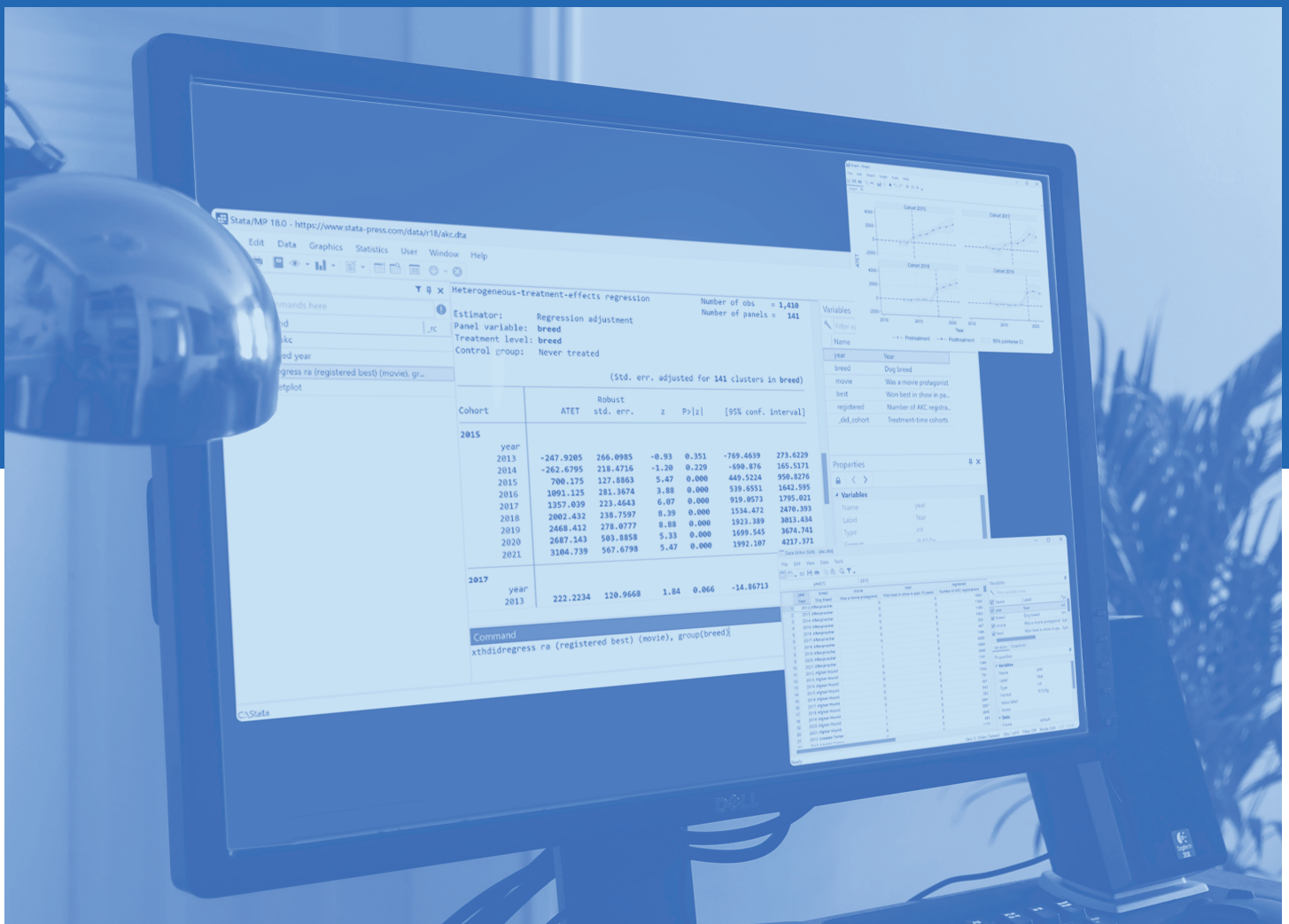# STATA®

Tools for all your data science needs

Fast. Accurate. Easy to use. Stata is your complete system for managing, graphing, and analyzing data. It is easy to learn through the extensive graphical interface yet completely programmable for the most demanding data manipulation and statistical requirements.

