAGING (BMA)



Traditionally, researchers choose a model and performs analysis based on this model. The results are conditional on the chosen model. In the presence of multiple plausible models, this approach may not however, be reliable. Model averaging allows one to perform analysis based on multiple models and thus account for model uncertainty in the results. BMA accounts for model uncertainty according to Bayesian principles, which can be applied universally to any data analysis. In the regression setting, model uncertainty describes the uncertainty about which predictors should be included in a regression model.

The new command **bmaregress** performs BMA for linear regression and can be used for inference, prediction, and, if desired, even model selection.

For instance,

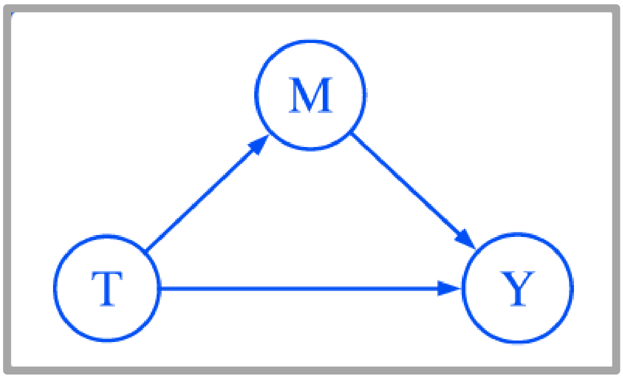
```
.bmaregress y x1 x2
```

considers all four possible models for outcome **y** that include or exclude predictors **x1** and **x2** and combines these models according to how likely each model is based on the observed data. Users can choose from a variety of prior distributions to explore the effect of assumptions about a model's and predictors' importance on the results. Postestimation commands allow users to estimate the probability of a model, identify important predictors, explore model complexity, obtain predictive means, evaluate predictive performance, and perform inference on regression coefficients.

CAUSAL MEDIATION ANALYSIS

Causal inference aims to identify and quantify the causal effect of a treatment on an outcome. In causal mediation analysis, researchers aim to further explore how this effect arises. Maybe exercise increases the level of a hormone that, in turn, increases well-being. Maybe an import quota increases the market power of local companies which, in turn, increases the prices of goods.

Relationships like these are often visualized with a causal diagram, for example,



With the new **mediate** command, one can estimate the total effect of a treatment on an outcome and decompose it into direct effects and indirect effects (via a mediator such as hormone level). In fact, multiple types of decompositions can be computed, depending on the hypothesis of interest. In addition, **estat proportion** reports the proportion of the total effect that occurs through the mediator.

mediate is very flexible–the outcome can be continuous, binary, or count; the mediator can be continuous, binary, or count; and the treatment can be binary, multivalued, or continuous.

The **mediate** command is very flexible. It supports 24 combinations of models for the outcome and mediator, so it can be applied to many situations that arise in real research.

TABLES OF DESCRIPTIVE STATISTICS

The new **dtable** command creates a table of descriptive statistics.

dtable reports summary statistics for continuous and categorical factor variables. Users can select which statistics they would like to report for each variable; select from the mean, standard deviation, median, interquartile range, percentage, proportion, and many others. One can also easily compare statistics across categories of another variable.

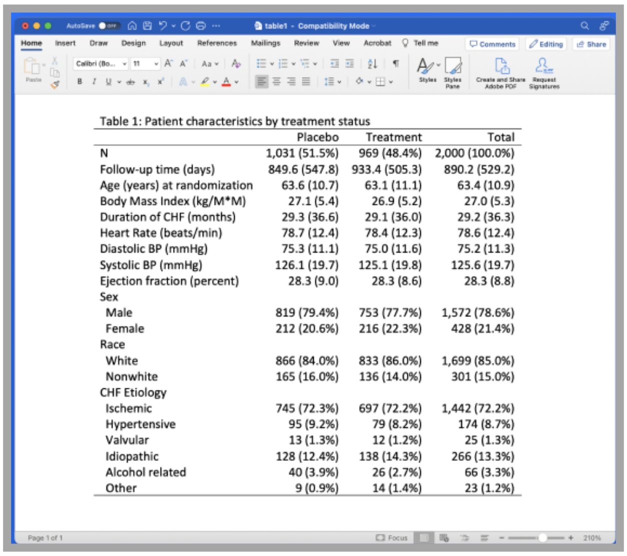


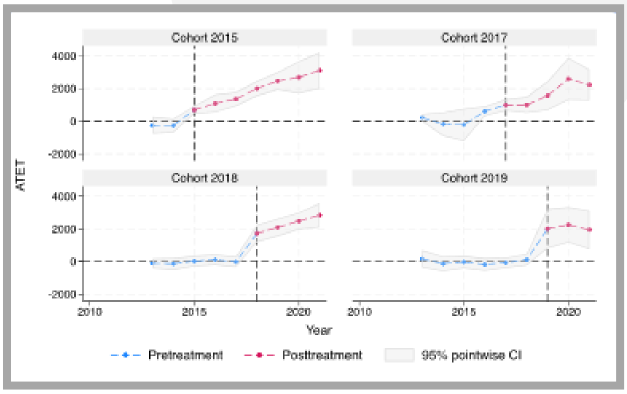
Table 1: Patient characteristics by treatment status

