

Capability Statement

Modern statistical analysis — desktop & web, AI-assisted, reproducible.

Generated from the engine handler inventory · 634 procedures

Statios unifies the analytical scope of SPSS, Stata and EViews in one modern application, with a built-in AI assistant and one-click, citation-ready reporting. Every figure is computed by an open, auditable Python engine that runs both on the desktop and entirely in the browser.

634

statistical procedures

11

analysis domains

600

fully supported

34

with documented limits

What we do

DOMAIN	PROCEDURES	SCOPE
Descriptive Statistics & Hypothesis Tests	37	Descriptive statistics and statistical hypothesis tests covering parametric (t-tests, ANOVA), non-parametric (Mann-Whitney, Kruskal-Wallis, Friedman), correlation, and categorical association tests.
Regression & Generalized Linear Models	142	Statios's model namespace covers 50+ regression and generalized linear estimation procedures, from OLS to latent-variable extensions. Methods range from standard parametric models (linear, logistic, count) to specialized estimators for endogenous/selected samples, censoring, and complex errors.
Post-Estimation & Diagnostics	47	Statios provides 31 post-estimation and regression diagnostic procedures spanning heteroskedasticity tests, serial correlation diagnostics, influence measures, marginal effects, contrasts, and model comparison tools.
Time Series & Macroeconometrics	84	Comprehensive time series analysis, DSGE models, and advanced forecasting. Includes unit root tests, ARIMA/seasonal models, VAR/cointegration, impulse responses, nonlinear models, forecast evaluation, and mixed-frequency regression.
Panel Data & Causal Inference	51	Statios supports canonical panel data estimators (FE, RE, Hausman test, 2SLS-IV), advanced dynamic panel methods (Arellano-Bond, System GMM), and modern causal inference designs (DiD, RDD, synthetic control, staggered DiD, IPW, mediation analysis).
Multivariate, Factor & Latent-Variable Analysis	35	Covers exploratory and confirmatory dimensionality reduction, clustering, reliability, latent structure discovery, and item response modeling — from classical PCA/factor analysis through modern IRT, SEM, and categorical methods.
Survival, Bayesian, Meta-Analysis & Resampling	68	Statistical methods for survival analysis, Bayesian inference, meta-analytic pooling, and resampling-based inference. Includes Kaplan-Meier curves, Cox proportional-hazards regression, parametric and competing-risks models, Bayesian linear/logistic/hierarchical regression with MCMC diagnostics, fixed- and random-effects meta-analysis, and bootstrap/jackknife/permutation methods.
Machine Learning & Choice Models	31	Statios provides machine learning and choice modeling tools: supervised learning with explainability (Random Forest, Gradient Boosting, SVM, Decision Trees, neural networks); model evaluation (cross-validation, ROC/AUC, confusion matrix, feature importance, SHAP); and conjoint analysis (traditional, choice-based, hierarchical Bayes) for preference modeling.
Spatial, Complex Survey & Missing Data	50	Statios provides comprehensive spatial econometrics, complex survey design analysis, and missing data handling through libpysal/spreg/esda backends. Covers spatial weights, autocorrelation tests (Moran's I, Geary's C, Getis-Ord), LISA clustering, spatial regression (SAR, SEM, SDM, SAC, GWR, MGWR), and advanced imputation (MICE, EM, PMM) with design-based inference.
Data Management, Transforms & Workflow	68	Data management, transforms and workflow procedures in Statios: dataset loading and multi-dataset workflows, direct cell editing with undo/redo, spreadsheet-style formulas for derived variables, variable metadata configuration (types, labels, measurement levels), data transformations (compute, recode, categorize, standardize, filter, lag/lead, dummy), data manipulation (merge, reshape, aggregate, collapse, append, sort, subset, deduplicate, rank, transpose, split/concat columns,

DOMAIN	PROCEDURES	SCOPE
		pivot), data-to-table formatting, and session/macro recording for reproducible workflows.
AI, Automation, Graphics & Reporting	21	AI-powered statistical guidance, automated analysis pipelines, publication-quality graphics and report generation, plus a visual rule-based workflow recommender (Guided Path) and a node-based pipeline editor (RLab).

Available, with known limitations

These work today, with a caveat worth knowing.

CAPABILITY	AREA	LIMITATION
Regression Discontinuity (Sharp RDD)	Regression & Generalized Linear Models	Bandwidth selection is ad-hoc (1.5*SD rule), not data-driven MSE-optimal.
Extended Regression Model — Probit (ERM-Probit)	Regression & Generalized Linear Models	IMR lambda SE not corrected for generated-regressor problem (Murphy-Topel correction not implemented).
Extended Regression Model — Ordered (ERM-Ordered)	Regression & Generalized Linear Models	Lambda SE uncorrected; no p-values reported due to generated-regressor issue.
Treatment effect contrasts (ATE/ATT/ATC)	Post-Estimation & Diagnostics	Uses OLS-based potential outcomes, not designed for instrumental variables.
Diffusion-index direct forecast (user-supplied factors)	Time Series & Macroeconometrics	NOT a true Stock-Watson Dynamic Factor Model: there is no latent-factor extraction (no PCA / Kalman / state-space estimation of common factors inside the method) — the factors must be supplied as columns. Automatic latent-factor extraction is a separate upcoming feature.
Unrestricted MIDAS regression	Time Series & Macroeconometrics	Currently uses same dataset for both frequencies (pending multi-dataset resolution).
Almon polynomial MIDAS	Time Series & Macroeconometrics	Uses single dataset for both frequencies (pending multi-dataset support).
Beta polynomial MIDAS	Time Series & Macroeconometrics	Single-dataset limitation; multi-dataset pending (Phase 18).
Exponential Almon MIDAS	Time Series & Macroeconometrics	Single dataset; multi-dataset architecture pending.
MIDAS nowcasting	Time Series & Macroeconometrics	Single-dataset implementation; true multi-frequency data awaits Phase 18.
MIDAS model comparison	Time Series & Macroeconometrics	Single-dataset; intended for separate high/low datasets pending multi-dataset manager.
Real Business Cycle model	Time Series & Macroeconometrics	Log-linearized first-order approximation only; no occasionally binding constraints.
New Keynesian 3-equation model	Time Series & Macroeconometrics	Simplified linearized form; no ELB, expectations-driven dynamics approximate.
DSGE calibration	Time Series & Macroeconometrics	Moment-matching heuristic only; not Bayesian estimation or SMM.
Regression Discontinuity (Sharp RDD)	Panel Data & Causal Inference	Bandwidth selection is ad-hoc (not Imbens-Kalyanaraman or Calonico-Cattaneo-Titiunik data-driven optimal); local-linear only.
Propensity Score Matching (PSM)	Panel Data & Causal Inference	Uses naive paired SE; does not apply Abadie-Imbens (2006) variance correction for estimated scores and matching with replacement.

CAPABILITY	AREA	LIMITATION
Hierarchical Agglomerative Clustering	Multivariate, Factor & Latent-Variable Analysis	O(n ²) memory constraint; refuses n > 20,000 rows.
IRT Fit Indices (Approximate)	Multivariate, Factor & Latent-Variable Analysis	Approximate method: fixed-1PL reference and diagonal weight matrix, not full Maydeu-Olivares M2.
Shared Frailty Cox Model	Survival, Bayesian, Meta-Analysis & Resampling	Gamma frailty only; variance estimated via martingale residuals.
Bayesian Hierarchical Model	Survival, Bayesian, Meta-Analysis & Resampling	Fallback uses OLS-shrinkage approximation without full MCMC; ICC computed from variance components.
Bayes Factor	Survival, Bayesian, Meta-Analysis & Resampling	BIC-based approximation; not exact marginal likelihood.
SHAP Values	Machine Learning & Choice Models	TreeExplainer unavailable in browser (WASM); fallback to KernelExplainer (sampled Shapley values, not exact TreeSHAP).
CHAID	Machine Learning & Choice Models	Stub implementation; returns placeholder.
QUEST	Machine Learning & Choice Models	Stub implementation; returns placeholder.
C5.0	Machine Learning & Choice Models	Stub implementation; returns placeholder.
Ensemble Trees Comparison	Machine Learning & Choice Models	Stub implementation; returns placeholder.
Tree Export	Machine Learning & Choice Models	Stub implementation; returns placeholder.
Compare Trees	Machine Learning & Choice Models	Stub implementation; returns placeholder.
Hierarchical Bayes CBC	Machine Learning & Choice Models	Empirical-Bayes / EM-style scheme, NOT Gibbs/Metropolis; samples are posterior modes, not draws; heterogeneity SD understates posterior uncertainty.
Multiscale GWR (MGWR)	Spatial, Complex Survey & Missing Data	Requires standardization; mgwr package optional.
SKATER clustering	Spatial, Complex Survey & Missing Data	spopt optional; manual implementation approximate.
Max-p regions	Spatial, Complex Survey & Missing Data	Requires spopt package and GeoDataFrame with geometry.
Session Undo	Data Management, Transforms & Workflow	Session-level undo not fully wired; edit.undo handles cell-level reversals.

CAPABILITY	AREA	LIMITATION
Session Redo	Data Management, Transforms & Workflow	Session-level redo not fully wired; edit.redo handles cell-level reversals.
Large-dataset/columnar mode: bounded analysis suite	data	WASM in-browser path capped at 1.5M rows or 900MB. CSV files up to 50M rows route to columnar store (Parquet/NPY), but only subset of analyses support column-on-demand (streaming). Binary formats (.xlsx/.sav/.dta) cannot be streamed.
WASM/Pyodide compatibility: numba and SHAP slower without JIT	ml	Web build stubs numba (Pyodide lacks LLVM JIT) and SHAP's <code>_kernel_lib</code> . Code runs interpreted (correct numerically, ~2–5× slower than compiled). KernelExplainer works; TreeExplainer unavailable in WASM.
Charts: 4 types shipped, 31 roadmap items incomplete	graphs	Implemented: histogram, scatter, boxplot, bar. Missing: violin, density, Q–Q, interaction, forest, ROC, funnel, survival curves, residual diagnostic panels, and 21 others from Phase 15 roadmap.
Extended regression diagnostics: some estimators in pre-release rigor improvements	regression	Arellano-Bond, GMM, DSGE solvers present but some advanced econometric estimators have rigor enhancements on unmerged branches. Shipped versions work; see code docstrings for limitations (e.g., 'simplified solver', 'bounded lag range').

In active development

ITEM	AREA	NOTES
Statios web pivot: Python engine to Pyodide/WASM	infrastructure	W1–W4 phases (June 2026): migrate sidecar to in-browser Pyodide via WorkerTransport. Desktop Tauri shell frozen. PWA deployment, OPFS file handling, web deploy pipeline. Completed handler count remains 445; no new features, transport refactor only.
Charts Expansion: 31 missing chart types	graphs	Violin, density (KDE), ridgeline, Q–Q, P–P, interaction plot, mosaic, ACF/PACF, IRF with CI, residual diagnostics (4-panel), forest/funnel, survival, ROC, Bland-Altman, coefficient, marginal effects, control chart, caterpillar. Phase 15 completion.
AI Orchestrator & Multi-Model Routing	ai	Post-W4: intelligent provider selection, prompt caching, fallback logic, cost tracking. Statios relay authentication for managed web tier.
Quality backlog: UI refinements, performance, test coverage	polish	Post-W4: Monaco editor for scripts, keyboard shortcuts, undo/redo on data transforms, canvas DataGrid for >10k rows, performance audits on large-dataset analyses.

On the roadmap

ITEM	AREA	NOTES
Advanced causal inference: bounds & sensitivity analyses	causal	Lee bounds, Rosenbaum sensitivity, E-values. Handlers present (causal.lee_bounds, etc.) but backlog for UI polish and help content.
Exotic time-series models: TAR/STAR, NARDL, nowcasting	timeseries	Handlers exist (ts.tar, ts.nardl, ts.nowcasting) but limited test coverage and optional dependencies (ruptures for Bai–Perron). Deferred to post-quality phase.
Multivariate: MANOVA, discriminant, canonical correlation	analysis	Handlers registered (test.manova, analysis.discriminant, analysis.canonical_correlation) but frontend UI tabs not yet exposed in AnalysisModal. Feature-complete backend; UI routing needed.
Bayesian posterior predictive checks & Bayes factors	bayes	Handlers: bayes.ppc, bayes.credible_interval, bayes.prior_posterior, bayes.bayes_factor. Deps (PyMC, ArviZ) optional. Lower priority; posterior work focuses on MCMC diagnostics.
Optimal scaling & ERM: categorical regression (CATREG), optimal transforms	analysis	Handlers: analysis.catreg, analysis.catpca, analysis.overals, analysis.catcorr. Phase 23 backends exist but semi-automated UI guidance and diagnostic plots deferred.

Appendix — full procedure list

Descriptive Statistics & Hypothesis Tests 37

Descriptive statistics and statistical hypothesis tests covering parametric (t-tests, ANOVA), non-parametric (Mann-Whitney, Kruskal-Wallis, Friedman), correlation, and categorical association tests.

PROCEDURE	WHAT IT DOES	STATUS
Descriptive statistics	Mean, median, variance, skewness, kurtosis, quantiles for numeric variables.	Full
Frequency distribution	Frequency tables with counts, percentages, cumulative percentages for categorical/numeric variables.	Full
One-sample t-test	Test if sample mean differs from hypothesized value; reports t-statistic, 95% CI, Cohen's d.	Full
Independent samples t-test	Compare means of two independent groups; includes Levene's test, optional Welch correction.	Full
Paired t-test	Compare means of two related/paired samples; reports difference mean, SE, 95% CI, effect size.	Full
One-way ANOVA	Compare means across 3+ groups; includes Levene's test, Tukey HSD post-hoc, eta/omega-squared.	Full
Welch's ANOVA	Heteroscedastic one-way ANOVA without equal-variance assumption; Games-Howell post-hoc.	Full
Two-way ANOVA	Test main effects and interaction of two factors; includes partial eta-squared effect sizes.	Full
Repeated measures ANOVA	Within-subjects ANOVA for 2+ repeated measures; Mauchly sphericity, GG/HF corrections.	Full
Yuen's trimmed t-test	Robust two-group t-test using 20% trimmed means; insensitive to heavy tails and outliers.	Full
Mixed ANOVA	Split-plot design: between-subject factor and within-subject repeated measures.	Full
Dunnett's test	Post-hoc test comparing each treatment group to a control group.	Full
Cochran-Armitage test	Trend test for proportions across ordered groups.	Full
Mann-Whitney U test	Non-parametric two-group test; exact/asymptotic p-values, rank-biserial effect size.	Full
Wilcoxon signed-rank test	Non-parametric paired test; exact/asymptotic p-values, effect size r.	Full
Kruskal-Wallis H test	Non-parametric 3+ group test; Dunn post-hoc with Bonferroni correction.	Full
Friedman test	Non-parametric repeated measures test; Kendall's W effect size.	Full
Chi-square test of independence	Association between two categorical variables; includes Cramér's V, Yates correction disclosure.	Full
Chi-square goodness-of-fit	Test if observed distribution matches expected frequencies.	Full
Fisher's exact test	Exact test for 2x2 contingency table; alternative: one/two-sided.	Full
Pearson correlation matrix	Pairwise Pearson/Spearman/Kendall correlations with p-values; listwise NA handling.	Full

PROCEDURE	WHAT IT DOES	STATUS
Correlation (pairwise)	Single correlation with confidence interval (Fisher z-transform for Pearson).	Full
Partial correlation	Correlation between two variables controlling for others; corrected df.	Full
Effect size calculator	Cohen's d/Hedges' g, eta/omega-squared, Cramér's V, Fisher's f-squared.	Full
Power analysis	Sample size / power calculation for t-test, ANOVA, correlation, chi-square, regression.	Full
Sign test	Non-parametric paired/one-sample test; binomial test on signs of differences.	Full
Runs test	Wald-Wolfowitz test for randomness; dichotomize by median/mean or threshold.	Full
McNemar test	Change detection in paired binary outcomes; 2x2 table symmetry.	Full
Cochran Q test	Non-parametric repeated measures for binary outcomes (3+ measurements).	Full
Jonckheere-Terpstra test	Non-parametric alternative to ANOVA for ordered groups (monotonic trend).	Full
Kendall's W	Coefficient of concordance for 2+ judges/raters across multiple items.	Full
Kendall's tau	Rank correlation between two variables (tau-b for ties).	Full
Mood's median test	Non-parametric k-group test based on median; robust to outliers.	Full
Permutation test	Permutation-resampling test (10000 permutations); mean difference or custom statistic.	Full
Kernel density estimate	Non-parametric density estimation; Scott/Silverman bandwidth, Gaussian kernel.	Full
MANOVA	Multivariate analysis of variance; Wilks' lambda, Pillai, Hotelling-Lawley, Roy's root.	Full
Distribution fitting + GOF	Fit Exponential-Family and related distributions (gamma, exponential, lognormal, Weibull, normal) by MLE, rank by AIC, and test fit with Kolmogorov-Smirnov. Matches R MASS::fitdistr + ks.test.	Full

Regression & Generalized Linear Models 142

Statios's model namespace covers 50+ regression and generalized linear estimation procedures, from OLS to latent-variable extensions. Methods range from standard parametric models (linear, logistic, count) to specialized estimators for endogenous/selected samples, censoring, and complex errors.

