PROGRAMMA | 17 SETTEMBRE

8:30-9:00 Registrazione dei partecipanti

9:00-9:45 SESSIONE I • INVITED SPEAKER

AI Agents for Research Workers • Scott Cunningham, Baylor University

In this year's invited session, Scott Cunningham turns to a development few applied researchers can now choose to ignore: the arrival of AI coding agents in the empirical workflow. Data cleaning, model estimation, visualization and replication are all tasks increasingly open to delegation. In this session, Scott discusses where these tools genuinely help, where they mislead, and what their use implies for the reliability and reproducibility of published research.

9:45-11:05 SESSIONE II • COMMUNITY CONTRIBUTED COMMANDS, I

fffuroot • **Giovanni Bruno**, Università Commerciale Luigi Bocconi di Milano e Chiara Oldani, Univerità degli Studi della Tuscia

Implementing in *Stata* unit-root and stationarity tests with smooth breaks approximated by flexible Fourier forms. I describe the *Stata* implementation of unit-root and stationarity tests with flexible Fourier forms as in Enders and Lee (2012a), (2012b) and Becker, Enders and Lee (2006).

xtthreshold: Panel Data Threshold Regression with Interactive Fixed Effects • **Jan Ditzen**, Libera Università di Bolzano; Yiannis Karavias, Brunel University of London e Joakim Westerlund, Lund University

Threshold regression provides a flexible framework for capturing regime - dependent relationships, and has become a widely used tool for uncovering structural breaks, nonlinearities, and state-dependent patterns in data across diverse fields. This presentation introduces a new community contributed command called **xtthreshold**, which provides researchers with a complete toolbox for analysing threshold regression in panel data with interactive fixed effects. The new command is able to estimate both discontinuous threshold regression and kink regression, both with interactive fixed effects. It further accommodates three alternative specifications of slope heterogeneity: fully homogeneous, fully heterogeneous, and semi-homogeneous models. Model selection is facilitated by a modified information criterion capable of discriminating among competing specifications, and inference is supported through confidence intervals and hypothesis testing for all model parameters, in a addition to bootstrap-based tests for the existence of non-linearity.

Earning While Learning: How to Run Batched Bandit Experiments • **Davud Rostam-Afschar** e Jan Kemper, University of Mannheim

Researchers typically collect experimental data sequentially, allowing early outcome observations and adaptive treatment assignment to reduce exposure to inferior treatments. This article reviews multi-armed-bandit adaptive experimental designs that balance exploration and exploitation. Because adaptively collected experimental data through bandit algorithms violate standard asymptotics, inference is challenging. We implement an estimator that yields valid heteroskedasticity-robust confidence intervals in batched bandit designs and compare coverage in Monte Carlo simulations. We introduce **bbandits** for *Stata*, a tool for designing experiments via simulation, running interactive bandit experiments, and implementing and analyzing adaptively collected data. **bbandits** includes three common assignment algorithms—ε-first, ε-greedy, and Thompson sampling—and supports estimation, inference, and visualization.

rdlasso: Regression Discontinuity with High-Dimensional Data • **Marianna Nitti** e Marco Ventura, Università degli Studi di Roma La Sapienza

We present a command, **rdlasso**, which enables the inclusion of highdimensional covariates in Regression Discontinuity Design (RDD) settings. This command is based on the paper "Inference in Regression Discontinuity Designs with High-Dimensional Covariates" by Kreiss and Rothe (2023). The command automates covariate selection using Lasso-based procedures, supports both sharp and fuzzy settings and integrates seamlessly with **rdrobust** for bandwidth selection and inference. The command relies on *Stata*'s native implementation of lasso for high-dimensional covariate selection and on **rdrobust** for bandwidth selection, estimation, and inference, making the methodology both accessible to *Stata* users and computationally feasible for applied researchers.

11:05-11:20 Pausa caffè

11:20-12:45 SESSIONE III • EXPLOITING THE POTENTIAL OF STATA 19, I

Generalized structural equation models: GLMs, multilevel, latent classes, and more • Kristin MacDonald, *StataCorp*

Stata's **gsem** command fits generalized structural equation models. This generalization allows user to fit structural equation models with continuous, binary, ordinal, categorical, count, and survival-time outcomes. Users can also fit models with continuous or categorical latent variables. Multilevel models are also supported. In this presentation, it will be demonstrated a wide variety of models that can be fit with **gsem**, from multilevel confirmatory factor models to latent class models. Along the way, participants will see both typical SEM applications and unique models that can be fit in the generalized SEM framework.

Beyond Evidence Synthesis: A Meta-Analytic Framework for Explaining Heterogeneous Dynamic Parameters • Maria Elena Bontempi, Università degli Studi di Bologna

Meta-analysis is traditionally used to synthesize evidence across independent studies. This paper illustrates a non-standard application of *Stata*'s multivariate meta-analysis framework to investigate parameter heterogeneity estimated from firm-level dynamic models. The empirical motivation comes from corporate finance, where heterogeneous adjustment dynamics make pooled panel specifications with numerous interactions difficult to interpret.

We first estimate fully heterogeneous error-correction models for individual firms, obtaining firm-specific parameters measuring the speed of leverage adjustment, the sensitivity to free cash flow, and debt-maturity interaction, together with their estimated standard errors. Rather than treating these parameters as final estimates, we use *Stata*'s multivariate random-effects meta-regression to explain their cross-sectional heterogeneity through firm characteristics, contractual features, and institutional changes.

The paper demonstrates how *Stata*'s **meta** commands can be extended beyond their conventional role of evidence synthesis to provide a flexible second-stage modelling framework for heterogeneous parameter estimates. The approach accommodates multiple correlated outcomes, accounts for estimation uncertainty through inverse-variance weighting, and avoids the overparameterization that often arises in pooled interaction models. Although illustrated using a novel dataset on corporate debt covenants extracted from SEC filings, the methodology is applicable to any context in which unit-specific parameters are estimated in a first stage and subsequently related to observed characteristics.