	Placebo	Treatment	Total
N	1,031 (51.5%)	969 (48.4%)	2,000 (100.0%)
Follow-up time (days)	849.6 (547.8)	933.4 (505.3)	890.2 (529.2)
Age (years) at randomization	63.6 (10.7)	63.1 (11.1)	63.4 (10.9)
Body Mass Index (kg/M*M)	27.1 (5.4)	26.9 (5.2)	27.0 (5.3)
Duration of CHF (months)	29.3 (36.6)	29.1 (36.0)	29.2 (36.3)
Heart Rate (beats/min)	78.7 (12.4)	78.4 (12.3)	78.6 (12.4)
Diastolic BP (mmHg)	75.3 (11.1)	75.0 (11.6)	75.2 (11.3)
Systolic BP (mmHg)	126.1 (19.7)	125.1 (19.8)	125.6 (19.7)
Ejection fraction (percent)	28.3 (9.0)	28.3 (8.6)	28.3 (8.8)
Sex			
Male	819 (79.4%)	753 (77.7%)	1,572 (78.6%)
Female	212 (20.6%)	216 (22.3%)	428 (21.4%)
Race			
White	866 (84.0%)	833 (86.0%)	1,699 (85.0%)
Nonwhite	165 (16.0%)	136 (14.0%)	301 (15.0%)
CHF Etiology			
Ischemic	745 (72.3%)	697 (72.2%)	1,442 (72.2%)
Hypertensive	95 (9.2%)	79 (8.2%)	174 (8.7%)
Valvular	13 (1.3%)	12 (1.2%)	25 (1.3%)
Idiopathic	128 (12.4%)	138 (14.3%)	266 (13.3%)
Alcohol related	40 (3.9%)	26 (2.7%)	66 (3.3%)
Other	9 (0.9%)	14 (1.4%)	23 (1.2%)

Tables created by **dtable** can be customized in many ways–statistics to be reported, numeric and string formats, notes, titles, labels, and more. The table can be exported directly to Microsoft Word, Microsoft Excel, HTML, Markdown, PDF, LaTeX, SMCL, or plain text.

dtable makes it easy to create what is commonly called a "Table 1" – the first table included in almost every research paper.

HETEROGENEOUS DIFFERENCE IN DIFFERENCES (DID)



DID models are used to estimate the average treatment effect on the treated (ATET) with repeated-measures data. A treatment effect can be an effect of a drug regimen on blood pressure or an effect of a training program on employment. Unlike with the standard cross-sectional analysis, provided by the existing **teffects** command, DID analysis controls for group and time effects when estimating the ATET, where groups identify repeated measures.

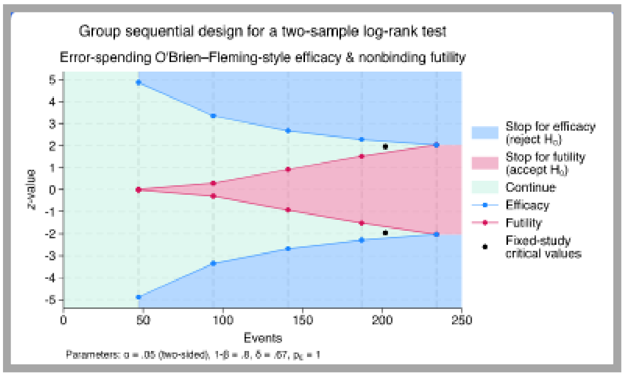
Heterogeneous DID models additionally account for variation in treatment effects arising from groups being treated at different points in time and effects varying over time within groups.

Suppose that several school districts introduce an exercise and a nutrition program to improve students' health outcomes. Different school districts introduce the program at different points in time. Is it sensible to assume the effect of the program on student's health outcomes does not change over time and is the same regardless of when the program was adopted? Maybe not! One can use heterogeneous DID models to account for the potential differences in effects.

The new commands **hdidregress** and **xthdidregress** fit heterogeneous DID models. **hdidregress** works with repeated-cross-sectional data, and **xthdidregress** works with both longitudinal/panel data.

GROUP SEQUENTIAL DESIGNS (GSDs)

GSDs are types of adaptive design that allow researchers to stop a trial early if they find compelling evidence that a treatment is effective or ineffective. Suppose one would like to design a study to test whether a type of chemeotherapy is effective for treating tumours and that one expects data to be collected over a few years' time. Rather than performing one analysis once all the data have been collected, GSDs allow users to perform interim analyses as the data are collected. Each interim analysis provides the opportunity to stop the trial or continue collecting data. The trial can be stopped early if there is strong evidence of efficacy. The trial can also be stopped early if there is strong evidence of futility; this avoids exposing additional participants to an inadequate treatment.



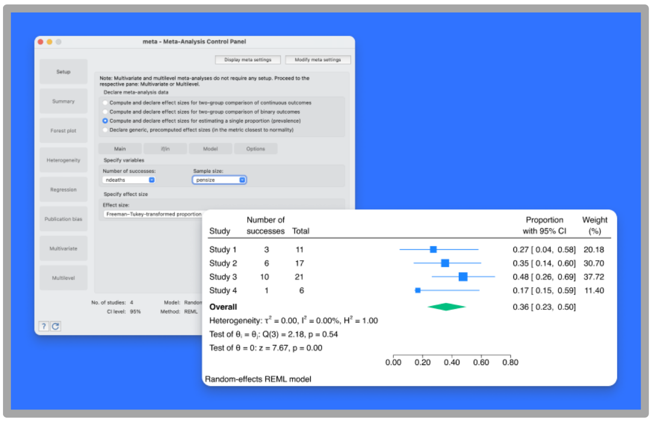
Stata 18 offers a suite of commands for GSDs. The new **gsbounds** command calculates efficacy and futility bounds based on the number of analyses (also called looks), the desired overall Type I error, and the desired power. Users can select from seven boundary-calculation methods—choose whether they want classical or error-spending methods and whether they want more conservative or less conservative bounds for early analyses. The new **gsdesign** command calculates efficacy and futility boundaries and provides sample sizes for the interim and final analyses for tests of means, proportions, and survivor functions. Graphs make it easy to visualize the boundaries across all interim and final analyses.

META-ANALYSIS FOR PREVALENCE

The **meta esize** command performs meta-analysis of two-sample binary or continuous data. Now, it also performs meta-analysis of one-sample binary data, also known as meta-analysis of proportions or meta-analysis of prevalence.

These types of data commonly appear in meta-analysis studies when pooling results from studies that each estimate one proportion. For instance, you may have studies reporting the prevalence of a particular disease or the proportion of students who drop out of high school. In this setting, effect sizes such as Freeman–Tukey-transformed proportions or logit-transformed proportions are typically used in the meta-analysis.