ANOVA/MANOVA
balanced and unbalanced designs • factorial, nested, and mixed designs • repeated measures • marginal means • contrasts • more

Basic statistics
summaries • cross-tabulations • correlations • z and t tests • equality-of-variance tests • tests of proportions • confidence intervals • factor variables • more

Bayesian analysis
thousands of built-in models • univariate and multivariate models • linear and nonlinear models • panel data • multilevel models • VAR • DSGE • continuous, binary, ordinal, and count outcomes • **bayes** prefix for 58 estimation commands • continuous univariate, multivariate, and discrete priors • add your own models • multiple chains • convergence diagnostics • posterior summaries • hypothesis testing • model fit • model comparison • predictions • dynamic forecast • impulse–response functions • more

Bayesian model averaging
full enumeration • MC3 and MH sampling • three model prior classes • fixed and random g-priors for coefficients • heredity rules • PIP for predictors • model ranking by PMP • BMA convergence • variable-inclusion maps • model-size distribution plots • jointness measures • log predictive-score • predictions • more

Binary, count, and categorical outcomes
logistic, probit, tobit • Poisson and negative binomial • conditional, multinomial, nested, ordered, rank-ordered, and stereotype logistic • multinomial probit • zero-inflated and left-truncated count models • selection models • marginal effects • more

Causal inference/treatment effects
inverse probability weight (IPW) • doubly robust methods • propensity-score matching • regression adjustment • covariate matching • DID • multivalued treatments • endogenous treatments • average treatment effects (ATEs) • ATEs on the treated (ATETs) • potential-outcome means (POMs) • continuous, binary, count, fractional, and survival outcomes • panel data • lasso • causal mediation analysis • more

Choice models
discrete choice • rank-ordered alternatives • conditional logit • multinomial probit • nested logit • mixed logit • panel data • case-specific and alternative-specific predictors • interpret results—expected probabilities, covariate effects, comparisons across alternatives • more

Cluster analysis
hierarchical clustering • kmeans and kmedians nonhierarchical clustering • dendrograms • stopping rules • user-extensible analyses • more

Contrasts, pairwise comparisons, and margins
compare means, intercepts, or slopes • compare with reference category, adjacent category, grand mean, etc. • orthogonal polynomials • multiple-comparison adjustments • graph estimated means and contrasts • interaction plots • more

Data manipulation
data transformations • data frames • match-merge • import/export data • JDBC • ODBC • SQL • Unicode • by-group processing • append files • sort • row–column transposition • labeling • save results • more

DSGE models
specify models algebraically • solve models • estimate parameters • identification diagnostics • policy and transition matrices • IRFs • dynamic forecasts • Bayesian • more

Epidemiology
standardization of rates • case–control • cohort • matched case–control • Mantel–Haenszel • pharmacokinetics • ROC analysis • ICD-10 • additive models of risk • more

Exact statistics
exact logistic and Poisson regression • exact case–control statistics • binomial tests • Fisher's exact test for r × c tables • more

Extended regression models (ERMs)
endogenous covariates • sample selection • nonrandom treatment • panel data • account for problems alone or in combination • continuous, interval-censored, binary, and ordinal outcomes • more

Finite mixture models (FMMs)
fmm prefix for 17 estimators • mixtures of a single estimator • mixtures combining multiple estimators or distributions • continuous, binary, count, ordinal, categorical, censored, truncated, and survival outcomes • more

Functions
statistical • random number • mathematical • string • date and time • regular expressions • Unicode • more

Generalized linear models (GLMs)
10 link functions • user-defined links • seven distributions • ML and IRLS estimation • nine variance estimators • seven residuals • more

Graphics
lines • bars • areas • ranges • contours • confidence intervals • interaction plots • survival plots • publication quality • customize anything • Graph Editor • more

Installation qualification
IQ report for regulatory agencies such as the FDA • installation verification

IRT (Item response theory)
binary (1PL, 2PL, 3PL), ordinal, and categorical response models • item characteristic curves • test characteristic curves • item information functions • test information functions • multiple-group models • differential item functioning (DIF) • more

Lasso
lasso • elastic net • model selection • prediction • inference • continuous, binary, count, and survival outcomes • cross-validation • adaptive lasso • double selection • partialing out • crossfit partialing out • double machine learning • endogenous covariates • treatment effects • more

Latent class analysis (LCA)
binary, ordinal, continuous, count, categorical, fractional, and survival items • add covariates to model class membership • combine with SEM path models • expected class proportions • goodness of fit • predictions of class membership • more

Linear models
regression • censored outcomes • endogenous regressors • bootstrap, jackknife, and robust and cluster–robust variance • wild cluster bootstrap • instrumental variables • three-stage least squares • constraints • quantile regression • GLS • DID • HDFE • more