PROCEDURE	WHAT IT DOES	STATUS
Linear Regression (OLS)	Standard ordinary least squares with diagnostics (R^2 , F-test, VIF, Durbin-Watson, residual normality).	Full
Linear Regression — Robust SEs (HCo–HC3)	OLS with heteroskedasticity-consistent standard errors; selectable HCo (White), HC1 (Stata default), HC2, or HC3 (MacKinnon & White 1985).	Full
Linear Regression — Newey-West HAC SEs	OLS with Newey-West heteroskedasticity- and autocorrelation-consistent (HAC) standard errors (Bartlett kernel, set hac_lags); standard for time-series regressions.	Full
Errors-in-Variables Regression	Method-of-moments correction for classical measurement error (attenuation bias) using per-regressor reliability ratios; reports naive OLS vs corrected coefficients.	Full
SIMEX (measurement-error correction)	Corrects attenuation bias from a regressor measured with known additive error variance by simulation-extrapolation: adds increasing extra noise, refits OLS at each lambda (averaging over B Monte-Carlo replicates), and extrapolates the coefficient quadratically back to lambda=-1 to recover the error-free estimate.	Full
Frisch-Waugh-Lovell Partial Slope	Partial out the controls from both Y and the focal regressor and regresses the residuals through the origin; numerically equals the focal coefficient in the full OLS, with the partial correlation.	Full
Omitted-Variable Bias Decomposition	Decomposes omitted-variable bias for each included regressor into short-model vs long-model coefficients and the exact $\gamma \cdot \delta$ contribution of each omitted variable.	Full
Log-Y Retransformation (Duan Smearing)	Fits a log-scale OLS and reports Duan's smearing factor $S = \text{mean}(\exp(\text{residuals}))$ plus naive vs bias-corrected (smeared) retransformed predictions on the original scale.	Full
Potential Outcomes (ATE/ATT/ATU)	Reports the average treatment effect (ATE), the effect on the treated (ATT) and on the untreated (ATU) in Rubin's counterfactual notation by regression adjustment under unconfoundedness: an OLS fit of the outcome on treatment, covariates and their interactions predicts each unit's pair ($Y(1)$, $Y(0)$), with delta-method or bootstrap standard errors.	Full
Selection Model (MNAR missingness)	Heckman (1979) two-step selection model for non-ignorable (MNAR) missingness: a probit models whether the outcome is observed, the outcome equation is corrected for selection on unobservables via the inverse Mills ratio, and rho (with its rho=0 ignorability test) quantifies the selection correlation, reported alongside naive complete-case OLS for comparison.	Full
Pattern-Mixture Model (MNAR)	Pattern-mixture model (Little 1993) for non-ignorable (MNAR) missing outcomes: partitions the sample by the outcome's missingness, fits the completers' regression, imputes the missing outcomes under a sensitivity restriction whose delta shifts the missing-group mean ($\delta = 0 = \text{MAR}$), and reports the pattern-averaged mean, fraction missing, and a sensitivity curve.	Full
Continuous-Treatment Dose-Response (GPS)	Estimates the average dose-response function of a continuous treatment with the Hirano-Imbens generalized propensity score: models the treatment given covariates as Normal, regresses the outcome on a flexible function of the treatment and the GPS, and averages to trace $E[Y(t)]$ over a grid of doses, returning the dose-response curve, the GPS-adjusted average marginal effect, and the naive unadjusted slope for comparison.	Full

PROCEDURE	WHAT IT DOES	STATUS
Abadie Kappa-Weighting (compliers)	Uses Abadie's (2003) kappa weights $\kappa = 1 - D(1-Z)/(1-p(X)) - (1-D)Z/p(X)$ to estimate the complier share, the complier mean outcome, the mean covariates among compliers versus the full sample, and a kappa-weighted LATE in a binary-instrument / binary-treatment IV setting.	Full
Marginal Treatment Effect (MTE)	Estimates the Heckman-Vytlacil marginal treatment effect via local IV / generalized Roy: fits the propensity score and a flexible control function in p , differentiates to trace the MTE over the unobserved-resistance dimension, integrates it with the appropriate weights to recover ATE/ATT/ATUT/LATE, and tests for essential heterogeneity.	Full
Synthetic DiD	Synthetic Difference-in-Differences (Arkhangelsky et al. 2021) combines synthetic-control unit weights with difference-in-differences time weights to estimate the ATT on a balanced treated/control panel, reporting the SDID effect, a placebo/jackknife standard error, the sparse non-negative unit weights and the pre-treatment fit.	Full
Changes-in-Changes (Athey-Imbens)	Nonlinear generalization of difference-in-differences (Athey & Imbens 2006) that maps the treated-pre outcome distribution through the control group's pre-to-post distributional change to identify the entire counterfactual treated distribution, reporting the ATT plus quantile treatment effects (0.25/0.5/0.75) that reveal heterogeneous effects ordinary DiD would miss.	Full
Doubly-Robust DiD (Sant'Anna-Zhao)	Estimates the ATT in a 2x2 difference-in-differences design by combining an outcome-regression DiD with an inverse-propensity-weighted DiD into the doubly-robust estimator of Sant'Anna & Zhao (2020), which stays consistent if either the propensity-score model or the untreated-outcome model is correct, reports influence-function standard errors with the IPW and OR components and the simple DiD for contrast, and reduces exactly to the canonical mean 2x2 DiD when no covariates are supplied.	Full
Honest DiD (parallel-trends sensitivity)	Rambachan & Roth (2023) sensitivity analysis for difference-in-differences: estimates the event-study coefficients, then bounds the post-treatment effect under a relative-magnitudes restriction (M_{bar}) on how far post-treatment trend violations can exceed observed pre-trends, reporting robust confidence sets across M_{bar} and the breakdown M_{bar} at which significance is lost.	Full
Sun-Abraham Event Study	Interaction-weighted (IW) event-study estimator that saturates the model with cohort-by-relative-period interactions and aggregates them by cohort shares, recovering the true dynamic treatment path (and clean pre-trends) under staggered adoption with heterogeneous effects where conventional dynamic TWFE is contaminated.	Full
Callaway-Sant'Anna Staggered DiD	Callaway & Sant'Anna (2021) staggered difference-in-differences estimating group-time average treatment effects $ATT(g,t)$ via clean 2x2 DiD against not-yet-treated/never-treated comparison units, aggregated to an overall ATT, an event-study (dynamic) profile, and per-cohort effects, with a naive two-way FE benchmark to expose heterogeneous-timing bias.	Full
Causal Forest (heterogeneous effects)	Honest generalized random forest (Wager-Athey) that estimates the conditional average treatment effect $\tau(x)$ via split-sample honest trees with R-learner local centering, returning the ATE with a forest-based SE, the distribution of individual CATEs, and the leading covariate driving effect heterogeneity.	Full
Double / Debiased ML	Estimates a debiased treatment effect in the partially-linear model $Y = \theta D + g(X) + e$ by cross-fitting gradient-boosted nuisance regressions for $E[Y X]$ and $E[D X]$ and	Full

PROCEDURE	WHAT IT DOES	STATUS
	solving the orthogonal (Neyman) moment on the residuals, reporting theta with robust SE/CI alongside the biased naive single-regression estimate for contrast.	
McCrary Density Test (RD manipulation)	McCrary (2008) density-discontinuity test that fits separate triangular-kernel local-linear density estimates on each side of an RD cutoff and tests whether the log-density of the running variable jumps at the threshold, detecting sorting/manipulation that would invalidate the RD design.	Full
Propensity-Score Subclassification	Estimates the ATE and ATT of a binary treatment by fitting a propensity-score logit, splitting units into quantile strata of the score, and combining the within-stratum treated-minus-control mean differences weighted by stratum size (ATE) or treated share (ATT), with per-stratum effects and a stratified standard error.	Full
Mahalanobis Matching (ATT)	Nearest-neighbour matching estimator of the Average Treatment effect on the Treated that matches each treated unit to its closest control(s) in Mahalanobis covariate distance (pooled-control inverse covariance), imputes the counterfactual from the matched controls, and reports the ATT with an approximate Abadie-Imbens standard error and covariate balance.	Full
Box-Cox Transformed Regression	Estimates the Box-Cox power-transform lambda maximizing the concentrated profile log-likelihood, then fits OLS on the transformed (strictly-positive) response.	Full
Stochastic Frontier Analysis	Maximum-likelihood Aigner-Lovell-Schmidt stochastic frontier (half-normal inefficiency) for production or cost functions, returning frontier coefficients, σ_v/σ_u , lambda, gamma, mean technical efficiency $E[\exp(-u)]$ and per-observation efficiency scores.	Full
Finite-Mixture Regression (latent class)	Fits a finite mixture of K linear regression regimes by EM, recovering each latent class's own coefficient vector, error variance, and mixing proportion together with the log-likelihood, BIC, and posterior class-assignment summary.	Full
Dirichlet-Process Mixture Density	Bayesian nonparametric density estimation that fits an infinite (Dirichlet-process) Gaussian mixture via truncated stick-breaking, automatically inferring the effective number of mixture components and returning their means, variances, weights, and the estimated density curve.	Full
Finite-Mixture Count (latent class)	Fits a finite mixture of K latent-class Poisson regressions by EM, recovering each regime's own log-link coefficient vector and mixing proportion together with the log-likelihood, BIC/AIC, and posterior class-assignment summary to capture unobserved heterogeneity and overdispersion in count data.	Full
Simultaneous Quantile Regression	Fits linear quantile regressions at several quantiles at once and uses a joint paired (xy) bootstrap covariance to report per-quantile coefficient tables and to test whether slopes change across the conditional distribution (e.g. tau=0.25 vs 0.75).	Full
Quantile Regression Forest	Estimates full conditional quantiles (Meinshausen 2006) with a random forest by reading the weighted empirical response distribution in each leaf, revealing heteroskedastic conditional intervals that widen with predictors plus an out-of-bag interval-coverage check and variable importance.	Full
Additive Quantile Regression (quantile GAM)	Fits a tau-quantile regression in which each predictor enters through a smooth cubic B-spline term, minimizing the asymmetric check-loss via linear programming to recover nonlinear conditional quantile curves.	Full

PROCEDURE	WHAT IT DOES	STATUS
Composite Quantile Regression	Estimates a single shared slope vector by pooling the check-loss across an equally-spaced grid of quantiles (with quantile-specific intercepts), giving slope estimates that match OLS under Gaussian errors and beat it in efficiency under heavy-tailed errors, with bootstrap standard errors.	Full
Expectile Regression	Newey-Powell asymmetric-least-squares expectile regression, fit by iteratively reweighted least squares to estimate any conditional tau-expectile (tau=0.5 reproduces OLS exactly), with heteroskedasticity-consistent sandwich standard errors and an asymmetry interpretation.	Full
Quantile Count Regression (jittering)	Estimates conditional quantiles of an overdispersed count response by the Machado-Santos Silva (2005) jittering method -- uniform noise is added to the integer counts, quantile regression is run on the log-transformed scale, and coefficients are averaged over many jitter draws to give stable, monotone-in-tau multiplicative quantile effects.	Full
Generated-Regressor Inference (Murphy-Topel)	Two-step estimation that regresses a stage-2 outcome on the fitted values from a first-stage OLS, reporting both the naive stage-2 standard errors and the Murphy-Topel/Pagan-corrected standard errors that propagate the first-stage estimation uncertainty into the generated regressor.	Full
Heteroskedastic Probit	Harvey (1976) binary probit in which the latent-error standard deviation is modelled as $\sigma_i = \exp(z_i'\gamma)$, estimated by BFGS maximum likelihood, returning separate mean (beta) and variance (gamma) coefficient tables plus a likelihood-ratio test of homoskedasticity ($H_0: \gamma=0$).	Full
Kernel Regression (Nadaraya-Watson)	Nonparametric Nadaraya-Watson local-constant kernel smoother estimating the conditional mean of one outcome on one predictor (Gaussian or Epanechnikov kernel, Silverman's rule-of-thumb or user bandwidth), returning fitted values, a 100-point smooth curve, an R-squared-like fit measure, and the bandwidth used.	Full
Kernel Conditional Density	Estimates the full conditional density $f(y x)$ with a double Gaussian kernel (Nadaraya-Watson weighting in x combined with a kernel in y), returning conditional 0.1/0.5/0.9 quantile curves and the conditional mean as functions of x so heteroskedastic spread, skew and multimodality are visible where a mean regression cannot show them.	Full
Single-Index Model (Ichimura)	Semiparametric single-index regression $y = g(x'\beta) + e$ estimated by Ichimura's semiparametric least squares, recovering the index direction beta (identified up to scale, unit-norm normalized) by minimizing the leave-one-out kernel-regression residual sum of squares while leaving the link $g(\cdot)$ fully nonparametric, and reporting standardized and raw-scale index coefficients, the data-driven bandwidth, a leave-one-out pseudo-R-squared, and the estimated link curve.	Full
Maximum Score (Manski)	Manski's semiparametric maximum-score estimator for binary choice, recovering the scale-normalized coefficient direction by maximizing correctly-predicted signs of $x'\beta$ under only a conditional-median error restriction, staying consistent where logit/probit fail under heteroskedasticity or non-normal errors.	Full
Translog Production/Cost Function	Estimates a second-order flexible (transcendental logarithmic) production or cost function by OLS on log-inputs, their squares and pairwise cross-products, reporting output elasticities at the sample mean, returns-to-scale, and an F-test of the Cobb-Douglas (no-second-order-terms) restriction.	Full

PROCEDURE	WHAT IT DOES	STATUS
Empirical Likelihood (mean)	Owen's nonparametric empirical-likelihood inference for a population mean or mean vector, returning the -2 log EL-ratio statistic with its chi-square p-value plus a distribution-free EL confidence interval that requires no normality assumption.	Full
Copula Fitting (bivariate dependence)	Fits Gaussian, Clayton, Gumbel and Frank copulas to two variables' rank pseudo-observations by maximum likelihood, reporting each family's parameter, log-likelihood and AIC, the AIC-selected copula, and Kendall's tau and Spearman's rho.	Full
Gaussian-Process Regression	Nonparametric Bayesian regression with an RBF + white-noise kernel that learns its length-scale and noise level by maximizing the log-marginal-likelihood, returning a smooth posterior mean plus per-point predictive uncertainty.	Full
BART (Bayesian Additive Regression Trees)	Flexible nonparametric regression that models the conditional mean as a regularized sum of trees fit by Bayesian backfitting MCMC, returning posterior-mean fitted values, predictive uncertainty, variable-inclusion importance, and in-sample R-squared.	Full
Gaussian-Process Classification	Fits a Gaussian-process classifier (RBF kernel, Laplace approximation) to a binary 0/1 outcome, reporting in-sample accuracy/log-loss/AUC, learned kernel hyperparameters, and a predicted-probability summary to capture smooth nonlinear class boundaries a linear logit would miss.	Full
k-Nearest-Neighbours Regression	Nonparametric regression that predicts each outcome as the average of its k nearest neighbours in standardized (Euclidean) predictor space, reporting an in-sample fit, a leave-one-out cross-validated RMSE and CV R-squared, and the chosen k.	Full
Spline Regression (B-splines)	Fits a flexible nonlinear curve by building a B-spline basis of one predictor (patsy bs() with given degrees of freedom and polynomial degree) and running OLS of the response on that basis, returning the fitted curve on a 100-point grid, the data-quantile knot locations, the basis dimension, R-squared and residual MSE.	Full
Generalized Additive Model (splines)	Fits $y = \text{intercept} + \sum_j f_j(x_j) + e$ with each term a penalized cubic B-spline (statsmodels GLMGam, Gaussian family), selecting the smoothing penalty by GCV and reporting per-term effective degrees of freedom, deviance, pseudo- R^2 , and a fitted-value summary.	Full
GAM (Poisson / Logit)	Generalized additive model for count (Poisson/log) or binary (Binomial/logit) outcomes that fits each predictor as a penalized cubic B-spline smooth via penalized IRLS with GCV/UBRE smoothing selection, reporting per-smooth effective degrees of freedom, deviance explained, fitted curves with confidence bands, and approximate significance.	Full
Local-Linear Regression	Nonparametric degree-1 local-polynomial kernel smoother that fits a kernel-weighted least-squares line at each point and takes its intercept as the conditional-mean estimate (Gaussian or Epanechnikov kernel, Silverman rule-of-thumb or user bandwidth), correcting the boundary bias of Nadaraya-Watson and returning fitted values, a 100-point smooth curve, effective degrees of freedom, and an R-squared-like fit measure.	Full
Local Polynomial Regression	Nonparametric kernel-weighted local polynomial smoother of arbitrary degree that generalizes Nadaraya-Watson (degree 0) and local-linear (degree 1) fits, returning the fitted curve, its estimated first derivative, and a smoother effective-df fit measure, with higher degrees reducing boundary bias.	Full
Isotonic Regression (PAVA)	Nonparametric monotone least-squares fit of a response on a single predictor via the pool-adjacent-violators algorithm, returning the fitted step function (level sets), R^2 ,	Full