After **meta esize**, one can use other commands in the **meta** suite for further analysis. For instance, researchers can create a forest plot with **meta forestplot**, perform subgroup analysis by adding the **subgroup()** option to **meta forestplot**, summarize meta-analysis data with **meta summarize**, or construct a funnel plot with **meta funnelplot**.



Note: Multivariate and multilevel meta-analyses do not require any setup. Proceed to the respective pane: Multivariate or Multilevel.

Declare meta-analysis data

○ Compute and declare effect sizes for two-group comparison of continuous outcomes
○ Compute and declare effect sizes for two-group comparison of binary outcomes
● Compute and declare effect sizes for estimating a single proportion (prevalence)
○ Declare generic, precomputed effect sizes (in the metric closest to normality)

Specify variables

Number of successes: ndeaths

Sample size: pensize

Specify effect size

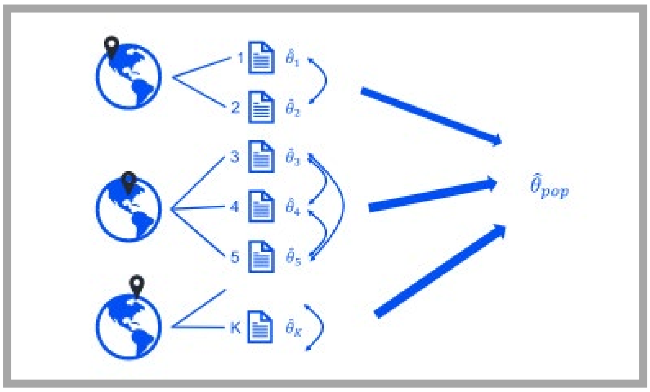
Effect size: Freeman–Tukey-transformed proportion

No. of studies: 4 Model: Random

CI level: 95% Method: REML

```
Study      Number of    Total              Proportion        Weight
           successes                       with 95% CI        (%)
Study 1        3          11         0.27 [ 0.04,  0.58]    20.18
Study 2        6          17         0.35 [ 0.14,  0.60]    30.70
Study 3       10          21         0.48 [ 0.26,  0.69]    37.72
Study 4        1           6         0.17 [ 0.15,  0.59]    11.40
Overall                              0.36 [ 0.23,  0.50]
Heterogeneity: τ² = 0.00, I² = 0.00%, H² = 1.00
Test of θi = θj: Q(3) = 2.18, p = 0.54
Test of θ = 0: z = 7.67, p = 0.00
Random-effects REML model
```

MULTILEVEL META-ANALYSIS



When researchers want to analyse results from multiple studies, they use meta-analysis to combine results and estimate an overall effect size. The existing meta suite is used to perform standard and multivariate meta-analysis.

Sometimes the reported effect sizes are nested within higher-level groupings, such as geographical locations (states or countries) or administrative units (school districts). Effect sizes within the same groups (for example, districts) are likely to be similar and thus dependent. In this case, you can use multilevel meta-analysis. The goal of multilevel meta-analysis is to not only synthesize an overall effect size but also account for this dependence and assess the variability among the effect sizes at different hierarchical levels. The new estimation commands **meta meregress** and **meta multilevel** are used to perform multilevel meta-analysis.

For example, in studies reporting effects (mean differences) of two teaching methods on math test scores, **y**, and sampling standard errors of **y** in **se**. The effect sizes are nested within schools, and schools are nested in districts. Users can fit a three-level random intercepts model with

.meta meregress y || district: || school:, essevariable(se)

or

.meta multilevel y, relevels(district school) essevariable(se)

If users have covariates and want to include random slopes, they can use **meta meregress:**

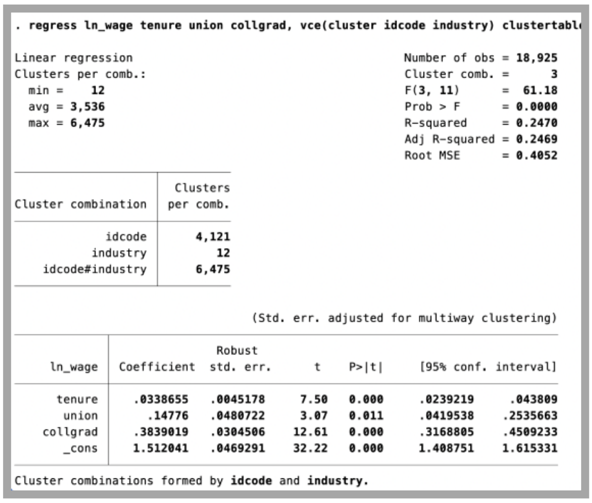
.meta meregress y x1 x2 || district: x1 x2 || school:, essevariable(se)

After fitting the model, postestimation commands are available for computing multilevel heterogeneity statistics, displaying estimated random-effects covariance matrices, and more.

Note: The syntax is the simplest of any package available. **meta meregress** is also the most flexible in terms of the of constraints that can be applied to the random effects.

ROBUST INFERENCE FOR LINEAR MODELS

Reliable standard errors are crucial to drawing appropriate inferences in research. *Stata* 18 offers new ways to obtain standard errors and confidence intervals for linear models fit with **regress**, **areg**, and **xtreg, fe**. The new methods aim to provide better inference when large-sample approximations do not work well. If researchers have clustered data with only a few clusters or with an uneven number of observations per cluster, they can now add the **vce(hc2 clustvar)** option to obtain HC2 cluster–robust standard errors. If there is more than one variable that identifies clusters in the data. Users can now add the **vce(cluster clustvar1 clustvar2** …) option to obtain multiway cluster standard errors.



```
. regress ln_wage tenure union collgrad, vce(cluster idcode industry) clustertabl

Linear regression                                    Number of obs = 18,925
Clusters per comb.:                                  Cluster comb. =      3
  min =     12                                       F(3, 11)      =  61.18
  avg = 3,536                                        Prob > F      = 0.0000
  max = 6,475                                        R-squared     = 0.2470
                                                     Adj R-squared = 0.2469
                                                     Root MSE      = 0.4052

-------------------------------
                   |  Clusters
Cluster combination|  per comb.
-------------------+-----------
             idcode|      4,121
           industry|         12
    idcode#industry|      6,475
-------------------------------

                              (Std. err. adjusted for multiway clustering)
---------------------------------------------------------------------------
             |               Robust
     ln_wage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+-------------------------------------------------------------
      tenure |   .0338655   .0045178     7.50   0.000     .0239219     .043809
       union |     .14776   .0480722     3.07   0.011     .0419538    .2535663
    collgrad |   .3839019   .0304506    12.61   0.000     .3168805    .4509233
       _cons |   1.512041   .0469291    32.22   0.000     1.408751    1.615331
---------------------------------------------------------------------------
Cluster combinations formed by idcode and industry.
```