Mata—Stata's serious programming language
interactive sessions • large-scale development projects • optimization • matrix inversions • decompositions • eigenvalues and eigenvectors • LAPACK engine • Intel® MKL • real and complex numbers • string matrices • interface to Stata datasets and matrices • numerical derivatives • object-oriented programming • more
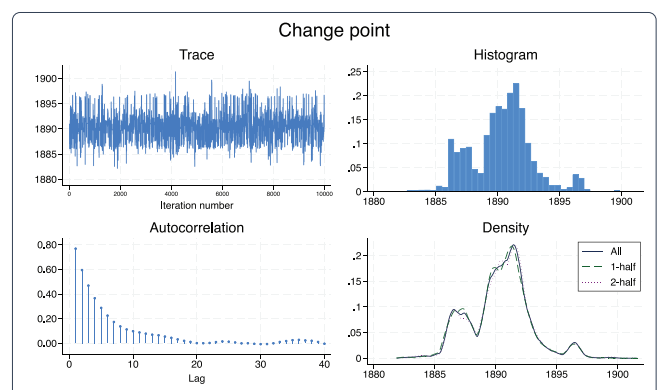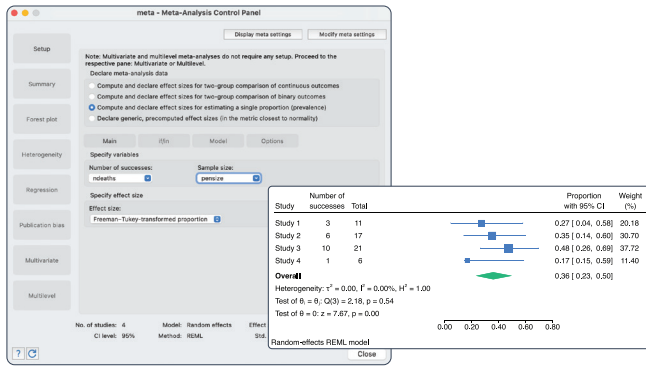
Meta-analysis

effect sizes • common, fixed, and random effects • forest, funnel, and more plots • subgroup, leave-one-out, and cumulative analysis • meta-regression • small-study effects • publication bias • multivariate • multilevel • more

Multilevel mixed-effects models

continuous, binary, count, and survival outcomes • two-, three-, and higher-level models • generalized linear models • nonlinear models • random intercepts • random slopes • crossed random effects • BLUPs of effects and fitted values • hierarchical models • residual error structures • DDF adjustments • support for survey data • more

Multiple imputation

nine univariate imputation methods • multivariate normal imputation • chained equations • explore pattern of missingness • manage imputed datasets • fit model and pool results • transform parameters • joint tests of parameter estimates • predictions • more

Multivariate methods

factor analysis • principal components • discriminant analysis • rotation • multidimensional scaling • Procrustean analysis • correspondence analysis • biplots • dendrograms • user-extensible analyses • more

Nonlinear regression, GMM and other systems of equations

generalized method of moments (GMM) • nonlinear regression • demand systems • more

Nonparametric methods

nonparametric regression • Wilcoxon–Mann–Whitney, Wilcoxon signed ranks, and Kruskal–Wallis tests • Cochran–Armitage and other trend tests • Spearman and Kendall correlations • Kolmogorov–Smirnov tests • exact binomial CIs • survival data • ROC analysis • smoothing • bootstrapping • more

Nonparametric regression

series estimation • kernel regression • multiple covariates • discrete and continuous covariates • point estimates and CIs • automatic optimal spline or bandwidth selection • slices of response surface with CIs and associated plots • more

Panel/longitudinal data

random and fixed effects with robust standard errors • HDFE • linear mixed models • random-effects probit • GEE • random- and fixed-effects Poisson • dynamic panel-data models • instrumental variables • endogenous treatment • sample selection • panel unit-root tests • DID • more

Power, precision, and sample size

power • sample size • effect size • minimum detectable effect • CI width • means • proportions • variances • correlations • ANOVA • regression • cluster randomized designs • case–control studies • cohort studies • contingency tables • survival analysis • balanced or unbalanced designs • results in tables or graphs • group sequential designs for clinical trials • more

Programming features

adding new commands • scripting • object-oriented programming • menu and dialog-box programming • dynamic documents • Markdown • Project Manager • Python integration • PyStata • Jupyter notebook • Java integration • Java plugins • H2O access • C/C++ plugins • more

Reporting

reproducible reports • customizable tables • graphical tables builder • Word • Excel • PDF • HTML • dynamic documents • Markdown • Stata results and graphs • SVG • EPS • PNG • TIF • more