PROCEDURE	WHAT IT DOES	STATUS
	and the imposed direction.	
Wavelet Regression (denoising)	Nonparametric denoising of a response treated as a signal ordered by a predictor via discrete wavelet transform with soft-thresholding (VisuShrink universal or hybrid SureShrink), returning the denoised fitted curve, robust MAD noise estimate, retained-coefficient sparsity, and reconstruction diagnostics.	Full
Wavelet Coherence	Computes squared wavelet coherence R^2 in $[0,1]$ between two time series across time and scale via a hand-rolled FFT-based Morlet continuous wavelet transform, smoothed cross-wavelet and power spectra (Torrence & Compo 1998; Grinsted et al. 2004), summarizing the dominant common period, coherence by short/medium/long scale bands, and the cone-of-influence-masked global coherence.	Full
Partially Linear (Robinson)	Robinson (1988) $\sqrt{n}$ -consistent semiparametric partially-linear regression $y = x'\beta + g(z) + e$, estimating the linear slopes β by the double-residual method (Nadaraya-Watson partialling-out of y and each x on z , then residual-on-residual OLS) while leaving $g(z)$ fully nonparametric.	Full
Overdispersion Test (Cameron-Trivedi)	Cameron-Trivedi (1990) regression-based test of the Poisson equidispersion assumption: fits a Poisson GLM, then runs auxiliary through-origin OLS of the studentised squared-residual statistic on $g(\mu)$ for $g(\mu)=\mu$ (NB1) and $g(\mu)=\mu^2$ (NB2), reporting per-variant $\hat{\alpha}$, a one-sided t-test of $H_0 \alpha=0$, and an overdispersion verdict.	Full
Conway-Maxwell-Poisson Regression	COM-Poisson GLM (log link on λ) with a jointly-estimated dispersion parameter ν that flexibly handles under- ($\nu>1$) and over-dispersed ($\nu<1$) counts, fit by maximum likelihood with a stable truncated normalizing constant and a Wald test of $\nu=1$ against the Poisson.	Full
Endogenous Count (control function)	Poisson count regression with an endogenous regressor corrected by two-stage residual inclusion (control function): a first-stage OLS residual on instruments is added to the count model, yielding the endogeneity-corrected effect, a control-function residual coefficient that doubles as an endogeneity test, and bootstrap standard errors.	Full
System OLS (multi-equation)	Estimates a system of linear equations equation-by-equation by OLS (intercept added), reporting per-equation coefficient tables (coef/se/t/p/CI), R^2 and n , plus the cross-equation residual covariance Σ (symmetric, PSD) and residual correlation on the common listwise-complete sample - the SUR/FGLS baseline.	Full
Joint Confidence Region	Fits OLS and reports the joint F test that a chosen set of coefficients is simultaneously zero, the exact confidence-ellipse geometry (centre, covariance submatrix, eigen-axes and half-axis lengths) for exactly two coefficients, and Bonferroni simultaneous intervals, contrasting the tilted exact region with the conservative axis-aligned box.	Full
General GLS (specified Omega)	Feasible generalized least squares (Aitken) under a user-specified error covariance: AR(1) Toeplitz errors, groupwise heteroskedasticity (one variance per group level), or power heteroskedasticity (var proportional to $ \text{fitted} ^\delta$), reporting the estimated structure parameter and an efficient coefficient table.	Full
Complementary Log-Log Regression	Binary regression with the asymmetric complementary log-log link $P(y=1)=1-\exp(-\exp(x'\beta))$ for rare-event data and grouped/discrete-time proportional-hazards (survival) models, returning a coef/se/z/p/CI table with log-likelihood and AIC/BIC.	Full

PROCEDURE	WHAT IT DOES	STATUS
GEE (Population-Averaged)	Generalized estimating equations fit a population-averaged (marginal) regression for clustered or longitudinal data, modelling within-cluster dependence through a working correlation (independence / exchangeable / AR-1) while reporting robust sandwich standard errors valid even under correlation misspecification.	Full
LIML (Limited-Information ML)	Limited-information maximum likelihood IV estimator that is less biased than 2SLS under weak or many instruments, reporting the k-class kappa eigenvalue, robust coefficient table with confidence intervals, first-stage instrument-strength diagnostics, and an overidentification test.	Full
Two-Sample IV (TS2SLS)	Two-sample two-stage least squares: estimates a single endogenous regressor's structural coefficient by fitting the first stage in one half of the data and carrying the fitted endogenous value into the reduced-form/structural equation in the other half (Inoue-Solon TS2SLS; Angrist-Krueger TSIV design), reporting the coefficient with a robust SE, first-stage strength, and sample sizes.	Full
Semiparametric IV (sieve)	Newey-Powell / Ai-Chen series (sieve) two-stage instrumental-variables estimation of an unknown dose-response $g(\text{endog})$ approximated by a polynomial sieve and instrumented by a sieve basis of the instruments, returning the estimated $g(\cdot)$ curve on a grid, the average derivative, HC1-robust standard errors, first-stage relevance, and a naive sieve-OLS comparison to expose endogeneity bias.	Full
Lewbel Heteroskedasticity IV	Estimates a regression with an endogenous regressor when no external instruments exist, by constructing Lewbel (2012) generated instruments from heteroskedasticity in the first-stage residual and running heteroskedasticity-robust 2SLS, reporting the first-stage F and an OLS-vs-Lewbel comparison.	Full
Generalized Empirical Likelihood (GEL)	Information-theoretic estimation of a linear IV / moment-condition model via the GEL saddle-point (empirical likelihood, exponential tilting, or continuous-updating), returning coefficients with sandwich SEs, an LR-type overidentification test, and nonnegative implied probabilities; reproduces 2SLS exactly when just-identified.	Full
Empirical-Likelihood Regression	Estimates linear-regression coefficients on the OLS estimating-equation moments $E[x(y - x'\beta)] = 0$ and tests/intervals them via Owen's empirical likelihood, giving distribution-free, generally-asymmetric EL confidence intervals plus chi-square-calibrated EL-ratio p-values whose point estimates coincide with OLS.	Full
Continuously-Updated GMM (CUE)	Continuously-updated GMM for a linear IV model that re-estimates the optimal weight matrix as a function of the parameters inside the objective and minimises jointly, returning CUE coefficients with robust standard errors, z/p-values, the Hansen J overidentification test, and (when just-identified) an exact 2SLS match.	Full
Optimal-Instrument GMM	Efficient GMM for a linear IV model that builds Chamberlain (1987) optimal instruments $A(z) = E[x z]/\text{Var}(e z)$ from a flexible projection of the regressors on the instruments plus a fitted conditional-variance model, reporting beta with efficient-GMM standard errors, first-stage fit, Hansen's J overidentification test, and an efficiency comparison versus plain 2SLS.	Full
Nonlinear GMM (exponential moments)	Two-step efficient GMM for the exponential / multiplicative-error model $y = \exp(x'\beta) \cdot \eta$ with instruments, using Mullahy's multiplicative moment $E[z^*(y \cdot \exp(-x'\beta) - 1)] = 0$ to identify possibly-endogenous regressors (a nonlinear analogue of 2SLS) and reporting the coefficient table plus Hansen's J overidentification test.	Full

PROCEDURE	WHAT IT DOES	STATUS
Method of Simulated Moments	Method of Simulated Moments / indirect inference for a univariate AR(1)-plus-measurement-noise model whose likelihood is intractable: parameters (ϕ , σ_η , σ_ϵ) are chosen so simulated auxiliary statistics (variance and first two autocovariances) match the data's, minimizing a scale-free weighted distance with common random numbers and reporting estimates, sandwich SEs with the $(1+1/S)$ simulation-noise inflation, the objective value, and the moment-match table.	Full
Minimum Distance (CMD)	Classical minimum-distance (minimum-chi-square) estimation that recovers a structural IV slope from the OLS reduced-form coefficients by minimizing $(\hat{\pi} - h(\hat{\theta}))' W (\hat{\pi} - h(\hat{\theta}))$ with the efficient weight $W = \text{inv}(\text{Var}(\hat{\pi}))$, reporting $\hat{\theta}$, CMD standard errors, the overidentification chi-square, and (in the just-identified case) exact agreement with 2SLS.	Full
IV Quantile Regression (Chernozhukov-Hansen)	Estimates the structural effect of an endogenous regressor on a chosen conditional quantile of the outcome using one or more instruments, via the Chernozhukov-Hansen inverse-quantile-regression grid search that drives the instrument coefficient to zero, correcting the endogeneity bias that contaminates ordinary quantile regression.	Full
Oaxaca-Blinder Decomposition	Decomposes the mean outcome gap between two groups from separate within-group OLS fits into the threefold endowments/coefficients/interaction components and the twofold pooled-reference (Neumark) explained vs unexplained split.	Full
Prais-Winsten AR(1) Regression	Iterative AR(1) FGLS that quasi-differences the data and retains a scaled first observation, correcting OLS for first-order serial correlation; reports ρ and before/after Durbin-Watson.	Full
Linear Probability Model	OLS of a binary 0/1 outcome with HC1 robust SEs; coefficients are marginal effects on $P(y=1)$, and the share of out-of-[0,1] fitted values is flagged.	Full
Generalized Linear Model (family x link)	Unified GLM: choose the exponential family (gaussian, poisson, binomial, gamma, inverse-Gaussian) and link (identity, log, logit, probit, inverse, sqrt). Reports coefficients, deviance, AIC/BIC.	Full
Linear Combination of Coefficients (lincom)	Tests an arbitrary linear combination $g'\beta$ of OLS coefficients with its standard error, t, p-value, and 95% CI (Stata-style lincom).	Full
Delta-Method (nonlinear combination)	Delta-method standard error for a sum / difference / ratio / product of two OLS coefficients, with t, p-value, and 95% CI.	Full
Restriction Test (F / Wald / LR)	General joint-exclusion test: F, Wald, and likelihood-ratio statistics for dropping a subset of regressors (unrestricted vs restricted OLS).	Full
Binary Logistic Regression	Binary outcome (0/1) modeled as probability; reports odds ratios, McFadden R^2 , classification accuracy.	Full
Probit Regression	Binary choice via normal CDF link; includes marginal effects and classification accuracy.	Full
Multivariate Probit	Jointly models 3+ correlated binary outcomes by maximum simulated likelihood with the GHK simulator, returning per-equation probit coefficients plus the full cross-equation error correlation matrix.	Full
Bivariate Probit	Jointly estimates two binary-outcome probit equations with correlated normal errors by maximizing the bivariate-normal log-likelihood, reporting both coefficient tables, the	Full

PROCEDURE	WHAT IT DOES	STATUS
	error correlation rho with a Wald rho=0 test that detects whether the outcomes can be modeled as two independent probits, the log-likelihood and n.	
Ordered Logit	Ordinal outcome; reports coefficients, threshold parameters, and McFadden pseudo-R ² .	Full
Generalized Ordered Logit	Fits an ordinal-outcome model that relaxes the proportional-odds (parallel-lines) assumption by letting each cumulative split have its own ML-estimated coefficient vector, reporting per-threshold coefficient tables, a Brant-style Wald test (with an LR cross-check) of parallel lines, log-likelihood, AIC/BIC, and an interpretation.	Full
Sequential / Continuation-Ratio Logit	Models an ordinal outcome as a sequence of binary continuation-ratio logits -- one per ordered transition, fit on the at-risk subset {Y>=r} -- yielding stage-specific coefficients for outcomes whose categories are stages literally passed through in order.	Full
Ordered Probit	Ordinal outcome via normal CDF; threshold parameters and pseudo-R ² included.	Full
Multinomial Logit	Multi-category outcome; reports coefficients and relative-risk ratios per category relative to base.	Full
Choice-Based Sampling (WESML)	Corrects a conditional/multinomial logit for endogenous (choice-based) stratified sampling using the Manski-Lerman WESML estimator, reweighting each observation by population-share / sample-share of its chosen alternative to recover consistent population coefficients with robust sandwich standard errors, alongside a side-by-side comparison to the biased unweighted fit.	Full
Multinomial Probit (GHK)	Estimates a random-utility multinomial probit on long-format choice data by maximum simulated likelihood with the Geweke-Hajivassiliou-Keane (GHK) recursive simulator, recovering attribute coefficients and the identified correlated-error covariance while relaxing the IIA restriction of the conditional logit.	Full
Nested Logit	Two-level nested logit for discrete choice estimated by FIML, relaxing IIA within nests via per-nest dissimilarity (lambda) parameters with a 0<lambda<=1 RUM-consistency check and a likelihood-ratio test against the conditional logit.	Full
Mixed (Random-Parameters) Logit	Random-parameters (mixed) logit estimated by simulated maximum likelihood on long-format discrete-choice data, drawing Normal mixing coefficients via Halton draws to recover preference means and the std devs that quantify unobserved heterogeneity.	Full
Maximum Simulated Likelihood	Fits a random-coefficient (mixed) binary-choice probit by Maximum Simulated Likelihood, approximating the intractable integral over unobserved slope heterogeneity with shifted-Halton low-discrepancy draws and reporting the mean coefficients, the heterogeneity standard deviation, the simulated log-likelihood, the number of draws, and a likelihood-ratio comparison to the pooled (no-heterogeneity) probit.	Full
Conditional Logit (McFadden)	Estimates McFadden's choice-specific conditional logit on long-format discrete-choice data, recovering how alternative-varying attributes drive the probability that each alternative is chosen within a choice occasion, with a full coefficient table, log-likelihood, and McFadden pseudo-R ² .	Full
Poisson Regression	Count data (non-negative integer); reports incidence-rate ratios (IRRs) and dispersion diagnostics; optional exposure/offset term for rate (frequency) models.	Full
Negative Binomial (NB1 / NB2)	Overdispersed count data; MLE-estimated dispersion (alpha). Selectable NB2 (quadratic variance, default) or NB1 (linear variance). Includes likelihood-ratio test of	Full

PROCEDURE	WHAT IT DOES	STATUS
	Poisson vs NB.	
Zero-Inflated Poisson (ZIP)	Count data with excess zeros via mixing; separate count and zero-inflation equations with Vuong test.	Full
Zero-Inflated Negative Binomial (ZINB)	Overdispersed count data with excess zeros; count + inflation + alpha parameter estimated jointly.	Full
Hurdle Model	Two-part: logit for zero vs nonzero + truncated Poisson for positive counts.	Full
Tobit Regression (Censored)	Outcome censored at known bound (lower/upper); MLE estimation with sigma parameter.	Full
Censored LAD (Powell CLAD)	Powell's (1984) Censored Least Absolute Deviations estimator for left-censored (Tobit-type) outcomes, fit by an iterated median-regression linear-programming scheme with bootstrap standard errors, consistent under conditional-median restrictions and robust to heteroskedastic or heavy-tailed errors where Tobit is biased.	Full
Truncated Regression	Sample truncated (observations above threshold retained); MLE with sigma.	Full
Heckman Selection Model (Two-Step)	Endogenous sample selection via probit + IMR augmentation; reports rho correlation coefficient.	Full
Endogenous Switching (Roy model)	Estimates a two-regime Roy switching regression with regime-specific outcome equations and a probit selection equation, correcting each regime for selection on unobservables via its own inverse Mills ratio and reporting the rho_0/rho_1 error-correlation terms.	Full
Beta Regression	Outcome in (0,1); precision parameter (phi) and mean-model coefficients estimated jointly.	Full
Fractional-response quasi-MLE (Papke-Wooldridge)	Outcome in [0,1]; Papke-Wooldridge (1996) Bernoulli quasi-MLE — GLM with binomial family + logit link; includes marginal effects.	Full
Quantile Regression	Conditional quantile (median or specified tau); pseudo-R ² and quantile-specific inference.	Full
Ridge Regression (L2)	L2 regularized OLS on standardized predictors; alpha controls shrinkage intensity.	Full
LASSO Regression (L1)	L1 regularized OLS; automatic feature selection via coefficient sparsity. Standardized fit.	Full
Elastic Net (L1+L2)	Combined L1/L2 penalty on standardized predictors; l1_ratio balances penalties.	Full
Elastic Net (cross-validated)	Elastic-net penalized linear regression that standardizes predictors and uses K-fold cross-validation to pick the penalty strength lambda at minimum CV-MSE and at the 1-SE rule, reporting the CV-MSE curve, original-scale selected coefficients, the number selected, and R ² .	Full
Robust Regression (Huber M-estimator)	M-estimator resistant to outliers; asymptotic-normal Wald inference (z-stat); no t-distribution df.	Full
Stepwise Selection	Greedy forward/backward/both; adds/removes by AIC or BIC until no improvement.	Full