WILD CLUSTER BOOTSTRAP

The wild cluster bootstrap provides another new option for robust inference when researchers have data with a few clusters, an uneven number of observations across clusters, or both.

The new **wildbootstrap** command computes wild cluster bootstrap *p*-values and confidence intervals for tests of simple and composite linear hypotheses about parameters from linear regression models. Users can type

```
. wildbootstrap regress y x1 x2 …
```

or

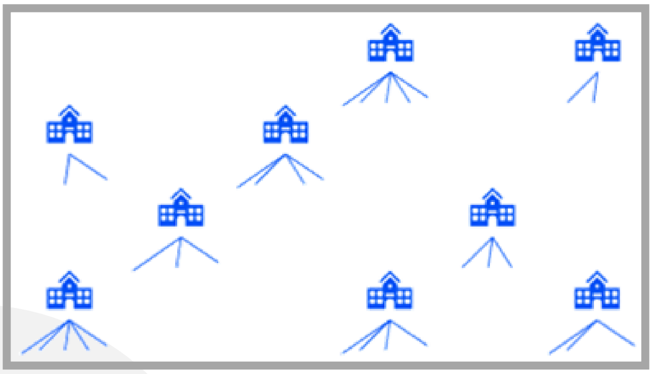
```
. wildbootstrap areg y x1 x2 …, absorb(x3)
```

or

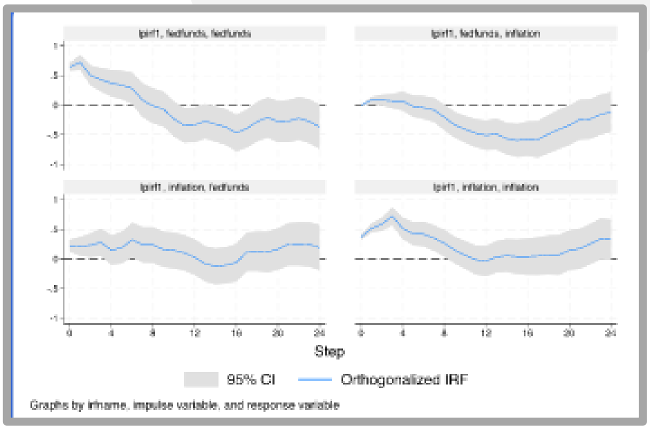
```
. xtset id
. wildbootstrap xtreg y x1 x2 …
```

to fit a linear regression model, a linear regression model with a large dummy-variable set, or a fixed-effects linear regression model for panel data, respectively, and to obtain the wild cluster bootstrap statistics.

This pairs well with the new standard errors options now available offering users many new tools for robust inference in linear models.



LOCAL PROJECTIONS FOR IMPULSE-RESPONSE FUNCTIONS (IRFs)



The new **lpirf** command provides local projections of IRFs. Local projections are used in time-series analysis to estimate the effect of shocks on outcome variables. For instance, researchers might evaluate the effect of an unexpected change in interest rates on a country's output and inflation rate. Type

```
. lpirf y1 y2
```

to obtain local projection estimates of IRFs for **y1** and **y2**. One can add the **exog()** option to estimate dynamic multipliers, which are responses of endogenous variables to a shock to an exogenous variable.

The new **lpirf** command works seamlessly with the existing **irf** commands, allowing users to create graphs and tables of IRFs, orthogonalized IRFs, and dynamic multipliers.

As with the linear models mentioned above, robust standard errors are often important in IRF estimation. Robust and Newey–West standard errors are available.

Local projections for IRFs provide an alternative to IRFs based on a vector autoregressive (VAR) model. Local projections are not constrained by a model; thus, they provide more flexible IRF coefficients. Local projections also allow for easier hypothesis testing.

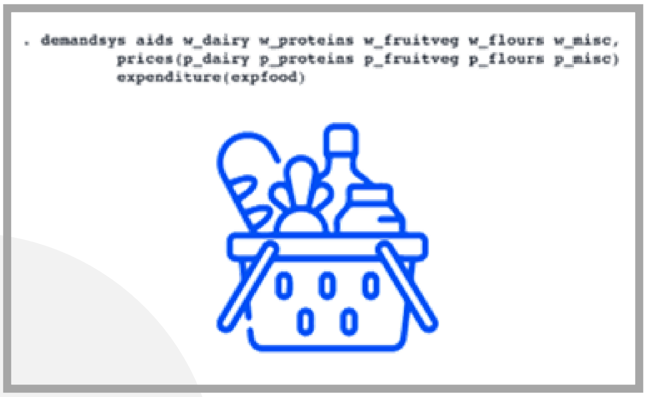
FLEXIBLE DEMAND SYSTEM MODELS

Often, researchers are interested in estimating demand for a basket of goods. The new **demandsys** command provides extensive tools for computing demand and measuring how sensitive demand for goods is to price and expenditure changes by calculating their corresponding elasticities.