Resampling and simulation methods

bootstrap • jackknife • Monte Carlo simulation • permutation tests • exact p-values • more

SEM (Structural equation modeling)

graphical path diagram builder • standardized and unstandardized estimates • modification indices • direct and indirect effects • continuous, binary, count, ordinal, and survival outcomes • multilevel models • random slopes and intercepts • factor scores, empirical Bayes, and other predictions • groups and tests of invariance • goodness of fit • handles MAR data by FIML • correlated data • survey data • more

Spatial autoregressive models

spatial lags of dependent variable, independent variables, and autoregressive errors • fixed and random effects in panel data • endogenous covariates • analyze spillover effects • more

Survey methods

multistage designs • bootstrap, BRR, jackknife, linearized, and SDR variance estimation • poststratification • calibration • raking • DEFF • predictive margins • means, proportions, ratios, totals • summary tables • almost all estimators supported • more

Survival analysis

Kaplan–Meier and Nelson–Aalen estimators • Cox regression (frailty) • parametric models (frailty, random effects) • competing risks • hazards • time-varying covariates • left-, right-, and interval-censoring • Weibull, exponential, and Gompertz models • more

Tests, predictions, and effects

Wald tests • LR tests • linear and nonlinear combinations • predictions and generalized predictions • marginal means • least-squares means • adjusted means • marginal and partial effects • forecast models • Hausman tests • more

Time series

ARIMA • ARFIMA • ARCH/GARCH • VAR • SVAR • ISVAR • VEC • multivariate GARCH • unobserved-components model • dynamic factors • state-space models • Markov-switching models • business calendars • tests for structural breaks • threshold regression • forecasts • impulse–response functions • local projections • unit-root tests • filters and smoothers • rolling and recursive estimation • more

Community-contributed features

search and download thousands of free additions • discover new features in the *Stata Journal* • share commands by posting to the SSC • discuss community-contributed commands on Statalist • more

Other statistical methods

kappa measure of interrater agreement • Cronbach's alpha • stepwise regression • tests of normality • more
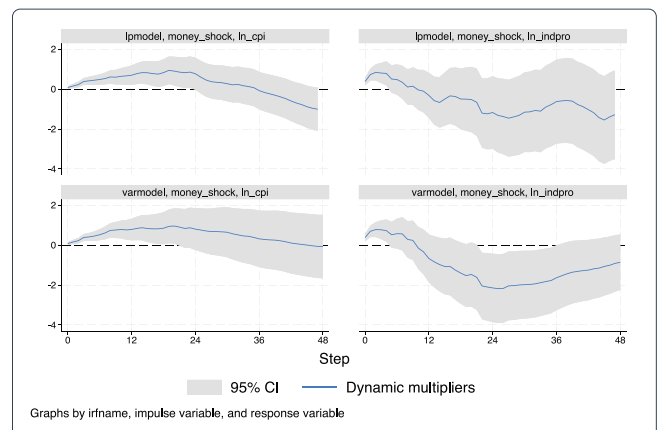
Statistical breadth and depth

You want a statistical package that does everything you need and also addresses the expanding range of statistical methods. Research professionals from all disciplines use Stata to manage and analyze their data—and so can you. Whether you are a student or a seasoned research professional, Stata gives you all the tools you need.

Easy to learn ...

With Stata's menus and dialogs, you get the best of both worlds. You can easily point and click or drag and drop your way to all of Stata's statistical, graphical, and data manipulation features. Use Stata's intuitive command syntax to quickly execute commands. You can even access the dialog boxes for each command directly from the online help system, which is a great way to explore all of Stata's capabilities.

Continuously updated

With Stata's continuous-release version, StataNow, you have access to new features as soon as they are ready. From the latest statistical methods to the best in reporting, data visualization, and the user interface, you do not have to wait until the next major release to begin using these new features.

Fully programmable ...

Stata has complete command-line scripting and programming facilities, including a full matrix programming language. You can easily script your analysis or even create new Stata commands that work just like those shipped with Stata. Stata's Do-file Editor allows you to create files of Stata commands, so analyses can be reproduced with the click of a button. With Stata's version control, your scripts will continue to work seamlessly when you upgrade to a new version.

Whether you enter commands directly or use the menus and dialogs, you can create a log of all actions and their results to ensure the reproducibility and integrity of your analyses.

And extensible

Easily search, add, and use any of the thousands of community-contributed commands with just a few mouse clicks.

Made for data science

stata.com/order