PROCEDURE	WHAT IT DOES	STATUS
Weighted Least Squares (WLS)	User-supplied or estimated weights (FGLS); corrects heteroscedasticity.	Full
Feasible Generalized Least Squares (FGLS)	Cochrane-Orcutt iterative GLSAR; corrects serial correlation in errors.	Full
Nonlinear Least Squares (NLS)	User-provided formula and starting values; <code>scipy.optimize curve_fit</code> minimizes residual sum of squares.	Full
Seemingly Unrelated Regressions (SUR)	Multiple equations fitted jointly accounting for cross-equation correlation; Breusch-Pagan test included.	Full
Three-Stage Least Squares (3SLS)	System of equations with endogenous right-hand-side variables; IV + SUR combined.	Full
FIML (simultaneous system)	Full-Information Maximum Likelihood joint estimation of a two-equation linear simultaneous-equations system, maximizing the system Gaussian log-likelihood concentrated over the residual covariance (with the $\ln \det(B) $ Jacobian) to recover structural coefficients with FIML standard errors, the residual covariance/correlation, the system log-likelihood, and a side-by-side comparison against equation-by-equation 2SLS.	Full
Generalized Method of Moments (GMM)	IV-GMM with moment conditions; J-test of overidentifying restrictions.	Full
Panel Fixed Effects	Within-entity estimator; removes time-invariant heterogeneity. Clustered SEs on entity. Optional two-way (entity + time) effects.	Full
Panel Random Effects	GLS random-effects estimator; assumes random effect uncorrelated with predictors.	Full
Linear Mixed Model (Random Intercepts/Slopes)	REML by default; fixed-effect Wald-z inference. AIC/BIC reported only under ML.	Full
Intraclass Correlation (ICC)	ICC(1) from one-way random-intercept model; variance-partition coefficient for clustering.	Full
Difference-in-Differences (DiD)	Treatment×Post interaction; cluster-robust or HC1 SEs depending on clustering.	Full
Regression Discontinuity (Sharp RDD)	Local-linear at cutoff with triangular or uniform kernel. Ad-hoc or user-supplied bandwidth.	Limited
RD Robust (CCT bias-corrected)	Sharp regression-discontinuity treatment effect via local-polynomial (triangular-kernel) fits on each side of the cutoff within an MSE-optimal bandwidth, reporting the conventional point estimate, the Calonico-Cattaneo-Titiunik (2014) bias-corrected estimate, conventional and robust standard errors, confidence intervals, and effective sample sizes.	Full
Propensity Score Matching (PSM)	Logit/probit propensity score; 1:1/1:k matching with replacement or caliper.	Full

PROCEDURE	WHAT IT DOES	STATUS
Extended Regression Model — Linear (ERM-Linear)	Combines endogeneity (IV-2SLS), sample selection (Heckman), treatment effects in one model.	Full
Extended Regression Model — Probit (ERM-Probit)	Endogeneity (control function) + selection (bivariate-probit IMR) for binary outcome.	Limited
Extended Regression Model — Ordered (ERM-Ordered)	Ordered probit with sample selection correction via IMR augmentation.	Limited
Extended Regression Model — Interval (ERM-Interval)	Outcome observed as bounds [lower, upper]; handles left/right/interval censoring via MLE.	Full
ERM Diagnostics	Specification tests for endogeneity, selection bias, and treatment effect significance within ERM.	Full
LOESS Regression	Local polynomial regression; non-parametric smoothing via weighted least squares.	Full
Gamma GLM (severity)	Gamma generalized linear model with log link for strictly-positive responses (e.g. claim severity/cost); supports an exposure/offset term. Dispersion via Pearson, matching R <code>glm(family=Gamma(link="log"))</code> .	Full
Tweedie GLM (pure premium)	Tweedie compound Poisson-Gamma GLM (log link) for non-negative data with a mass at zero — single-model pure-premium pricing; configurable variance power p in $(1,2)$.	Full

Post-Estimation & Diagnostics 47

Statios provides 31 post-estimation and regression diagnostic procedures spanning heteroskedasticity tests, serial correlation diagnostics, influence measures, marginal effects, contrasts, and model comparison tools.

PROCEDURE	WHAT IT DOES	STATUS
Normality tests (comprehensive suite)	Shapiro-Wilk, Kolmogorov-Smirnov, D'Agostino-Pearson, Anderson-Darling; includes skewness/kurtosis and Q-Q plot data.	Full
Homogeneity of variance test	Levene and Bartlett tests for equality of variances across groups; concludes on Levene (robust).	Full
Outlier detection (multiple methods)	IQR, Z-score ($ z > 3$), and modified Z-score (MAD-based) methods with index identification.	Full
Multicollinearity check	Variance Inflation Factor (VIF) analysis with concern classification (low/moderate/high).	Full
Automatic test suggestion	Recommends appropriate statistical tests based on data characteristics.	Full
Breusch-Pagan test for heteroskedasticity	LM and F variants test if residual variance depends on predictors; requires $n > k+2$.	Full
White test for heteroskedasticity	General heteroskedasticity test not assuming specific form; more flexible than Breusch-Pagan.	Full
Goldfeld-Quandt heteroskedasticity test	Sorts by first predictor, splits at the midpoint, drops a block of observations starting at the midpoint, and F-tests the lower vs upper group residual variances.	Full
Breusch-Godfrey serial correlation test	LM and F tests for autocorrelation up to specified lags in OLS residuals.	Full
Durbin-Watson autocorrelation statistic	DW near 2 indicates no autocorrelation; < 1.5 positive, > 2.5 negative autocorrelation.	Full
Ljung-Box autocorrelation test	Tests joint significance of autocorrelations at specified lags; generic for series/residuals.	Full
Heteroskedasticity-robust LM test	Wooldridge robust LM test that a set of regressors are jointly zero, valid under heteroskedasticity of unknown form ($n - SSR$ auxiliary regression on residualized products).	Full
AR(1) serial-correlation test	Wooldridge's t-test for first-order serial correlation: regress OLS residuals on their own lag (plus the regressors) and test $\rho = 0$.	Full
Durbin's h Test	Durbin's h test for first-order serial correlation in OLS models that include a lagged dependent variable, where the Durbin-Watson statistic is biased toward 2.	Full
Moulton Factor (clustered SE inflation)	Moulton (1986) standard-error inflation factor: estimates the residual intra-class correlation across groups and reports naive OLS vs Moulton-adjusted standard errors.	Full
Vuong Test (non-nested)	Vuong's (1989) likelihood-ratio test for two non-nested models (OLS/logit/Poisson), comparing per-observation log-likelihoods to decide which specification is closer to the true data-generating process, with AIC/BIC complexity-corrected variants.	Full

PROCEDURE	WHAT IT DOES	STATUS
Cox Test (non-nested)	Pesaran-Cox test discriminating between two non-nested linear regression models via the difference in log-likelihoods adjusted by its expectation under each model, reported in both directions as standard-normal statistics with a preferred-model verdict.	Full
Encompassing Test (non-nested)	Mizon-Richard (1986) encompassing test that asks, by artificial nesting, whether one non-nested linear model explains everything its rival adds: each direction adds the rival's non-overlapping regressors to the base model and F-tests their joint significance (insignificant => the base model encompasses the rival), returning F/df/p, per-direction verdicts, and a preferred model.	Full
J Test (non-nested models)	Davidson-MacKinnon J test that discriminates between two non-nested linear specifications by adding each model's OLS fitted values to the other and testing their t-statistic in both directions, returning per-direction rejection verdicts and a preferred model.	Full
Information-Matrix Test (White)	White's (1982) omnibus information-matrix specification test for OLS: $IM = n \cdot R^2$ of the auxiliary regression of squared residuals on the unique squares and cross-products of the regressors (chi-square), detecting general misspecification — wrong functional form, heteroskedasticity, or non-normality — not just heteroskedasticity.	Full
Harvey Multiplicative Heteroskedasticity	Harvey's (1976) LM test for multiplicative heteroskedasticity, regressing the log of squared OLS residuals on chosen variables to estimate the variance model $Var(e) = \exp(z' \alpha)$ and test whether the variance depends on them.	Full
LM / Score Test (omitted regressors)	Classic Lagrange-multiplier (score) test for omitted regressors: regresses the residuals of the restricted OLS (dependent on kept) on the full regressor set and reports $LM = n \cdot R^2$ -squared of that auxiliary regression as a chi-square(q) test that the tested coefficients are jointly zero, with the equivalent F-form and a reject/not verdict.	Full
Multivariate Normality (Doornik-Hansen)	Tests whether a set of variables (or VAR residuals) is jointly multivariate normal via Mardia's multivariate skewness and kurtosis statistics and the Doornik-Hansen omnibus $\chi^2(2p)$ test, returning each statistic with its df and p-value, per-variable skew/kurtosis, and an overall normal/non-normal verdict.	Full
Ramsey RESET specification test	Tests for functional form misspecification by adding polynomial terms of fitted values.	Full
Hosmer-Lemeshow goodness-of-fit test	Groups logit predictions by deciles, compares observed vs expected event counts via chi-square.	Full
Variance Inflation Factor (VIF) analysis	VIF per predictor with concern thresholds (VIF > 5 moderate, > 10 high); includes mean VIF.	Full
Cook's distance influence measure	Identifies observations with disproportionate influence on regression; threshold default $4/n$.	Full
Leverage (hat values) analysis	Detects high-leverage observations (unusual X values); threshold $2 \cdot (k+1)/n$.	Full
DFBETAS parameter-specific influence	Measures change in each coefficient when one observation removed; threshold $2/\sqrt{n}$.	Full

PROCEDURE	WHAT IT DOES	STATUS
Marginal effects (logit/probit)	Average Marginal Effects (AME) or Marginal Effects at Means (MEM) for binary choice models.	Full
Predicted probabilities with CI	Logit/probit predictions on observed data or at specified covariate values with 95% CI.	Full
Post-hoc multiple comparisons	Tukey, Bonferroni, Holm step-down, Scheffe, Games-Howell with p-value adjustments.	Full
Effect sizes across designs	Cohen's d, Glass's delta, Hedges' g, eta-squared, partial eta-squared, omega-squared, Cohen's f.	Full
Planned contrasts (pairwise/polynomial)	Pairwise, polynomial (Legendre), Helmert, deviation contrasts with t-tests and CI.	Full
Prediction intervals (OLS)	Confidence and prediction intervals for fitted values; wider than CI due to residual variance.	Full
Coefficient table (with VIF/robust SE)	Formatted OLS results with standardized betas, VIF, HC3 robust SE option; significance stars.	Full
Model comparison (R-sq, AIC, BIC)	Side-by-side OLS models with delta AIC/BIC relative to best; RMSE and F-stat per model.	Full
Marginal effects at specified values	Conditional marginal effects (OLS/logit/probit/Poisson) at user-supplied covariate values.	Full
Margins contrast (difference in effects)	Compare marginal effects of one predictor across two or more levels; delta-method SE.	Full
Conditional marginal effects (interaction)	Marginal effect of focal variable at 10th, 25th, 50th, 75th, 90th percentiles of moderator.	Full
Nested contrasts (hierarchical)	Two-sample Welch t-tests on specified pairs within a grouping variable.	Full
Elasticities (full/partial/arc)	$dy/dx \cdot x/y$ (full), $dy/dx \cdot x$ (partial), or arc elasticity across specified predictor values.	Full
Semi-elasticities ($dy/dx/y$ and $dy/dx \cdot x$)	Proportional change in y and linear change in y per unit change in x.	Full
Predicted means over groups (Stata-style)	Marginal cell means with 95% CI; pairwise comparisons across factor levels.	Full
Adjusted predictions (holding covariates constant)	Linear predictor at specified covariate values (mean/median/min/max); 95% CI included.	Full
Probability profile (logit/probit curve)	Binary outcome probability as function of one predictor; other predictors at mean.	Full
Treatment effect contrasts (ATE/ATT/ATC)	Average treatment effect via regression-based g-computation; balanced or flexible interaction.	Limited

Time Series & Macroeconometrics 84

Comprehensive time series analysis, DSGE models, and advanced forecasting. Includes unit root tests, ARIMA/seasonal models, VAR/cointegration, impulse responses, nonlinear models, forecast evaluation, and mixed-frequency regression.

PROCEDURE	WHAT IT DOES	STATUS
Autocorrelation & Partial Autocorrelation (ACF / PACF)	Sample ACF and PACF (Durbin-Levinson / Yule-Walker) with white-noise confidence bands and per-lag Ljung-Box Q statistics, for ARMA order identification.	Full
HAC Long-Run Variance	Estimates the heteroskedasticity- and autocorrelation-consistent (HAC) long-run variance $S = \gamma_0 + 2 \sum k(j/L) \gamma_j$ of a time series using the Bartlett (Newey-West) or Quadratic-Spectral (Andrews) kernel, with the Newey-West automatic plug-in bandwidth, reporting the variance ratio S/γ_0 and the HAC standard error of the mean.	Full
Structural VECM (long-run restrictions)	Fits a Johansen VECM at a given cointegration rank and identifies structural shocks via long-run (Granger-representation) restrictions, reporting the reduced-rank long-run impact matrix, the lower-triangular structural long-run matrix (K-r permanent / r transitory shocks), and few-horizon structural impulse responses.	Full
Time-Varying-Parameter VAR	Estimates a VAR whose coefficients drift over time as random walks, cast as a linear-Gaussian state-space model and recovered by the Kalman filter and Rauch-Tung-Striebel smoother, returning smoothed coefficient paths (start/end/mean), state-innovation variances, an LR test of constant coefficients, and the log-likelihood.	Full
Markov-Switching VAR	Regime-switching vector autoregression (Hamilton) whose intercepts, dynamics, and innovation covariance switch between latent Markov regimes, estimated by EM (Hamilton filter + Kim smoother) returning per-regime VAR coefficients, the regime transition matrix, smoothed regime probabilities, and log-likelihood.	Full
Threshold VAR (TVAR)	Estimates a two-regime multivariate threshold VAR by grid-searching the threshold that splits the sample on a transition variable's lag, fits a separate VAR in each regime, and tests for threshold nonlinearity with a bootstrap-calibrated sup-LR statistic.	Full
Quantile VAR	Vector autoregression estimated at a chosen conditional quantile via equation-by-equation quantile regression of each variable on the full lagged vector, returning per-lag tau-quantile coefficient matrices alongside the mean (OLS) VAR for comparison, to capture tail/regime-dependent dynamics.	Full
Local Projections (Jorda IRF)	Estimates Jorda (2005) local-projection impulse responses by running a separate OLS regression of y_{t+h} on the shock at t (plus controls) at each horizon with Newey-West HAC standard errors, yielding an impulse-response function with confidence bands that is robust to dynamic misspecification relative to VAR IRFs.	Full
Unobserved Components (structural TS)	Harvey structural time-series model: decomposes a univariate series into stochastic level, optional local-linear-trend slope, optional stochastic seasonal, and irregular via Kalman-filter maximum likelihood, reporting component variances, smoothed series, log-likelihood/AIC, and a short forecast.	Full
Vector Moving-Average (VMA)	Fits a multivariate VMA(q) model $y_t = \mu + \epsilon_t + \Theta_1 \epsilon_{t-1} + \dots + \Theta_q \epsilon_{t-q}$ by conditional Gaussian maximum likelihood, returning the MA coefficient matrices, innovation covariance, finite (zero-beyond-lag-q) impulse responses, log-likelihood and information criteria.	Full
VARMA Model	Estimates a Vector ARMA(p,q) for a small multivariate time series by conditional maximum likelihood (via VARMAX), reporting the AR and MA coefficient	Full

PROCEDURE	WHAT IT DOES	STATUS
	matrices, residual covariance, log-likelihood and AIC/BIC.	
Intervention Analysis (interrupted TS)	Box-Tiao interrupted-time-series analysis fitting an AR(p) model with a deterministic step (permanent level shift) or pulse (transient shock) intervention dummy at a known date, reporting the estimated intervention effect with its t-test, the long-run steady-state level shift, pre/post means, and a plain-language interpretation.	Full
Subset / Restricted VAR	Estimates a vector autoregression then imposes data-driven zero restrictions on coefficients whose full-VAR $ t $ falls below a threshold, re-estimating the sparser system by restricted EGLS and reporting retained coefficients, the number of restrictions, and the SC/HQ/AIC improvement over the full VAR.	Full
Periodic VAR (seasonal coefficients)	Fits a periodic vector autoregression / periodic autoregression (PAR) in which the intercepts and lag coefficient matrices vary deterministically by season, returns the per-season coefficient matrices, and tests the constant-coefficient (standard-VAR) restriction via per-equation F tests and a system likelihood-ratio test of periodicity.	Full
Threshold Cointegration (Enders-Granger)	Tests for threshold cointegration with asymmetric error correction by fitting Enders-Granger TAR and momentum-TAR models to the cointegrating-residual changes, reporting the two regime adjustment speeds, the Phi no-cointegration statistic, and an F-test of symmetric vs. asymmetric adjustment.	Full
Generalized IRF (Koop-Pesaran-Potter)	Computes ordering-invariant generalized impulse-response functions for a VAR (Koop-Pesaran-Potter 1996; Pesaran-Shin 1998), tracing each variable's response to a shock in the chosen variable without imposing a recursive Cholesky ordering.	Full
LSTAR (Smooth-Transition AR)	Fits a logistic smooth-transition autoregression by nonlinear least squares, estimating two linear AR regimes plus the smoothness (gamma) and threshold (c) of a logistic transition over a lagged level, with the Luukkonen-Saikkonen-Terasvirta LM test of linearity against smooth-transition nonlinearity.	Full
Spectral Granger Causality (Geweke)	Decomposes Granger causality from one series to another by frequency using Geweke's (1982) spectral measure of linear feedback, reporting causality at low / business-cycle / high frequencies, the dominant frequency, and the frequency-averaged measure that matches the time-domain Granger test by Geweke's integral identity.	Full
Tsay Nonlinearity Test	Tsay (1986) F-test for threshold/quadratic nonlinearity in an AR(p) model: augments the linear AR with the squares and pairwise cross-products of the lagged values and tests their joint significance, returning the F statistic, degrees of freedom, p-value, and a linear/nonlinear verdict.	Full
BDS Test (nonlinearity)	Brock-Dechert-Scheinkman-LeBaron test for serial independence: computes the correlation-integral BDS statistic at embedding dimensions 2..max_dim and returns per-dimension $N(0,1)$ statistics with p-values plus an overall iid / non-iid verdict, detecting both linear and nonlinear dependence (e.g. in model residuals).	Full
Kalman Smoother (state-space)	Fits a local-level (or local-linear-trend) unobserved-components state-space model by maximum likelihood and returns both the real-time Kalman-filtered and the full-sample RTS-smoothed level (and slope), with estimated irregular/level/slope variances, log-likelihood and smoothed-state standard errors.	Full