Users can use **demandsys** to fit eight different demand system models:

- Linear expenditure system
- Basic translog
- Generalized translog
- Almost ideal demand
- Generalized almost ideal
- Quadratic almost ideal
- Generalized quadratic almost ideal

With the **estat elasticities** command, one can estimate various elasticities–expenditure elasticities, uncompensated own-price and cross-price elasticities, and compensated own-price and cross-price elasticities–to explore how sensitive demand is to changes in prices and expenditures. With eight demand systems to choose from, the **demandsys** command gives researchers much flexibility to choose the demand system technique that aligns with their empirical assumptions.



```
. demandsys aids w_dairy w_proteins w_fruitveg w_flours w_misc,
        prices(p_dairy p_proteins p_fruitveg p_flours p_misc)
        expenditure(expfood)
```

ALIAS VARIABLES ACROSS FRAMES

Since *Stata* 16, *Stata* has supported multiple datasets in memory. Each dataset resides in a frame. When datasets are related, users can link their frames by using the **frlink** command and identifying the variables that match the observations in the current frame with observations in the related frame.

In *Stata* 18, users can use the new **fralias add** command to create alias variables across linked frames and easily perform analyses using variables stored in separate frames.

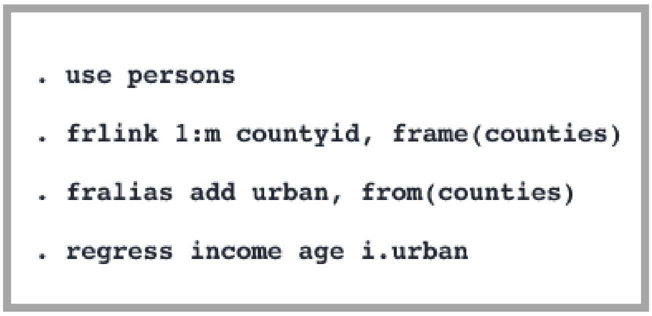
Alias variables behave as if they were copied from one frame into another, but because they are stored in the original frame, they take up very little memory. To see how easy alias variables are to use, say that **y** is a variable in the current frame and that **x** is a variable from a linked frame named **frame2**. To create an alias to **x** in the current frame, type

```
.fralias add x, from(frame2)
```

One can then fit a regression by typing

```
.regress y x
```

just as if **x** were stored in the current frame.



```
. use persons

. frlink 1:m countyid, frame(counties)

. fralias add urban, from(counties)

. regress income age i.urban
```

TVCs WITH INTERVAL-CENSORED COX MODEL

In event-time data, interval-censoring occurs when the time to an event of interest, such as recurrence of cancer, is not directly observed, but is known to lie within an interval. The existing **stintcox** command fits semiparametric interval-censored Cox proportional hazards models. In *Stata* 18, **stintcox** allows time-varying covariates.

stintcox now supports multiple-record-per-subject interval-censored data, which include a record for each examination time for each subject. This format makes it easy to accommodate time-varying covariates; the data record the values of the covariates at each examination time. Multiple-record-per-subject data also provide a convenient way to specify current status data.

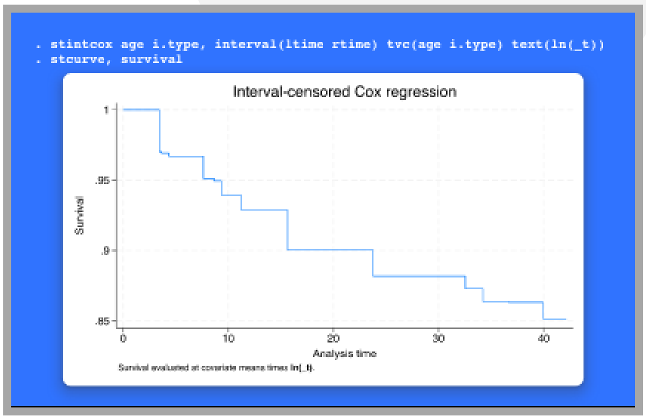
stintcox also has new options **tvc(***varlist_t***)** and **texp(***exp***)** that provide a convenient way to include time-interacted covariates, which are formed by the interaction between covariates specified in **tvc()** and a deterministic function of time specified in **texp()**.

After fitting a model, the standard and special-interest postestimation features are available and appropriately account for the time-varying covariates. Users can use the new **estat gofplot** command to produce a goodness-of-fit plot. One can predict the relative hazard. and use **stcurve** to plot survivor and related functions. Researchers working with multiple record data can use the new **stcurve** option

> **atmeans** to evaluate the function at time-specific means of the covariates or the new option
>
> **atframe(***framename***)** to evaluate the function at values of variables specified in the *framename* frame.

Note: Genuine semiparametric modelling of interval-censored event-time data was not available until the methodological advances made in recent years, which are implemented in the **stintcox** command.

Methodology continues to advance with the extensions for incorporating time-varying covariates, which are now available in this command.



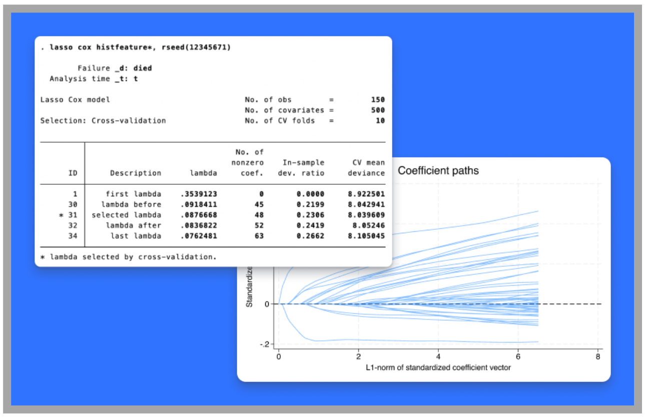
```
. stintcox age i.type, interval(ltime rtime) tvc(age i.type) text(ln(_t))
. stcurve, survival
```

Interval-censored Cox regression
Survival
1
.95
.9
.85
0
10
20
30
40
Analysis time
Survival evaluated at covariate means times ln(_t).

LASSO FOR COX MODEL

Lasso is used for prediction and model selection in the presence of many potential covariates. Where many means hundreds, thousands, or more!. The **lasso** command was previously introduced to perform lasso for linear, logit, probit, and Poisson models. New in *Stata* 18 is lasso for Cox proportional hazards models. **lasso cox** can be used to select covariates using lasso and fit a Cox model to survival-time data. **elasticnet cox** can similarly be used to select covariates using elastic net and fit a Cox model.

After **lasso cox** and **elasticnet cox**, researchers can use **predict** to predict the hazard ratio; use **stcurve** to plot the survivor, hazard, or cumulative hazard functions; or use any of the other postestimation tools available after **lasso** and **elasticnet** to examine the lasso results.

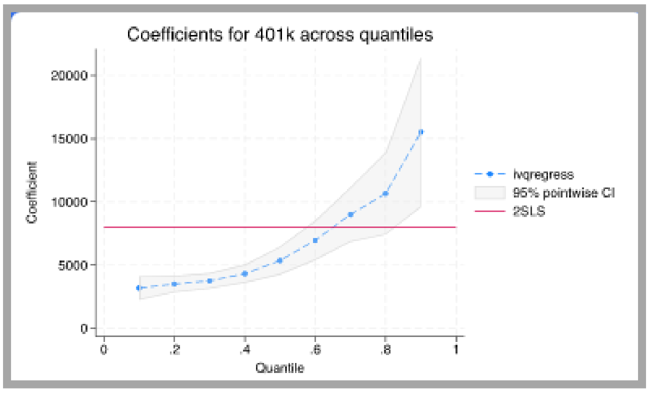


```
. lasso cox histfeature*, rseed(12345671)

        Failure _d: died
  Analysis time _t: t

Lasso Cox model                          No. of obs        =        150
                                         No. of covariates =        500
Selection: Cross-validation              No. of CV folds   =         10

--------------------------------------------------------------------------
         |                                No. of
         |                               nonzero    In-sample      CV mean
      ID |     Description      lambda     coef.    dev. ratio    deviance
---------+----------------------------------------------------------------
       1 |    first lambda    .3539123         0       0.0000     8.922501
      30 |   lambda before    .0918411        45       0.2199     8.042941
    * 31 | selected lambda    .0876668        48       0.2306     8.039609
      32 |    lambda after    .0836822        52       0.2419      8.05246
      34 |     last lambda    .0762481        63       0.2662     8.105045
--------------------------------------------------------------------------
* lambda selected by cross-validation.
```

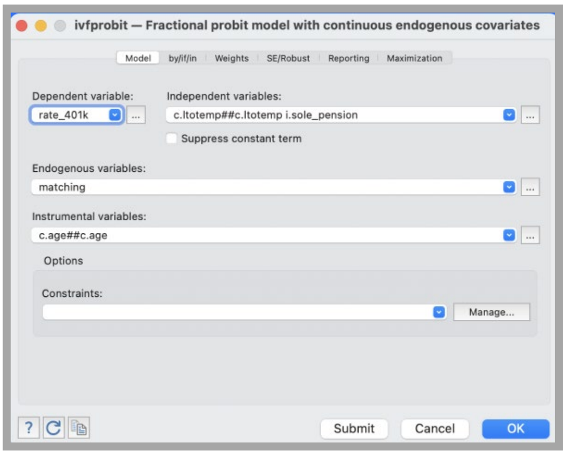
INSTRUMENTAL-VARIABLES QUANTILE REGRESSION

When researchers want to study the effects of covariates on different quantiles of the outcome, not just on the mean, they use quantile regression. For instance, they might be interested in modelling the grade distribution of students and how it is affected by changes in covariates. The existing **qreg** command fits quantile regression models, but what if they suspect that one of their covariates is endogenous? This endogeneity might arise for reasons such as self-selection of study participants, omission of a relevant variable from the model, or measurement error. The new **ivqregress** command allows them to model quantiles of the outcome and, at the same time, control for problems that arise from endogeneity using IV.

After fitting an IV quantile regression model, one can plot the coefficients across quantiles with the **estat coefplot** command. Researchers can test for endogeneity using the **estat endogeffects** command and estimate dual confidence intervals that are robust to weak instruments using **estat dualci**.



IV FRACTIONAL PROBIT MODEL


ivfprobit — Fractional probit model with continuous endogenous covariates

Dependent variable: rate_401k

Independent variables: c.ltotemp##c.ltotemp i.sole_pension

□ Suppress constant term

Endogenous variables: matching

Instrumental variables: c.age##c.age

Options

Constraints:

Fractional outcomes are common. One might be modelling participation rates in a 401(k) pension plan, the pass rate on standardized tests, expenditure shares, or the like.

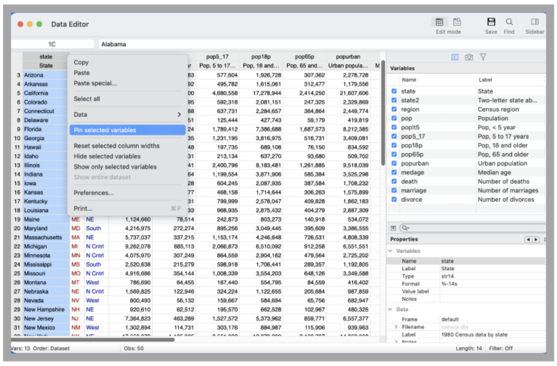
Fractional response models are a flexible and intuitive way to model outcomes that lie between 0 and 1.

They do not have the problem of linear models that will yield predictions outside 0 and 1 or the problem of log-odds models that are undefined at 0 and 1. Fractional response models can be fit using the existing **fracreg** command.

What if researchers are concerned that one or more of their model covariates are endogenous? With the new **ivfprobit** command, they can fit a model for a fractional dependent variable and account for endogeneity in one or more of the covariates.