PROCEDURE	WHAT IT DOES	STATUS
Time-Varying-Parameter Regression	Estimates regression coefficients that evolve over time as random walks via a linear-Gaussian state-space model fit by the Kalman filter/smoother, returning the smoothed time-varying coefficient paths, signal/state variances, and log-likelihood to detect structural change.	Full
IGARCH (Integrated GARCH)	Integrated GARCH(p,q) estimated by maximum likelihood with the variance persistence (sum of ARCH+GARCH coefficients) constrained to exactly 1, so volatility shocks are permanent rather than mean-reverting.	Full
GARCH-in-Mean (ARCH-M)	Estimates a GARCH-in-mean (ARCH-M) model in which the conditional volatility (std. dev.) or conditional variance enters the mean equation as a risk-premium term, reporting the lambda risk premium with significance alongside the omega/alpha/beta variance dynamics and persistence.	Full
FIGARCH (long-memory volatility)	Estimates a fractionally-integrated GARCH(p,d,q) by ML over a truncated ARCH-infinity expansion, recovering the long-memory parameter d in (0,1) that places volatility persistence between the GARCH (d=0) and IGARCH (d=1) extremes, with conditional-volatility summary, log-likelihood, AIC/BIC and a long-memory verdict.	Full
GARCH Portmanteau Diagnostics	Fits a GARCH(1,1) by Gaussian QMLE, forms standardized residuals, and tests them for remaining structure via Ljung-Box Q on z_t and z_t^2 and Engle ARCH-LM on z_t^2 at every lag, returning the variance parameters, per-lag statistics with p-values, and a pass/fail adequacy verdict for the volatility specification.	Full
Realized GARCH	Estimates the Realized GARCH(1,1) model of Hansen, Huang & Shek (2012), jointly fitting the variance recursion $h_t = \omega + \beta \cdot h_{t-1} + \gamma \cdot x_{t-1}$ and the measurement equation $x_t = \xi + \phi \cdot h_t + \tau(z_t) + u_t$ by Gaussian QMLE, reporting omega/beta/gamma, measurement parameters, persistence ($\beta + \gamma \cdot \phi$), log-likelihood, and the fitted conditional-volatility path (building a realized proxy from squared returns when no realized measure is supplied).	Full
CCC Multivariate GARCH	Bollerslev (1990) Constant Conditional Correlation multivariate GARCH: fits a univariate GARCH(1,1) to each series, estimates the constant correlation matrix R from the standardized residuals, and forms conditional covariances $H_t = D_t R D_t$ that are positive-definite by construction.	Full
HAR-RV (realized volatility)	Corsi (2009) Heterogeneous Autoregressive model of Realized Volatility: regresses next-period volatility on the daily (1-day), weekly (5-day) and monthly (22-day) trailing averages of a realized-volatility proxy, reporting component coefficients/SE/t/p, a joint significance F-test, R-squared and a one-step-ahead forecast.	Full
BEKK Multivariate GARCH	Fits a diagonal BEKK(1,1) multivariate GARCH (Engle & Kroner 1995) by Gaussian ML to 2-3 return series, estimating the diagonal A/B ARCH/GARCH parameters, implied per-series persistence, the time-varying conditional covariance/correlation path (guaranteed positive-definite at every t), the mean dynamic correlation, and the log-likelihood.	Full
Stochastic Volatility	Fits a stochastic-volatility model whose latent log-variance follows its own AR(1) state, estimated by Harvey-Ruiz-Shephard quasi-ML via the Kalman filter on log-	Full

PROCEDURE	WHAT IT DOES	STATUS
	squared returns, returning the persistence (ϕ), vol-of-vol (σ_{η}), long-run variance level (μ), and the Kalman-smoothed conditional-volatility path.	
Causality-in-Variance (Cheung-Ng)	Cheung-Ng (1996) test for volatility spillover between two series: GARCH-filters each series, then uses the n-scaled sum of squared cross-correlations of the squared standardized residuals to test, in each direction, whether past volatility of one series predicts the other's future volatility.	Full
FMOLS Cointegrating Vector	Phillips-Hansen Fully-Modified OLS estimates a single cointegrating vector among I(1) series, correcting OLS for serial correlation and regressor endogeneity via a Bartlett long-run covariance to yield an asymptotically efficient, unbiased long-run coefficient table with valid standard errors.	Full
DOLS Cointegrating Vector	Stock-Watson (1993) single-equation Dynamic OLS estimate of a cointegrating vector, regressing the I(1) dependent on the I(1) regressors plus leads and lags of their first differences, with HAC (Newey-West) standard errors on the asymptotically efficient long-run coefficients.	Full
Markov Chain (transition/stationary)	Estimates a first-order discrete Markov chain from a state sequence by maximum-likelihood transition counting, returning the row-normalized transition matrix, the stationary (invariant) distribution solving $\pi = \pi P$, the n-step transition matrix P^n , and Kac mean recurrence / mean first-passage times.	Full
Phillips-Ouliaris Cointegration Test	Residual-based test of the no-cointegration null among I(1) series: runs the cointegrating OLS, then applies the Phillips-Perron Z-alpha and Z-t unit-root statistics to the residuals (Newey-West long-run-variance corrected) against MacKinnon response-surface cointegration critical values to deliver Z-alpha, Z-t, an approximate p-value, and a cointegration verdict.	Full
Blanchard-Quah SVAR	Identifies permanent and transitory structural shocks in a bivariate (or n-variable) VAR via Blanchard-Quah long-run restrictions, returning the lower-triangular long-run impact matrix, the contemporaneous impact matrix, orthonormal structural impulse responses, and the forecast-error variance decomposition.	Full
VAR Residual Diagnostics (multivariate)	Fit a VAR(p) and run a battery of multivariate residual checks - the Hosking (1980) portmanteau/Ljung-Box test for serial correlation, a multivariate ARCH-LM serial-correlation test, and the Doornik-Hansen omnibus normality test - returning each statistic, df, p-value, and an overall model-adequacy verdict.	Full
VAR Lag-Order Selection (IC table + LR)	Fits a VAR(p) for every lag $p=0..\text{maxlags}$ on a common fixed sample and reports the AIC, Hannan-Quinn, Schwarz/BIC and FPE information criteria plus a sequential likelihood-ratio test, returning each criterion's argmin and the LR-selected lag.	Full
ARCH-LM Test (ARCH effects)	Engle's LM test for conditional heteroskedasticity (volatility clustering): n·R ² auxiliary regression of the squared series on its own lags. Run before fitting a GARCH model.	Full
Time-trend regression (detrending)	Regress a series on a polynomial deterministic time trend; reports trend coefficients, R ² , overall trend significance, and detrended-residual summary (trend- vs difference-stationary).	Full

PROCEDURE	WHAT IT DOES	STATUS
Spurious-regression diagnostic	Regress y on x and flag the classic spurious-regression signature (high R^2 , very low Durbin-Watson, non-stationary residuals via Engle-Granger ADF) for integrated, non-cointegrated series.	Full
Seasonal ARIMA (SARIMA)	Fit seasonal ARIMA with forecasting, including Ljung-Box and Jarque-Bera diagnostics.	Full
Holt-Winters Exponential Smoothing	Exponential smoothing with additive/multiplicative seasonal component and trend.	Full
Error-Trend-Seasonal (ETS) model	Automated error-trend-seasonal decomposition with AIC/BIC selection.	Full
Unobserved Components / State Space	Kalman filter state-space model with level, trend, and seasonal components.	Full
Structural VAR (SVAR)	Structural VAR with estimated A and B matrices for shock identification.	Full
Impulse Response Functions	Orthogonalized and non-orthogonalized IRF from VAR with cumulative responses.	Full
Forecast Error Variance Decomposition	Variance decomposition of VAR forecasts by shock source over horizons.	Full
Markov Switching Autoregression	Regime-switching model with time-varying transition probabilities and smoothed probabilities.	Full
Threshold Autoregression (TAR)	Nonlinear autoregression with endogenous threshold detection and linearity test.	Full
Dynamic Conditional Correlation GARCH	DCC-GARCH model for time-varying correlations (Engle 2002) via quasi-MLE.	Full
Zivot-Andrews unit root test	Unit root test with one endogenous structural break in level or trend.	Full
DF-GLS unit root test	Elliott-Rootenberg-Stock feasible GLS-detrended unit root test.	Full
ERS Point Optimal unit root test	Efficient detrending with GLS and point-optimal deviations under the null.	Full
Change-in-mean structural break detection	Detects shifts in the MEAN of one series via dynamic-programming least-squares (L2) segmentation, with the number of breaks chosen by BIC; reports segment means and variances.	Full
Quandt-Andrews Break Test (sup-Wald)	Tests for a single structural break at an unknown date in a linear regression by computing the Chow/Wald F for a full-coefficient break at every trimmed candidate point and reporting the sup-F (QLR), average-F and exp-F (Andrews-Ploberger) statistics, the estimated break fraction, Andrews (1993) asymptotic critical values, and a Hansen (1997) approximate p-value.	Full
CUSUM and CUSUM-squared stability test	Recursive residual stability test for parameter constancy over time.	Full

PROCEDURE	WHAT IT DOES	STATUS
Particle Filter (SMC state-space)	Bootstrap particle filter (sequential Monte Carlo) for a univariate local-level state-space model: estimates the process/observation variances by maximum likelihood on the exact Kalman likelihood, then runs an SIR filter with systematic resampling to return the filtered latent-state mean, a 90% credible band, the one-step predictive log-likelihood, and the effective sample size over time, benchmarked against the exact Kalman filter.	Full
Koyck Distributed Lag	Estimates a Koyck (geometric) infinite distributed-lag model $y_t = \alpha + \beta x_t + \lambda y_{t-1}$ via OLS on the Koyck transform, reporting the impact multiplier, geometric decay rate, long-run multiplier $\beta/(1-\lambda)$, mean/median lag, the implied lag-weight profile, and a Durbin's-h serial-correlation diagnostic.	Full
Toda-Yamamoto Granger causality	Granger causality test in levels with integration-order adjustment; Wald test on VAR lags.	Full
Nonlinear ARDL (NARDL)	Autoregressive distributed lag with asymmetric positive/negative decomposition; symmetry test.	Full
Pesaran-Shin-Smith ARDL bounds test	Cointegration test via F-test on lagged levels in error-correction form.	Full
Diebold-Mariano forecast comparison	Test comparing forecast accuracy of two models (MSE or MAE) with Newey-West variance.	Full
Forecast combination	Combine forecasts via equal weight, inverse MSE, OLS, or Bates-Granger method.	Full
Diffusion-index direct forecast (user-supplied factors)	Direct multi-step diffusion-index forecast $y_{t+h} = a + \sum g_p y_{t-p} + b'F_t$ using factor columns you supply (e.g. PCA component scores from <code>analysis.pca</code>), with an out-of-sample backtest against an AR benchmark.	Limited
Dynamic Factor Model (EM/Kalman)	Estimate a latent dynamic factor model on multiple standardized time series by EM with the Kalman filter/RTS smoother (Doz-Giannone-Reichlin / Watson-Engle), extracting common AR factors and reporting loadings, the smoothed factor path, factor AR coefficients, variance shares, log-likelihood and iteration count.	Full
Spectral analysis	Periodogram or Welch PSD with dominant frequency and period identification.	Full
Hamilton Filter (trend/cycle)	Hamilton (2018) regression-based business-cycle filter that decomposes a series into trend (the h-step-ahead forecast fitted from a constant and p lagged levels) and cycle (the forecast residual), reporting coefficients, R^2 and the cycle standard deviation as a sound, spurious-dynamics-free replacement for the Hodrick-Prescott filter.	Full
Spectral Coherence (cross-spectrum)	Estimates the squared coherency and phase spectrum between two time series via a smoothed Welch/averaged-periodogram cross-spectrum with a Hann taper, reporting coherence and phase by frequency, the peak-coherence frequency, and an approximate significance threshold.	Full
Unrestricted MIDAS regression	Mixed-frequency regression: unrestricted lags of high-frequency predictor on low-frequency outcome.	Limited

PROCEDURE	WHAT IT DOES	STATUS
Almon polynomial MIDAS	MIDAS with polynomial lag weighting via NLS; information criterion comparison across lag lengths.	Limited
Beta polynomial MIDAS	MIDAS with beta lag structure for smooth, parsimonious high-frequency aggregation.	Limited
Exponential Almon MIDAS	MIDAS using exponential-Almon lag polynomial for flexible decay pattern.	Limited
MIDAS forecast	Forecast using fitted MIDAS model; extends from low-freq observations with high-freq predictors.	Full
MIDAS nowcasting	Real-time estimate of current low-frequency value using available high-frequency data.	Limited
MIDAS model comparison	Compare U-MIDAS, Almon, Beta, and Exponential MIDAS across lag range via information criteria.	Limited
Real Business Cycle model	Canonical RBC with technology shock; analytical steady state, IRF, variance decomposition.	Limited
New Keynesian 3-equation model	IS–Phillips–Taylor system; parameterizable demand, supply, monetary shocks with closed-form IRF.	Limited
Linear rational expectations solver	Klein-Sims QZ method for general linear RE systems; determines Blanchard-Kahn determinacy.	Full
DSGE simulation	Forward simulation of solved DSGE policy matrices under specified or random shock sequence.	Full
DSGE impulse response	Impulse response to single shock in solved DSGE model; orthogonal identification.	Full
DSGE variance decomposition	Forecast error variance share by shock at specified horizons in DSGE.	Full
DSGE calibration	Heuristic moment-matching calibration for RBC and NK; suggests parameters given targets.	Limited

Panel Data & Causal Inference

51

Statio supports canonical panel data estimators (FE, RE, Hausman test, 2SLS-IV), advanced dynamic panel methods (Arellano-Bond, System GMM), and modern causal inference designs (DiD, RDD, synthetic control, staggered DiD, IPW, mediation analysis).