DATA EDITOR ENHANCEMENTS

The Data Editor has many enhancements in *Stata* 18:



1C | Alabama

	state					pop5_17	pop18p	pop65p	popurban
	State				r	Pop, 5 to 17...	Pop, 18 and...	Pop, 65 and...	Urban popula...
3	Arizona				83	577,604	1,926,728	307,362	2,278,728
4	Arkansas				92	495,782	1,615,061	312,477	1,179,556
5	California				00	4,680,558	17,278,944	2,414,250	21,607,606
6	Colorado				95	592,318	2,081,151	247,325	2,329,869
7	Connecticut				88	637,731	2,284,657	364,864	2,449,774
8	Delaware				51	125,444	427,743	59,179	419,819
9	Florida				24	1,789,412	7,386,688	1,687,573	8,212,385
10	Georgia				35	1,231,195	3,816,975	516,731	3,409,081
11	Hawaii				48	197,735	689,108	76,150	834,592
12	Idaho				31	213,134	637,270	93,680	509,702
13	Illinois				41	2,400,796	8,183,481	1,261,885	9,518,039
14	Indiana				64	1,199,554	3,871,906	585,384	3,525,298
15	Iowa				28	604,245	2,087,935	387,584	1,708,232
16	Kansas				77	468,158	1,714,644	306,263	1,575,899
17	Kentucky				31	799,999	2,578,047	409,828	1,862,183
18	Louisiana				33	968,935	2,875,432	404,279	2,887,309
19	Maine	ME	NE	1,124,660	78,514	242,873	803,273	140,918	534,072
20	Maryland	MD	South	4,216,975	272,274	895,256	3,049,445	395,609	3,386,555
21	Massachusetts	MA	NE	5,737,037	337,215	1,153,174	4,246,648	726,531	4,808,339
22	Michigan	MI	N Cntrl	9,262,078	685,113	2,066,873	6,510,092	912,258	6,551,551
23	Minnesota	MN	N Cntrl	4,075,970	307,249	864,559	2,904,162	479,564	2,725,202
24	Mississippi	MS	South	2,520,638	215,279	598,918	1,706,441	289,357	1,192,805
25	Missouri	MO	N Cntrl	4,916,686	354,144	1,008,339	3,554,203	648,126	3,349,588
26	Montana	MT	West	786,690	64,455	167,440	554,795	84,559	416,402
27	Nebraska	NE	N Cntrl	1,569,825	122,946	324,224	1,122,655	205,684	987,859
28	Nevada	NV	West	800,493	56,132	159,667	584,694	65,756	682,947
29	New Hampshire	NH	NE	920,610	62,512	195,570	662,528	102,967	480,325
30	New Jersey	NJ	NE	7,364,823	463,289	1,527,572	5,373,962	859,771	6,557,377
31	New Mexico	NM	West	1,302,894	114,731	303,176	884,987	115,906	939,963

Vars: 13 Order: Dataset | Obs: 50

Variables

	Name	Label
☑	state	State
☑	state2	Two-letter state ab...
☑	region	Census region
☑	pop	Population
☑	poplt5	Pop, < 5 year
☑	pop5_17	Pop, 5 to 17 years
☑	pop18p	Pop, 18 and older
☑	pop65p	Pop, 65 and older
☑	popurban	Urban population
☑	medage	Median age
☑	death	Number of deaths
☑	marriage	Number of marriages
☑	divorce	Number of divorces

Properties

Variables	
Name	state
Label	State
Type	str14
Format	%-14s
Value label	
Notes	
Data	
Frame	default
Filename	census.dta
Label	1980 Census data by state
Notes	

Length: 14 | Filter: Off

Pinnable rows and columns. Pinned rows or columns do not scroll with the rest of the data so they will stay in view as you scroll through the data. This is useful for doing visual comparisons with some other data that might be visible only as you scroll. An ID variable would be a natural thing to pin.

Resizable cell editor for string data. When one edits string variables, the cell editor can be resized so that more of the string is visible while editing without it scrolling out of view of the cell editor.

Tooltips for truncated text. Any cell value that is too wide to fit within its cell's column width is truncated to fit. Hovering the mouse pointer over a cell with truncated text will display a tooltip with the cell's value without truncation.

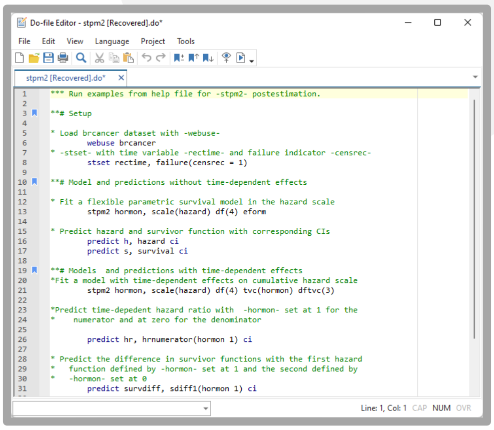
Proportional-width font support. The Data Editor now supports proportional-width fonts. This improves readability of the data and allows more variables to be displayed at a time without requiring scrolling. A monospaced font can still be used if preferred.

Show variable labels in column header. Variable labels can now be displayed in the column headers right below their variable names. This can be very useful for viewing datasets with short and nondescriptive variable names that have variable labels.

New keyboard shortcut for hiding or showing value labels. Quickly toggle between viewing numeric values and their corresponding labels.

DO-FILE EDITOR ENHANCEMENTS

The Do-file Editor also has enhancements in *Stata* 18:



```
*** Run examples from help file for -stpm2- postestimation.

**# Setup

* Load brcancer dataset with -webuse-
        webuse brcancer
* -stset- with time variable -rectime- and failure indicator -censrec-
        stset rectime, failure(censrec = 1)

**# Model and predictions without time-dependent effects

* Fit a flexible parametric survival model in the hazard scale
        stpm2 hormon, scale(hazard) df(4) eform

* Predict hazard and survivor function with corresponding CIs
        predict h, hazard ci
        predict s, survival ci

**# Models  and predictions with time-dependent effects
*Fit a model with time-dependent effects on cumulative hazard scale
        stpm2 hormon, scale(hazard) df(4) tvc(hormon) dftvc(3)

*Predict time-depedent hazard ratio with  -hormon- set at 1 for the
*    numerator and at zero for the denominator

        predict hr, hrnumerator(hormon 1) ci

* Predict the difference in survivor functions with the first hazard
*   function defined by -hormon- set at 1 and the second defined by
*   -hormon- set at 0
        predict survdiff, sdiff1(hormon 1) ci
```

Automatic backups. Documents that are open in the Do-file Editor are periodically saved to a backup file on disk. This includes new documents that have not been saved to disk yet. If one's computer were to lose power or crash before changes to the document were saved, the unsaved changes can still be recovered. To recover the unsaved changes, open the document in the Do-file Editor again. If a backup file is found in the same location as the document, users will be prompted to recover the backup file or open the document that was last saved to disk. Recovering the backup file will simply load it into the Do-file Editor; it will not overwrite the document that was saved to disk unless you choose to do so.

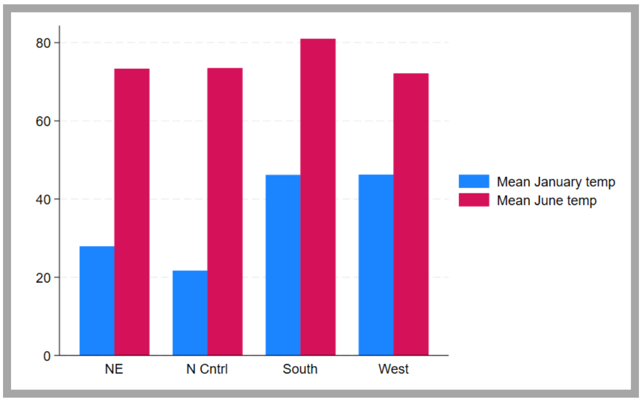
Syntax-highlight user-defined keywords. *Stata*'s Do-file Editor now includes the ability to syntax-highlight user-defined keywords. This will allow users to syntax-highlight favorite community-contributed commands. Users simply create a specially named keyword definition file containing a list of keywords, and *Stata* will syntax-highlight those keywords using a settable color and font styles such as bold or italics. One can even create a global keyword definition file that can be shared with all users of the same computer. Each user can still create their own local keyword definition file, and the keywords from both the global file and the local file will be loaded into the Do-file Editor.

ALL NEW GRAPH STYLE

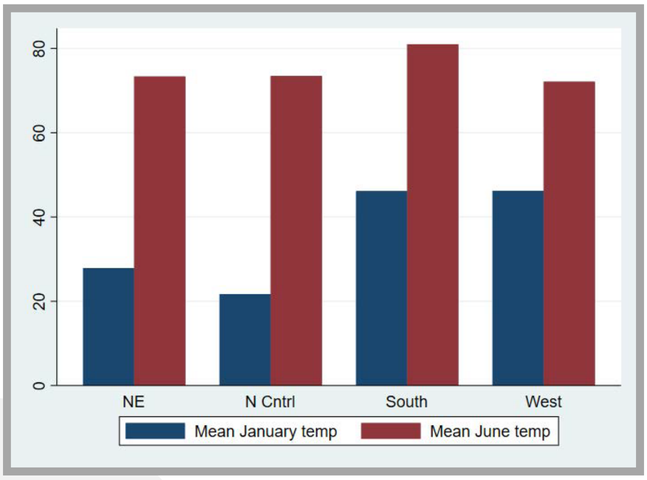
Graphs in *Stata* 18 have a new look.

- The new default graph scheme (or, the new appearance for *Stata* graphs) includes the following much-requested features:
- White background
- Updated color palette with bright colors
- Horizontal *y*-axis labels
- Wider aspect ratio
- Dynamic legend placement for certain graphs
- And more

As one example, a bar graph in with the new scheme now looks like this:

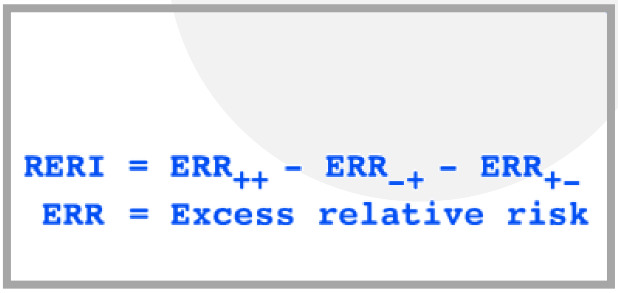


Instead of



In fact, four new graph schemes have been introduced: **stcolor**, **stcolor_alt**, **stgcolor**, and **stgcolor_alt**. The new default is **stcolor** and the other schemes are variations on **stcolor** that provide different widths and legend placements.

RELATIVE EXCESS RISK DUE TO INTERACTION (RERI)



$$\text{RERI} = \text{ERR}_{++} - \text{ERR}_{-+} - \text{ERR}_{+-}$$

$$\text{ERR} = \text{Excess relative risk}$$

Epidemiologists often need to determine how two exposures interact to put subjects at a higher risk of experiencing an outcome of interest. For example, one might want to investigate how exposures to cigarette smoke and asbestos interact to increase the risk of lung cancer. With the new **reri** command, users can measure two-way interactions in an additive model of relative risk, while accounting for other risk factors.

Researchers can choose from various supported models, such as logistic, binomial generalized linear, Poisson, negative binomial, Cox, parametric survival, interval-censored parametric survival, and interval-censored Cox models. They can evaluate an additive model for the interaction of smoke and asbestos by using three related statistics: RERI, attributable proportion, and synergy index.

TSTAT S.r.l
via Rettangolo, 12/14 | 67039 Sulmona (AQ)
Tel +39 0864 210101
tstat@tstat.it
www.tstat.it | www.tstattraining.eu

***Stata* Official Trainer in Europe**

***Stata* Official *Stata* Distributor for:**
Albania • Bosnia and Herzegovina • Croatia • Greece • Italy • Kosovo • North Macedonia • Malta • Montenegro • Serbia • Slovenia •Slovak Republic