PROCEDURE	WHAT IT DOES	STATUS
Fixed Effects Panel Regression	Within-transformation OLS eliminating time-invariant unobserved heterogeneity. Optional two-way (entity + time) fixed effects.	Full
Random Effects Panel Regression	GLS panel estimator assuming random effects uncorrelated with regressors.	Full
Hausman Specification Test	Tests FE vs RE using Moore-Penrose generalized inverse when covariance difference non-positive semidefinite (finite-sample robust).	Full
Two-Stage Least Squares (2SLS) IV	Instrumental variables regression with heteroskedasticity-robust first-stage F-test and weak-instrument diagnostics.	Full
Arellano-Bond Difference GMM	Dynamic panel estimator using lagged levels as instruments for differenced equations; includes Sargan and AR(1)/AR(2) tests.	Full
Blundell-Bond System GMM	Two-step efficient GMM stacking level and difference equations; Hansen J test and AR tests.	Full
Hausman-Taylor Estimator	Estimates time-varying exogenous, time-invariant exogenous, and endogenous coefficients using within and between equations.	Full
Between Estimator	OLS on entity-level means; removes within-group variation.	Full
First-Difference Estimator	Eliminates time-invariant heterogeneity via differencing; OLS on changes.	Full
Panel-Corrected Standard Errors (Beck-Katz PCSE)	Kernel-based covariance estimator robust to contemporaneous and lagged cross-sectional dependence.	Full
Driscoll-Kraay Standard Errors	Robust to both cross-sectional and temporal dependence via time clustering.	Full
Panel Cointegration Tests	Pedroni, Kao, or Fisher method tests for long-run equilibrium relationships in panel data.	Full
Panel Granger Causality (Fisher Combined)	Granger test per entity combined via Fisher's method across units.	Full
Cross-Sectional Dependence Test	Pesaran CD, Frees, or Friedman test for contemporaneous cross-unit correlation.	Full
Panel ARDL (Mean Group / Dynamic Fixed Effects)	Panel autoregressive distributed lag models: Mean Group (averages unrestricted per-entity long-run estimates) and Dynamic Fixed Effects (pooled with entity intercepts). Reports short-run, long-run, and ECM terms.	Full
Mundlak (Correlated Random Effects)	RE model augmented with entity means; joint test on means equivalent to Hausman test; robust to cluster heteroskedasticity.	Full

PROCEDURE	WHAT IT DOES	STATUS
Panel RE Test (Breusch-Pagan LM)	Breusch-Pagan Lagrange-multiplier test for entity random effects in a panel: from pooled-OLS residuals it forms the two-sided BP-LM (chi-squared, 1 df) and the more powerful one-sided Honda statistic (normal) from grouped within-entity residual sums, handles balanced and unbalanced panels, and verdicts whether pooled OLS is adequate or a random/fixed-effects estimator is required.	Full
Panel Serial-Correlation Test (Wooldridge)	Wooldridge (2002) F-test for first-order serial correlation in the idiosyncratic errors of a linear panel model -- first-differences the equation, regresses the differenced residuals on their own lag and tests whether the slope equals the no-serial-correlation value of -0.5 (cluster-robust, as Stata xtserial), and also reports the Baltagi-Wu LBI.	Full
Anderson-Hsiao Dynamic Panel	Anderson-Hsiao IV estimator for dynamic panels: first-differences out the fixed effect, then instruments the differenced lagged dependent with its t-2 level via 2SLS, recovering a consistent persistence coefficient where pooled OLS and within are biased.	Full
FOD-GMM Dynamic Panel (Helmert)	Arellano-Bover (1995) forward-orthogonal-deviations (Helmert) GMM for dynamic panels: removes fixed effects via the FOD transform (which keeps the idiosyncratic errors serially uncorrelated, so the one-step weight is simply $(Z'Z)^{-1}$), instruments the transformed lagged dependent with the expanding set of lagged levels, and reports coefficient estimates (coef/se/z/p) with entity-clustered robust SEs and a Hansen J over-identification test - markedly less biased than the within estimator for short T.	Full
Bias-Corrected FE (dynamic panel)	Within (LSDV) fixed-effects estimation of a dynamic panel with a lagged dependent variable, then the analytical Nickell (1981) / Kiviet (1995) correction of the finite-T within (Nickell) bias subtracted from the autoregressive coefficient, returning the uncorrected and bias-corrected persistence plus the bias estimate.	Full
Parks-Kmenta FGLS (long panel)	Feasible GLS for long balanced panels ($T > N$) that simultaneously corrects for panel heteroskedasticity, contemporaneous cross-sectional correlation, and panel-specific AR(1) serial correlation, returning efficient slope estimates, per-entity AR(1) coefficients, and Parks-Kmenta standard errors (with the standard over-optimism caveat).	Full
Common Correlated Effects (CCE)	Pesaran (2006) CCE estimator for heterogeneous panels with cross-sectional dependence: augments each unit's regression with cross-sectional averages of the dependent and regressors to proxy and absorb the unobserved common factors, then reports the Mean-Group (CCEMG) or Pooled (CCEP) slopes, which remain consistent where plain fixed effects is biased by the factor.	Full
Panel Quantile Regression (Canay)	Estimates fixed-effects panel quantile regression via Canay's (2011) two-step estimator - removing each entity's location-shift fixed effect from a within mean regression, then running ordinary quantile regression at tau on the de-fixed data -- yielding slope effects on the conditional tau-quantile that are robust to entity fixed effects (unlike biased pooled quantile regression).	Full
Error Components + AR(1) (Baltagi-Li)	One-way random-effects panel whose idiosyncratic error follows an AR(1) process: estimates the autocorrelation parameter rho, applies the Baltagi-Li Prais-Winsten transform and Sherman-Morrison FGLS, reports the variance components, the FGLS coefficient table, and a test of rho=0 (no serial correlation).	Full
Pooled Mean Group (PMG)	Estimates a dynamic heterogeneous panel ARDL error-correction model with common (pooled) long-run coefficients and unit-specific short-run dynamics and adjustment speeds via the Pesaran-Shin-Smith concentrated-likelihood iteration, reporting pooled	Full

PROCEDURE	WHAT IT DOES	STATUS
	long-run coef/se/z/p, the average error-correction speed, per-unit speeds, and a Hausman MG-vs-PMG poolability test.	
MINQUE/MIVQUE Variance Components	Estimates the between-entity and idiosyncratic variance components of a one-way error-components panel by Rao's non-iterative Minimum-Norm/Minimum-Variance Quadratic Unbiased estimator (handles unbalanced panels cleanly, unlike ANOVA), reports the implied intraclass correlation, and forms the FGLS slope estimates.	Full
Random-Trend Panel	Fixed-effects panel model that gives every entity its own intercept AND its own linear time trend (swept out via a double-within detrending transform), recovering the common predictor slopes purged of the unit-specific trends that would bias a plain one-way FE estimator.	Full
BLUP of Random Effects	Predicts each entity's unobserved random effect after a one-way random-effects (GLS) panel fit by shrinking its mean composite residual toward zero with the Henderson BLUP factor $\theta_i = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2 / T_i)$, reporting per-entity BLUPs, the shrinkage factor, and the variance components.	Full
Two-Way Random Effects	Two-way error-components GLS panel model with random entity AND time effects, estimating the entity, time, and idiosyncratic variance components (Wallace-Hussain) and reporting the Fuller-Battese theta transforms and FGLS coefficients.	Full
Panel Unit Root (IPS / LLC)	First-generation panel unit-root tests on a long panel: the Im-Pesaran-Shin $\bar{W}$ -bar statistic (standardised average of per-entity augmented Dickey-Fuller t-stats, allowing heterogeneous AR roots) plus the Levin-Lin-Chu pooled t^* (common- ρ , bias-adjusted), each returned with its $N(0,1)$ p-value, the per-entity ADF t-stats and a stationary/unit-root verdict.	Full
Panel Cross-Sectional Dependence (BP-LM)	Tests panel errors for cross-sectional dependence from within-model residuals: the Breusch-Pagan LM (sum of T-weighted squared pairwise residual correlations $\sim \chi^2(N(N-1)/2)$), the Pesaran (2004) scaled LM, the Pesaran-Ullah-Yamagata (2008) bias-corrected scaled LM, and the reference Pesaran CD statistic, each with p-values and a verdict.	Full
CIPS Panel Unit Root (2nd-gen)	Pesaran (2007) cross-sectionally augmented IPS panel unit-root test, robust to cross-sectional dependence from a common factor: it averages per-entity CADF t-stats (each ADF regression augmented with the cross-section means of the lagged level and the differences) and compares the CIPS statistic against the non-standard Pesaran critical values.	Full
Panel Poolability (Chow-F)	Chow-type F test of panel poolability comparing restricted pooled-OLS SSR against the sum of unrestricted per-entity OLS SSRs to assess whether intercepts and slopes are homogeneous across entities, $F \sim F((G-1)(k+1), N-G(k+1))$.	Full
Difference-in-Differences (DiD)	2x2 DiD with cluster-robust or HC1 SEs; includes group means and specification.	Full
Regression Discontinuity (Sharp RDD)	Local polynomial with triangular or uniform kernel; ad-hoc $1.5 \times \text{SD}$ bandwidth (not MSE-optimal) or user-specified.	Limited
Propensity Score Matching (PSM)	Logistic propensity score with 1-NN matching on logit scale; caliper on logit scale (Austin 2011); naive paired SE (Abadie-Imbens not applied).	Limited

PROCEDURE	WHAT IT DOES	STATUS
Chow Structural Break Test	Tests equality of regression coefficients across two subsamples; F-statistic with df protection.	Full
Synthetic Control Method	Weighted combination of donors minimizing pre-treatment MSE; RMSPE ratio and period-level treatment effects.	Full
Event-study (dynamic two-way fixed effects)	Conventional dynamic TWFE event study: two-way FE with relative-time dummies; pre-treatment trend test; reference period omitted.	Full
Staggered DiD (Callaway & Sant'Anna 2021)	Group-time ATT estimates with not-yet-treated comparison group; bootstrap SEs; dynamic, group, and calendar aggregations.	Full
IV Diagnostics	First-stage F-tests, Cragg-Donald Wald F, Sargan over-identification, Durbin endogeneity test.	Full
Inverse-Probability Weighting (IPW)	Logistic propensity score weighting; ATE, ATT, or ATU estimands.	Full
Doubly Robust Estimation	Outcome model + propensity weighting; consistent if either model correct.	Full
Local Average Treatment Effect (LATE)	IV-based LATE for compliers only; asymptotic SE under 2SLS.	Full
Classical Mediation Analysis	Classical Baron-Kenny path decomposition (direct/indirect effects) with a Sobel test and a percentile bootstrap CI for the indirect effect.	Full
Sensitivity Analysis	E-value or Rosenbaum bounds for unobserved confounding robustness.	Full
Lee Bounds	Nonparametric bounds under sample selection on unobservables.	Full
Manski Treatment Bounds	Worst-case partial-identification (Manski 1990) bounds on the average treatment effect of a binary treatment, with monotone-treatment-response and monotone-treatment-selection refinements that tighten the identified ATE interval.	Full
Sharp Regression Kink Design (RKD)	Sharp RKD: estimates the change in the outcome's slope at the threshold via piecewise-linear local regression (assumes a known unit kink in the assignment-to-policy map).	Full
Bunching Estimator (normalized excess mass)	Estimates excess mass at a policy threshold by fitting a counterfactual polynomial density excluding the bunching region; reports the normalized excess mass (bunching ratio).	Full

Multivariate, Factor & Latent-Variable Analysis 35

Covers exploratory and confirmatory dimensionality reduction, clustering, reliability, latent structure discovery, and item response modeling — from classical PCA/factor analysis through modern IRT, SEM, and categorical methods.

PROCEDURE	WHAT IT DOES	STATUS
Principal Component Analysis	Eigenvalue-based dimensionality reduction via correlation-matrix SVD with Kaiser criterion for retention.	Full
Exploratory Factor Analysis	ML factor extraction with Varimax/Promax rotation, KMO and Bartlett's sphericity test.	Full
K-Means Clustering	Standardized iterative partition with silhouette score; sample-based for n > 10k.	Full
Hierarchical Agglomerative Clustering	Distance-based clustering with Ward/complete/average linkage; limited to n ≤ 20k (O(n ²) memory).	Limited
Cronbach's Alpha	Internal consistency reliability via item correlations with item-total and alpha-if-deleted.	Full
McDonald's Omega (Total)	ML single-factor CFA reliability coefficient without tau-equivalence assumption; requires k ≥ 3 items.	Full
Split-Half Reliability	Odd-even split with Spearman-Brown and Guttman corrections.	Full
Parallel Analysis	Horn's method: compares observed eigenvalues against 95th percentile of random-data eigenvalues.	Full
Multivariate Analysis of Variance	One-way MANOVA with Wilks' lambda, Pillai trace, Hotelling-Lawley, Roy's root and F-approximations.	Full
Linear Discriminant Analysis	Fisher discriminant functions and linear classification functions with resubstitution accuracy.	Full
Canonical Correlation Analysis	Maximum correlation between two variable sets via CCA with Rao's F-approximation for Wilks' lambda.	Full
Confirmatory Factor Analysis	Measurement model via semopy with factor loadings, residuals, and model fit (CFI, TLI, RMSEA, SRMR).	Full
Structural Equation Modeling	Full latent-variable SEM (measurement + structural paths) via semopy with standardized estimates.	Full
Path Analysis	Single-level directional relationships with path coefficients, R ² per endogenous, and overall fit.	Full
Rasch Model (1PL IRT)	Joint MLE for dichotomous items; person abilities, item difficulties, infit/outfit, reliability.	Full
2PL IRT Model	Two-parameter logistic with discrimination + difficulty; uses girth if available, else manual L-BFGS.	Full
3PL IRT Model	Three-parameter logistic (discrimination, difficulty, guessing) via girth or manual EM.	Full
Graded Response Model (Samejima)	Ordinal polytomous items with cumulative thresholds per item; category response curves and item info.	Full
Partial Credit Model (Masters)	Polytomous step parameters without item discrimination; flexible by-item category counts.	Full

PROCEDURE	WHAT IT DOES	STATUS
Rating Scale Model (Andrich)	Polytomous with common thresholds across items and item-specific location parameters.	Full
Nominal Response Model (Bock)	Unordered categorical responses with per-category slopes and intercepts per item.	Full
Differential Item Functioning Analysis	Mantel-Haenszel, logistic, or Lord's chi-squared DIF with ETS delta classification.	Full
Classical Test Theory Analysis	Item difficulty, corrected item-total correlation, and alpha-if-deleted per item.	Full
IRT Fit Indices (Approximate)	Limited-information chi-square, approximate RMSEA/CFI, Q3 local independence (NOT M2-based).	Limited
Multidimensional Scaling	Metric or non-metric embedding via Euclidean distances; stress and Shepard diagram.	Full
Simple Correspondence Analysis	SVD on contingency-table chi-square distances for two categorical variables.	Full
Multiple Correspondence Analysis	Indicator-matrix CA for $k \geq 2$ categorical variables; uses prince or manual SVD.	Full
Gaussian Finite Mixture Model	k-component Gaussian mixture with BIC/AIC selection, posteriors, and component weights.	Full
Latent Class Analysis	EM-estimated discrete latent classes for binary/categorical indicators; entropy and model fit.	Full
Latent Profile Analysis	Diagonal-covariance GMM for continuous indicators with entropy, classification table.	Full
Categorical Regression (CATREG)	ALS optimal scaling for nominal/ordinal predictors with regression fit on quantified variables.	Full
Categorical PCA (CATPCA)	Optimal scaling PCA with ALS alternation; preserves ordinal/nominal structure.	Full
OVERALS (Non-linear CCA)	Multi-set non-linear canonical correlation via ALS with set-wise quantifications.	Full
Optimal Scaling Transform	Applies nominal/ordinal/spline quantifications to variables without modeling.	Full
Categorical Correlation	Pearson correlations on optimally-scaled categorical variables.	Full

Survival, Bayesian, Meta-Analysis & Resampling 68

Statistical methods for survival analysis, Bayesian inference, meta-analytic pooling, and resampling-based inference. Includes Kaplan-Meier curves, Cox proportional-hazards regression, parametric and competing-risks models, Bayesian linear/logistic/hierarchical regression with MCMC diagnostics, fixed- and random-effects meta-analysis, and bootstrap/jackknife/permutation methods.

PROCEDURE	WHAT IT DOES	STATUS
Kaplan-Meier Survival Curves	Non-parametric survival estimation with Greenwood pointwise 95% CI and log-rank/multivariate log-rank test.	Full
Cox Proportional Hazards	Semi-parametric hazard regression with Efron tie handling, PH assumption test (Grambsch-Therneau), and concordance index.	Full
Mixture Cure Model	Fits a Berkson-Gage mixture cure model by EM, separating a logistic incidence submodel for the long-term-survivor (cured) fraction from a Weibull AFT latency submodel for the susceptible subjects who can still experience the event.	Full
Piecewise-Exponential Survival	Proportional-hazards survival model with a piecewise-constant baseline hazard, fitted by the Poisson / piecewise-exponential equivalence on person-interval exposure with covariate log-hazard-ratios; reduces to the exponential model with one interval.	Full
Discrete-Time Hazard (cloglog)	Fits a discrete-time (grouped-duration) proportional-hazards model by expanding subjects into person-period rows and estimating a complementary-log-log binomial GLM, yielding constant-across-time log-hazard-ratios that mirror the Cox model at low per-period hazard, plus a nonparametric or polynomial baseline-hazard summary.	Full
Actuarial Life Table	Cutler-Ederer life table with conditional survival/death probabilities, cumulative hazard rate, and Greenwood SE.	Full
Parametric Survival Models	Univariate or AFT parametric models (Weibull, exponential, lognormal, log-logistic, Gompertz) with parameter estimates.	Full
Accelerated Failure Time	AFT regression with covariates, time ratio estimates, concordance index, and AIC.	Full
Competing Risks (Fine-Gray)	Subdistribution-hazards model with cumulative incidence functions per event type and subdistribution HR.	Full
Shared Frailty Cox Model	Cluster-robust Cox with gamma frailty variance estimate and LR test vs no-frailty baseline.	Limited
Schoenfeld Residuals Test	PH assumption test per covariate and globally using scaled Schoenfeld residuals with interpretation.	Full
Nelson-Aalen Cumulative Hazard	Non-parametric cumulative hazard estimator with 95% pointwise confidence bands, optionally grouped.	Full
Stratified Log-Rank Test	Log-rank test with optional stratification, multivariate for 3+ groups, and pairwise comparisons.	Full
Restricted Mean Survival Time	RMST up to user-specified tau with SE and 95% CI; difference test and CI for 2 groups.	Full
Time-Varying Cox Model	Cox regression with time-varying covariates (start-stop format), HR with CI and partial log-likelihood.	Full
Cox Residuals for Diagnostics	Five residual types (martingale, deviance, Schoenfeld, scaled Schoenfeld, score) for model diagnostics.	Full
Bayesian Linear Regression	MCMC-based (PyMC or emcee fallback) with weakly informative priors, posterior samples, PPC, and diagnostics (R-hat, ESS).	Full

PROCEDURE	WHAT IT DOES	STATUS
Bayesian Unit-Root Analysis	Closed-form Student-t posterior for the AR(1) root ρ under a flat/Jeffreys prior, reporting the posterior mean/sd, a 95% credible interval, and the posterior probabilities $P(\rho \geq 1)$ (unit root/explosive) and $P(\rho < 1)$ (stationary).	Full
Bayesian VAR (Minnesota prior)	Estimates a reduced-form VAR(p) with the Minnesota (Litterman) shrinkage prior imposed by dummy observations, reporting posterior-mean coefficient matrices that center the own first lag at 1 (levels) or 0 (stationary), shrink cross-variable lags harder than own lags by $1/\text{lag}$, plus the shrinkage hyperparameters, per-equation in-sample fit, and a multi-step forecast.	Full
Bayesian VECM (cointegration)	Bayesian Vector Error Correction Model for cointegrated series: given a chosen cointegrating rank and lag order, places a weakly-informative Normal-inverse-Wishart prior on the short-run dynamics and adjustment coefficients and Gibbs-samples the posterior, reporting posterior means and 95% credible intervals for the speed-of-adjustment loadings (α), the normalised cointegrating vectors (β), the short-run matrices, and error-correction evidence.	Full
Bayesian CFA (one-factor)	Fits a one-factor confirmatory measurement model $x_{ij} = \lambda_j \cdot \eta_i + \epsilon_{ij}$ by a three-block Gibbs sampler (factor scores, loadings, residual variances), returning posterior loading summaries with 95% credible intervals, residual (uniqueness) variances, implied communalities, and an SRMR-based fit summary.	Full
Bayesian IRT (2PL)	Bayesian two-parameter logistic item-response model fitted by a Metropolis-within-Gibbs MCMC sampler, returning posterior item discriminations and difficulties (with 95% credible intervals) plus latent person-ability summaries on the $N(0,1)$ -identified scale.	Full
Bayesian Meta-Analysis	Fits a random-effects meta-analysis ($\theta_i \sim N(\mu, \tau^2)$, $y_i \sim N(\theta_i, \text{se}_i^2)$) by a direct Gibbs sampler with a weakly-informative half-Cauchy prior on the between-study sd, returning the posterior overall effect μ with credible interval, between-study heterogeneity τ , shrunk per-study estimates, and the posterior probability the effect is positive.	Full
Conjugate Bayesian Regression (Normal-Gamma)	Analytic conjugate Bayesian linear regression with a Normal-Gamma (Normal-inverse-Gamma) prior that returns closed-form multivariate-t posteriors for the coefficients and an inverse-Gamma posterior for the error variance, reducing exactly to OLS under a flat prior and shrinking slopes toward zero under a tight prior, with no MCMC.	Full
Bayesian Quantile Regression	Estimates the conditional tau-quantile of an outcome via the asymmetric-Laplace likelihood, sampling the posterior of the quantile-regression coefficients with the Kozumi-Kobayashi (2011) location-scale-mixture Gibbs sampler and reporting posterior means, sds and 95% credible intervals.	Full
Student-t (robust) Regression	Outlier-robust linear regression with iid Student-t(ν) errors fitted by an EM / IRLS scale-mixture ML algorithm, downweighting extreme residuals so the fit tracks the bulk of the data rather than the outliers, and reporting coefficients/SEs, the error scale, the (optionally ML-estimated) degrees of freedom, log-likelihood and per-observation robustness weights.	Full
Metropolis-Hastings (Bayesian regression)	Random-walk Metropolis-Hastings MCMC sampler for the joint posterior of regression coefficients and the error variance (β , $\log \sigma^2$) in Bayesian linear regression, with a proposal auto-tuned toward the optimal ~ 0.234 acceptance rate, reporting posterior means/sds/95% credible intervals, the acceptance rate, and effective sample sizes.	Full

PROCEDURE	WHAT IT DOES	STATUS
Gibbs Sampler (Bayesian regression)	Conjugate Gibbs sampler for Bayesian linear regression that alternates Normal draws of the coefficients given σ^2 and inverse-Gamma draws of σ^2 given the coefficients, returning posterior means/sds/95% credible intervals, effective sample size and a Geweke convergence diagnostic (acceptance is identically 1), with weak-prior posteriors that coincide with OLS.	Full
Laplace Approximation (Bayesian)	Approximates the posterior of a Bayesian linear or logistic regression by a Gaussian centered at the posterior mode (MAP) with inverse-Hessian covariance, reporting approximate posterior means/SDs/credible intervals, the Laplace approximation to the log marginal likelihood (model evidence), and a Tierney-Kadane fully-exponential refined posterior mean.	Full
Importance Sampling (Bayesian)	Bayesian posterior inference for a linear regression by importance sampling / sampling-importance-resampling: proposals drawn from a heavy-tailed multivariate-t centred at OLS (covariance inflated) under a weakly-informative conjugate Normal-inverse-Gamma prior are reweighted to give importance-weighted posterior means, 95% credible intervals for the coefficients and error variance, the weights' effective sample size, and a resampled (SIR) posterior draw.	Full
Bayesian Poisson Regression	Fits a log-linear Poisson count model by random-walk Metropolis-Hastings under a weak Normal prior, returning posterior means, standard deviations and 95% credible intervals for each coefficient, the incidence-rate ratios $\exp(\text{coef})$, and the Metropolis acceptance rate (tuned toward the ~ 0.234 optimal scaling).	Full
Bayesian Tobit	Left-censored (Tobit) regression estimated by Gibbs sampling with data augmentation of the censored latent responses, returning posterior means and 95% credible intervals for the coefficients and error standard deviation plus the censored fraction.	Full
Bayesian Probit (Albert-Chib)	Fits a binary probit model by Gibbs sampling with Albert-Chib truncated-normal latent-utility data augmentation under a diffuse Normal prior, returning posterior means, SDs, 95% credible intervals, effective sample sizes, and the frequentist MLE as a reference.	Full
Horseshoe Regression (sparse)	Bayesian linear regression with a horseshoe shrinkage prior (Carvalho-Polson-Scott 2010) sampled by the fully-conjugate Makalic-Schmidt (2016) inverse-gamma scale-mixture Gibbs sampler, reporting posterior means and 95% credible intervals per coefficient, per-coefficient shrinkage weights $\kappa = 1/(1 + n \cdot \tau^2 \cdot \lambda^2)$, and which coefficients are effectively selected (CI excludes 0) so that on sparse designs the null coefficients are shrunk hard toward 0 while genuine signals stay near their true values.	Full
Spike-and-Slab Variable Selection	Bayesian stochastic search variable selection (George-McCulloch SSVS) that places a two-component spike-at-zero / diffuse-slab prior on every regression coefficient and Gibbs-samples Bernoulli inclusion indicators to report each predictor's posterior inclusion probability (PIP), Bayesian-model-averaged coefficients, and the most probable visited models.	Full
Bayesian Lasso	Bayesian lasso regression (Park & Casella) places a Laplace prior on the coefficients and samples it with a scale-mixture-of-normals Gibbs sampler, returning shrunk posterior-mean coefficients with 95% credible intervals, the posterior of the shrinkage parameter λ , and which predictors are effectively selected (CI excludes 0).	Full
Adaptive Lasso	Weighted lasso (Zou 2006) whose per-coefficient penalties come from an initial OLS/ridge estimate, delivering oracle variable selection: truly-zero coefficients are driven	Full

PROCEDURE	WHAT IT DOES	STATUS
	to exactly 0 while genuine signals stay nearly unbiased, with the penalty chosen by reproducible cross-validation.	
Savage-Dickey Bayes Factor	Computes the Savage-Dickey density-ratio Bayes factor (BF ₀₁ /BF ₁₀) for the sharp null H ₀ : $\beta = 0$ on one regression coefficient by taking the ratio of the marginal posterior to the prior density at zero under a conjugate Normal-prior Bayesian regression, returning posterior mean/sd, the prior sd, and a Jeffreys-scale verdict.	Full
Bayesian IC (DIC/WAIC/LOO)	Bayesian predictive information criteria for a conjugate-Gibbs Bayesian linear regression, reporting DIC (Spiegelhalter), WAIC (Watanabe) and Pareto-smoothed importance-sampling LOO (Vehtari-Gelman-Gabry) on the deviance scale with their effective-parameter counts (p_DIC, p_WAIC, p_LOO) and Pareto-k diagnostics, for principled out-of-sample model comparison.	Full
Bayesian Model Averaging	Enumerates all 2^k linear-regression covariate subsets, scores each by BIC to obtain approximate posterior model probabilities, and reports per-predictor posterior inclusion probabilities (PIP), BMA posterior means/sds of the coefficients (averaged over models), and the top models.	Full
Bayesian Logistic Regression	Bernoulli likelihood with logit link, posterior odds ratios, in-sample accuracy, and MCMC diagnostics.	Full
Bayesian Hierarchical Model	Random-intercept multilevel model (PyMC, emcee, or OLS shrinkage fallback) with ICC, varying intercepts per group.	Limited
MCMC Diagnostics	Gelman-Rubin R-hat, ESS via autocorrelation, trace summaries with interpretations of convergence.	Full
Bayes Factor	Model comparison via BIC approximation with Jeffreys-scale interpretation (decisive/strong/moderate/anecdotal).	Limited
Prior vs Posterior	Visualization comparison of prior and posterior distributions per parameter.	Full
Posterior Predictive Check	Bayesian p-value and visual comparison of observed vs simulated data from posterior predictive.	Full
Credible Interval	Highest density interval (HDI) and equal-tailed interval (ETI) at 94% credibility.	Full
Fixed-Effects Meta-Analysis	Inverse-variance weighting, Cochran Q, I ² , per-study and pooled CIs, assumes homogeneous effects.	Full
Random-Effects Meta-Analysis	DerSimonian-Laird tau ² estimation, I ² , 95% prediction interval (Higgins-Thompson-Spiegelhalter), per-group weights.	Full
Egger's Test for Publication Bias	Funnel-plot asymmetry test via standardized-effect intercept regression, t-test on k-2 df.	Full
Forest Plot Data	Per-study effects with 95% CI, pooled estimate (fixed or random), I ² , and heterogeneity for visualization.	Full
Bootstrap Standard Error	Bootstrap distribution of a statistic (mean/median/std/var/quantiles) with SE, bias, and summary.	Full
Wild Bootstrap (OLS)	Wild (and wild-cluster) bootstrap standard errors, percentile CIs, and p-values for OLS coefficients, robust to heteroskedasticity and the standard tool for few-cluster inference.	Full

PROCEDURE	WHAT IT DOES	STATUS
Percentile-t Bootstrap (coefficient)	Studentized (bootstrap-t) confidence interval for a single OLS regression coefficient, built by pairs-resampling the pivot $t^* = (b^* - b) / se^*$ and inverting it via the Hall pivot-flip formula for second-order-accurate, skewness-corrected coverage, contrasted against the normal-theory $b \pm 1.96 se$ interval.	Full
Block Bootstrap (time series)	Moving-block and circular-block bootstrap that resamples overlapping blocks of consecutive observations to deliver dependence-robust standard errors and percentile confidence intervals for the mean, median, variance, or lag-1 autocorrelation of a time series.	Full
Parametric Bootstrap (OLS)	Resamples OLS errors from the fitted homoskedastic-Gaussian model ($y^* = X^* \beta_{\text{hat}} + N(0, \sigma^2)$), refits each synthetic sample, and reports per-coefficient bootstrap SEs and percentile 95% CIs that converge to the classical OLS standard errors as replications grow.	Full
Residual Bootstrap (OLS)	Model-based (residual) bootstrap that holds the design and fitted values fixed, resamples OLS residuals with replacement to build synthetic responses, refits each, and reports per-coefficient bootstrap standard errors and percentile 95% confidence intervals without assuming any error distribution.	Full
Subsampling Inference	Politis-Romano-Wolf subsampling: draws subsamples of size $b < n$ without replacement, recomputes a statistic on each, and uses the $\sqrt{b}$ -rescaled subsampling distribution to build a confidence interval and standard error that are stable as b varies.	Full
Bayesian Bootstrap (Rubin)	Rubin's Bayesian bootstrap for OLS coefficients: draws Dirichlet(1,...,1) weights over the observations and refits weighted least squares each replication to produce a posterior distribution (mean, sd, 95% credible interval) for every coefficient, first-order equivalent to the ordinary bootstrap.	Full
BCa Bootstrap Interval	Bias-corrected and accelerated (BCa) bootstrap 95% confidence interval for a chosen scalar statistic (mean, median, std, variance, quartiles, IQR) of one numeric variable, correcting the plain percentile interval for median bias (z_0) and skewness (jackknife acceleration a).	Full
Cluster Bootstrap (OLS)	Pairs cluster bootstrap for OLS coefficient inference that resamples whole clusters with replacement and refits to produce cluster-robust bootstrap standard errors and percentile confidence intervals valid under arbitrary within-cluster error correlation.	Full
Bootstrap Confidence Interval	Four CI methods (percentile, basic, BCa, studentized) with distribution summary and downsampled samples.	Full
Bootstrap Regression	Row-level resampling with OLS per iteration; bootstrap SE, bias, CI, and comparison to classical estimates.	Full
Jackknife Estimation	Leave-one-out resampling, SE, bias correction, and bias-corrected estimate.	Full
Permutation Test (Two-Sample)	Exact resampling test for mean/median/t-stat difference with null distribution and two-tailed p-value.	Full
Permutation Test (Correlation)	Exact test for Pearson/Spearman/Kendall correlation with null distribution and significance.	Full

PROCEDURE	WHAT IT DOES	STATUS
Cross-Validation	K-fold CV (sklearn-compatible) for linear/logistic/ridge/lasso with multiple scoring metrics.	Full
Bayesian Bootstrap	Dirichlet-weighted resampling for posterior distribution of a statistic; Bayesian alternative to classical bootstrap.	Full

Machine Learning & Choice Models 31

Statios provides machine learning and choice modeling tools: supervised learning with explainability (Random Forest, Gradient Boosting, SVM, Decision Trees, neural networks); model evaluation (cross-validation, ROC/AUC, confusion matrix, feature importance, SHAP); and conjoint analysis (traditional, choice-based, hierarchical Bayes) for preference modeling.

PROCEDURE	WHAT IT DOES	STATUS
Random Forest	Fit random forest for classification or regression with feature importances and OOB score.	Full
Gradient Boosting	Fit gradient boosted model (XGBoost if available) with learning curves and feature importances.	Full
Support Vector Machine	Fit SVM for classification or regression with standardized features and support vector counts.	Full
Decision Tree	Fit decision tree with tree rules, feature importances, and depth/leaf metrics.	Full
Cross-Validation	K-fold cross-validation with train/test scores and timing metrics across folds.	Full
ROC/AUC Analysis	Binary classification ROC curve with AUC and Youden's J optimal threshold.	Full
Confusion Matrix	Multi-class confusion matrix with per-class precision/recall/F1, normalizations, and macro/weighted averages.	Full
Classification Metrics	Accuracy, precision, recall, F1, Matthews correlation, Cohen's kappa, and log-loss.	Full
Feature Importance	Built-in and permutation-based feature importances with standard errors.	Full
SHAP Values	SHAP values via TreeExplainer (exact TreeSHAP) or KernelExplainer (model-agnostic) with expected values.	Limited
SHAP Summary Plot	SHAP summary (beeswarm) plot data with feature importance ranking and per-sample contributions.	Full
Train/Test Predict	Fit model on train set and return test predictions, actuals, residuals/probabilities.	Full
MLP Classifier	Scikit-learn multilayer perceptron for classification with configurable hidden layers and activation.	Full
MLP Regressor	Scikit-learn multilayer perceptron for regression with configurable hidden layers and activation.	Full
Deep Neural Network	Keras/TensorFlow deep network for classification or regression with custom architecture, epochs, batch size.	Full
Autoencoder	Unsupervised autoencoder for dimensionality reduction and feature learning.	Full
CNN for Tabular Data	Convolutional neural network adapted for tabular classification or regression.	Full
Neural Network Hyperparameter Tuning	Grid search over neural network hyperparameters with cross-validation.	Full
CHAID	Chi-squared automatic interaction detection for tree-based classification.	Limited
QUEST	Quick, unbiased, efficient statistical tree method for classification.	Limited
C5.0	C5.0 decision tree and rule learner for classification.	Limited
Ensemble Trees Comparison	Side-by-side comparison of multiple tree-based ensemble methods.	Limited

PROCEDURE	WHAT IT DOES	STATUS
Tree Export	Export trained tree model in formats suitable for deployment.	Limited
Compare Trees	Comparative metrics across tree models (accuracy, depth, interpretability).	Limited
Traditional Conjoint Analysis	Rating-based conjoint via OLS with dummy-coded attributes and part-worths.	Full
Choice-Based Conjoint	Aggregate multinomial logit conjoint from choice data with part-worths and hit rate.	Full
Hierarchical Bayes CBC	Empirical-Bayes iterated-MAP CBC with individual part-worths (not true HB-MCMC).	Limited
Market Simulator	Market share simulation given part-worths and scenarios (logit, first-choice, randomized).	Full
Attribute Importance	Relative attribute importance from part-worth ranges (% allocation).	Full
Conjoint Design	Fractional experimental design generation (full, orthogonal, D-optimal).	Full
Conjoint Segmentation	Segment respondents using K-means on individual part-worths.	Full

Spatial, Complex Survey & Missing Data 50

Statios provides comprehensive spatial econometrics, complex survey design analysis, and missing data handling through libpysal/spreg/esda backends. Covers spatial weights, autocorrelation tests (Moran's I, Geary's C, Getis-Ord), LISA clustering, spatial regression (SAR, SEM, SDM, SAC, GWR, MGWR), and advanced imputation (MICE, EM, PMM) with design-based inference.

PROCEDURE	WHAT IT DOES	STATUS
Load spatial geodata	Load shapefile, GeoJSON, or GeoPackage; auto-reproject to WGS84 and extract numeric columns.	Full
Create spatial weights matrix	Queen/Rook contiguity, KNN, or distance-band weights; auto-fallback to KNN if islands detected.	Full
Global Moran's I test	Spatial autocorrelation test with normality, randomization, and permutation p-values disclosed.	Full
Local Moran's I (LISA)	Spatial clustering analysis with per-location p-values; Benjamini-Hochberg FDR correction optional.	Full
Geary's C test	Spatial autocorrelation test (emphasizes local differences); two-tailed normality/randomization/permutation p-values.	Full
OLS with spatial diagnostics	OLS regression plus LM Lag, LM Error, robust LM tests with model recommendation.	Full
Spatial Lag Model (SAR)	Spatial autoregressive model via maximum likelihood; reports rho, coefficients, and fit indices.	Full
Spatial Error Model (SEM)	Spatial error autoregressive model; reports lambda parameter and standard ML fit statistics.	Full
Getis-Ord Gi* hot-spot analysis	Per-location hot/cold spot classification (Gi* with self-inclusion); uses permutation p-values.	Full
Spatial Durbin Model (SDM)	Lag model with spatially-lagged predictors; includes LR test vs OLS baseline.	Full
SAC / SARAR model	Combined spatial lag + error model via GMM; estimates rho and lambda jointly.	Full
Geographically Weighted Regression (GWR)	Local regression with adaptive/fixed bandwidth kernel; auto-selects via cross-validation.	Full
Multiscale GWR (MGWR)	GWR with per-predictor bandwidth selection; standardizes variables prior to fit.	Limited
Bivariate LISA	Local spatial correlation between X and neighbor-average of Y; quadrant classification.	Full
Local Geary's C	Local test for both positive and negative spatial association; detects clustering vs dissimilarity.	Full
SKATER clustering	Spatially-constrained clustering via MST pruning; manual fallback if spopt unavailable.	Limited
Max-p regions	Form maximum contiguous regions meeting threshold on a variable; spopt-dependent.	Limited
Rate smoothing	Spatial smoothing of crude rates (event/base) via empirical/spatial Bayes.	Full
Spatial filter	Getis spatial filter decomposing variable into spatial and non-spatial components.	Full

PROCEDURE	WHAT IT DOES	STATUS
Moran bivariate scatter	Data for Moran scatter plot (z-scores, lag, quadrants); includes global I and p-value.	Full
Survey design definition	Specify strata, clusters, sampling weights, and finite population correction; summarize design.	Full
Survey-adjusted means	Design-based estimates with Taylor linearization variance; includes DEFF per variable.	Full
Survey-adjusted totals	Weighted population totals with design-adjusted standard errors and DEFFs.	Full
Survey proportions	Design-based category proportions with confidence intervals and design effects.	Full
Survey-weighted regression	Linear or logistic regression via WLS with design-adjusted sandwich covariance.	Full
Design Effect (DEFF)	Ratio of design variance to SRS variance for a mean; includes interpretation guide.	Full
Replicate weights analysis	Jackknife, BRR, or bootstrap SE/CI from replicate weight methodology.	Full
Survey t-test	Two-sample design-adjusted t-test on difference of weighted means.	Full
Rao-Scott chi-square	Chi-square test for categorical association; first-order design effect correction.	Full
Missing value summary	Comprehensive report of missing counts, patterns, and complete-case counts per variable.	Full
Little's MCAR test	Test whether missingness is completely random; EM-based ML with chi-square inference.	Full
Simple imputation	Mean, median, or mode imputation for a single variable.	Full
KNN imputation	Impute missing via k nearest neighbors; standardizes features to avoid scale-dependence.	Full
MICE imputation	Multiple imputation by chained equations; pooled estimates via Rubin's rules.	Full
EM imputation	EM algorithm for missing data assuming multivariate normality; returns imputed dataset.	Full
Predictive Mean Matching (PMM)	Imputation via OLS prediction + nearest-neighbor donor draw; preserves observed distribution.	Full
Missing data patterns	Identify missing patterns; check monotone vs non-monotone structure; Little's MCAR suggestion.	Full
MNAR sensitivity analysis	Shift imputed values by delta range; track regression coefficient changes to find tipping point.	Full
Cartogram	Distort polygon areas proportional to values; supports Dorling, diffusion, non-contiguous methods.	Full
Conditional map	Small-multiples choropleth panels conditioned on quantile bins of another variable.	Full

PROCEDURE	WHAT IT DOES	STATUS
Differential LISA	LISA on change (t2-t1) with dynamic regime classification (emerging, persistent, declining).	Full
Co-location map	Detect spatial co-occurrence of two categorical variables; permutation-based significance.	Full
Space-time cube	3D space-time voxel hot-spot analysis; ESRI-style trajectory classification.	Full
Bivariate choropleth	Two-variable simultaneous visualization via n×n class palette (canonical 3×3 or interpolated).	Full
Rate map	Visualize crude, empirical Bayes, spatial Bayes, or SMR rates with classification.	Full
Standardized Moran scatter	Interactive Moran scatter with z-scores, quadrants, global I, and per-point leverage.	Full
Local G* map	Getis-Ord G* with adaptive distance-band optimization; classification at confidence levels.	Full
Spatial animation	Frame-by-frame data for animated choropleth/LISA showing quintiles and cluster trajectories.	Full
Conditional scatter	Small-multiples scatter plot partitioned by quantile bins; OLS line and correlation per panel.	Full
Enhanced Moran scatter	Moran scatter with regression line, leverage, residuals, and outlier flags ($ z > 2$ or leverage $> 3 \times \text{mean}$).	Full

Data Management, Transforms & Workflow 68

Data management, transforms and workflow procedures in Statios: dataset loading and multi-dataset workflows, direct cell editing with undo/redo, spreadsheet-style formulas for derived variables, variable metadata configuration (types, labels, measurement levels), data transformations (compute, recode, categorize, standardize, filter, lag/lead, dummy), data manipulation (merge, reshape, aggregate, collapse, append, sort, subset, deduplicate, rank, transpose, split/concat columns, pivot), data-to-table formatting, and session/macro recording for reproducible workflows.

PROCEDURE	WHAT IT DOES	STATUS
Load Dataset	Load data from file (CSV, TSV, XLSX, SPSS .sav, Stata .dta, JSON) into active workspace.	Full
Get Dataset Info	Retrieve metadata for active dataset: row/column counts, dtypes, missing cells, columnar status.	Full
Get Data Window	Fetch windowed rows (e.g., 100 rows at offset 0) for grid rendering.	Full
Get Full Columns	Retrieve whole columns for chart building; large datasets sampled with fixed seed for row alignment.	Full
Get Variables	Inferred variable metadata (type, measure, labels, value labels) merged with engine-owned overrides.	Full
List Datasets	List all open datasets and current active dataset ID.	Full
Switch Active Dataset	Set a dataset as active (default target for all analyses and edits).	Full
Close Dataset	Remove a dataset from memory; clears associated undo/redo stacks.	Full
Duplicate Dataset	Create independent copy of dataset with all metadata and current state.	Full
Rename Dataset	Change display name (in-memory; original file unchanged).	Full
Export Dataset	Write active dataset to file (CSV, XLSX, SAV, DTA, JSON) with variable metadata.	Full
Set Cell Value	Edit single cell; coerces value to column dtype; generates undo entry.	Full
Undo Cell Edit	Reverse last cell edit (stack cap 200 per dataset).	Full
Redo Cell Edit	Restore undone cell edit.	Full
Insert Row	Insert blank row at index; structural edit clears undo stacks.	Full
Delete Row	Remove row at index; structural edit clears undo stacks.	Full
Add Variable	Append new column (all-NaN or default name auto-generated).	Full
Delete Variable	Remove column; structural edit clears undo stacks.	Full
Sort Dataset	Sort by column(s) ascending/descending; structural edit clears undo stacks.	Full
Evaluate Formula	One-off scalar formula (e.g., =A1+B1); returns value or Excel-style error code.	Full
Apply Formula to Column	Fill formula down entire column as derived variable; stored for recomputation.	Full
Recompute Derived Columns	Recalculate all derived columns after data edits in dependency order; reports #REF! for missing inputs.	Full
Get Formulas	Retrieve stored formula metadata: source, refs (letter→column), anchor row, dependencies.	Full
Set Variable Type	Convert column dtype (numeric/string/date/boolean) with honest failure accounting.	Full
Set Measurement Level	Assign measure (nominal/ordinal/scale) for SPSS Variable View parity.	Full

PROCEDURE	WHAT IT DOES	STATUS
Set Value Labels	Map raw values to display labels (e.g., 1→'Yes', 0→'No'); empty list clears labels.	Full
Auto-Encode Categorical	SPSS-style: text categories → numeric codes with inverse map as value labels.	Full
Compute Variable	Create new variable from AST-safe expression (log, sqrt, abs, sin, cos, mean, std, + −*/). Stores result in new column.	Full
Recode Variable	Map values by pattern (e.g., '1-30'→'low', '31+'→'high') into new column.	Full
Standardize Variable	Transform to z-scores (mean 0, SD 1), min-max [0,1], or robust (median, IQR).	Full
Categorize / Bin	Convert numeric to categories: equal-width, quantile-based, or custom breaks.	Full
Filter Cases (Read-Only)	Report which rows match a condition (e.g., age > 30); read-only—use data.subset to drop.	Full
Create Dummy Variables	Expand categorical to 0/1 indicators; optionally drop first to avoid collinearity.	Full
Lag / Lead	Shift values k periods forward/backward for time-series predictors.	Full
Merge Datasets	SQL-like join (inner/left/right/outer/cross) on key column(s); reports match/unmatch counts.	Full
Reshape Long	Convert wide → long: multiple value columns become one 'value' + one 'variable' column.	Full
Reshape Wide	Convert long → wide: spread key column values across columns.	Full
Aggregate Data	Group by column(s) and compute summaries (mean/sum/count/etc.); produces new dataset.	Full
Collapse Data	Similar to aggregate; summarize grouped data (default stats: mean, SD, N).	Full
Append Datasets	Vertically stack 2+ datasets (row-bind); common columns aligned.	Full
Sort Dataset	Sort by column(s) ascending/descending; optionally replace active or create new dataset.	Full
Subset / Filter	Keep rows matching condition (AST-safe); replace active dataset.	Full
Deduplicate	Remove duplicate rows (keep first/last/none); optionally subset columns first.	Full
Transpose Dataset	Flip rows ↔ columns; optionally use one column as new row index.	Full
Split Column	Split text column by separator into multiple new columns.	Full
Concatenate Columns	Join 2+ columns into one (e.g., first + last → fullname).	Full
Rank Variable	Convert to ranks (1,2,3...); methods: average (ties split), min/max, dense, ordinal.	Full
Crosstab Counts	Build contingency table (row × col) from two categorical variables; include marginals.	Full
Pivot Longer (Complex)	Advanced wide→long: regex on column names extracts multiple groups into separate columns.	Full

PROCEDURE	WHAT IT DOES	STATUS
Save Session	Persist all datasets, variable metadata, and output history to .stios JSON.	Full
Load Session	Restore datasets, metadata, and outputs from .stios; repopulates DatasetManager.	Full
Session Info	Read session metadata (creation date, version, dataset names, row/col counts).	Full
Session Diff	Compare two .stios files; report differences in datasets and outputs.	Full
Start Macro Recording	Begin recording user actions (transforms, analyses) for later replay.	Full
Stop Macro Recording	End current macro recording session.	Full
Save Macro	Write recorded actions to .sfmacro JSON file.	Full
Load Macro	Read .sfmacro file; return actions ready to replay.	Full
Replay Macro	Execute recorded actions in sequence (transform.compute, data.*, etc.) on current dataset.	Full
Session Undo	Reverse a logged action by ID (currently stub; per-dataset cell undo used instead).	Limited
Session Redo	Restore undone action by ID (currently stub; per-dataset cell redo used instead).	Limited
Export Script	Convert recorded macro actions to Python or R script (reproducible workflow).	Full
Regression Table	Format regression results (multiple model specs) in APA/HTML/LaTeX table.	Full
Descriptives Table	Format descriptive statistics by groups in APA/HTML/LaTeX table.	Full
Correlation Table	Format correlation matrix in APA/HTML/LaTeX; supports Pearson, Spearman, Kendall.	Full
Crosstab Table	Format contingency table (observed, expected, residuals) in APA/HTML/LaTeX.	Full
Model Comparison Table	Side-by-side regression models with fit statistics and significance stars.	Full
ANOVA Table	Format one-way/repeated-measures ANOVA table in APA/HTML/LaTeX.	Full
Factor Loading Table	Format EFA/PCA loadings with rotation (varimax, promax, etc.).	Full

AI, Automation, Graphics & Reporting 21

AI-powered statistical guidance, automated analysis pipelines, publication-quality graphics and report generation, plus a visual rule-based workflow recommender (Guided Path) and a node-based pipeline editor (RLab).

PROCEDURE	WHAT IT DOES	STATUS
Statios AI assistant	Ask the built-in assistant in plain language; it chooses an analysis, runs it on the local engine, and interprets the result. Routed through the Statios relay — your raw data rows never leave your device.	Full
Summarize iteration	Summarizes executed analyses into key findings as structured JSON (used by the Research Path); routed through the Statios relay.	Full
Auto data quality check	Single-call data quality report: missing patterns, duplicates, type consistency, cardinality anomalies.	Full
Auto assumption validation	Detect violations of key statistical assumptions (normality, homogeneity, linearity, etc.) for a given test type.	Full
Auto exploratory data analysis	One-command EDA: univariate summaries, correlation matrix, key visualizations, and insight extraction.	Full
Compare candidate models	Fit multiple regression models (OLS, ridge, lasso, elastic net, robust) on same outcome; rank by IC/CV.	Full
Histogram	Distribution visualization for a numeric variable with Sturges' bin count and descriptive stats.	Full
Scatter plot	Bivariate X–Y plot with optional grouping and Pearson correlation in response.	Full
Box plot	Median, quartiles, and outliers per group (or ungrouped); includes min/max and sample size.	Full
Bar chart	Categorical value counts or aggregated numeric summaries, with optional value labels.	Full
Dashboard slicer metadata	Returns filter domain for a variable: categorical value list or numeric min/max for dashboard UI.	Full
Dashboard tile data	Executes a dashboard tile query (chart or KPI) with optional filters; sampled for large datasets.	Full
Export statistical table	Write a result table to DOCX (APA style), LaTeX, HTML, Markdown, or CSV.	Full
Generate full report	Build a report document from analysis sections with optional citations; export as Markdown, HTML, or LaTeX.	Full
Build report document	Assemble a deterministic report model from journal entries, dataset metadata, and optional guided plan; handles APA 7 formatting and citation aggregation.	Full
Export report to file	Render and write a report model as HTML, Markdown, or BibTeX; optionally saves .bib sidecar.	Full
Recommend analysis path	Profile-based deterministic method selector: returns step plan with matched rule, alternatives, and aggregated bibliography.	Full
Generate executable plan DAG	Convert study profile + matched rule into a directed graph with nodes (analyses), edges (unconditional backbone + conditional branches), and bound parameters.	Full
Profile dataset	Introspect active dataset and return partial StudyProfile (data-determined fields only); leaves interview unknowns and variable candidates for UI.	Full

PROCEDURE	WHAT IT DOES	STATUS
AI-guided interview turn	Constrained AI conversation that fills StudyProfile fields only via strict-JSON validation; forbids methodological advice.	Full
Generate methods section	Render a plan DAG as academic methods prose (intro, phase paragraphs with inline citations, alphabetical APA list, BibTeX export).	Full

Statios — Capability Statement. This document is generated from the software's own engine inventory and reflects the current build. Figures and procedures are subject to change as the product evolves.