Bibliometric Analysis in Stata: Text Mining, Network Community Detection, and Unsupervised Learning • **Carlo Drago**, Università degli Studi Niccolò Cusano e Gentian Hoxhalli, Luarasi University & Armed Forces Academy

This study presents a bibliometric analysis workflow that combines *Stata* with its integrated Python environment to map the thematic structure of a large OpenAlex corpus and a dataset of abstracts in Scopus on the theme of energy policy. Relevant keywords are extracted from article titles through transparent linguistic preprocessing, domain-specific filtering, and frequency-based selection. The resulting article-by-keyword matrix is analyzed through principal component analysis to reduce dimensionality and identify the main latent thematic dimensions of the literature. K-means clustering is then applied to the retained component scores, assigning each article to a homogeneous research cluster. *Stata* manages data preparation, descriptive statistics, graphical outputs, and result export, while Python provides scalable text mining and machine learning procedures. The workflow produces an interpretable keyword dictionary, PCA loadings and scores, cluster memberships for individual articles, and cluster-level bibliometric profiles. A different approach based on the analysis of abstracts classifies them using cosine distance in an unsupervised framework. Finally, the co-occurrence matrix is analyzed using network analysis to identify the most relevant "cores" in the literature. The contribution is a practical workflow for conducting transparent, computationally efficient bibliometric analyses entirely within a *Stata*-centered research environment.

12:45-13:10 SESSIONE IV - *STATA* TIPS AND TRICKS

Building Desktop Applications for *Stata* Using Python • Giovanni Cerulli, Consiglio Nazionale delle Ricerche, CNR-IRCrES

Stata is one of the most widely used environments for statistical analysis, but many workflows still require users to interact directly with the command line or do-files. This presentation illustrates how Python can be used to build lightweight desktop graphical user interfaces (GUIs) that interact seamlessly with *Stata*, allowing users to launch complex analyses through intuitive point-and-click applications. The presentation will discuss the architecture of the communication between Python and *Stata*, practical implementation details, and possible extensions.

adoadd • Jan Ditzen, Libera Università di Bolzano

adoadd is a package developed to control *Stata*'s ado environment.

13:10-14:15 Pranzo

14:15-15:15 SESSIONE V - COMMUNITY CONTRIBUTED COMMANDS, II

wqsreg - a *Stata* command for Weighted Quantile Sum regression • **Marta Ponzano**, Università degli Studi Link di Roma; Stefano Renzetti, Università degli Studi di Parma e Andrea Bellavia, T.H. Chan School of Public Health, Harvard Medical School

Weighted Quantile Sum (WQS) regression is a statistical method for quantifying the association between a set of possibly correlated predictors and a health outcome, estimating the joint effect of the predictors as well as their individual contributions to the total effect. We present **wqsreg**, the first *Stata* command for WQS regression, implemented for continuous, binary and count outcomes. The execution of the command involves two sequential steps: 1) estimating the weights and constructing the WQS index under specific constraints; 2) modeling its association with the outcome. **wqsreg** integrates several flexible components of the framework such as bootstrap, training/validation, and repeated holdout procedures; it returns regression estimates as well as graphical displays of the individual weights. We present an application of the command on exposome data exploring the association between 38 exposures and a continuous outcome while adjusting for a set of covariates. To the best of our knowledge, **wqsreg** provides the first command to conduct WQS regression in *Stata*. We anticipate that our contribution will further promote the use of appropriate statistical methods for handling multiple correlated predictors.

Assessing the functional form of the Cox model: the **stfform** command • **Daniele Spinelli** e Rino Bellocco, Università degli Studi di Milano-Bicocca

We introduce **stfform**, a post-estimation command for **stcox** that allows users to assess the functional form of continuous covariates in Cox proportional hazards regression models using cumulative sums of martingale residuals. By complementing graphical diagnostics *with a* formal statistical test, **stfform** provides a principled assessment of the adequacy of the specified functional form.

catmetrics: computing 300 measures of association, similarity and forecast evaluation for categorical data • **Andrei Sirchenko**, Nyenrode Business University; Dragoş Bînzari e Konrad Wrębiak, University of Amsterdam

This article introduces **catmetrics**, a new *Stata* command for computing various measures of association, similarity, and deterministic and probabilistic forecast evaluation for categorical data. The command accommodates binary, nominal, and ordinal multicategory data. It reports over 300 distinct measures, providing their alternative names across various disciplines and their attainable ranges. Furthermore, in addition to overall metrics, **catmetrics** calculates class-specific quantities for each category, as well as their macro and weighted averages.

15:15-16:40 SESSIONE VI - EXPLOITING THE POTENTIAL OF *STATA* 19, II

Conditional Average Treatment-effects estimation using *Stata* • Di Liu, *StataCorp*

Treatment effects estimate the causal effects of a treatment on an outcome. The effect may be heterogeneous. Average treatment effects conditional on a set of variables (CATEs) help us understand heterogeneous treatment effects. By construction, they are useful to evaluate how different treatment-assignment policies affect different groups in the population.

In this talk, we will show how to use *Stata* 19's new command **cate** to answer questions such as the following:

1. Are the treatment effects heterogeneous?
2. How do the treatment effects vary with some variables?
3. Do the treatment effects vary across prespecified groups?
4. Are there unknown groups in the data for which treatment effects differ?
5. Which is best among possible treatment-assignment rules?

Optimal Policy Learning under Constraints in *Stata* • Giovanni Cerulli, Consiglio Nazionale delle Ricerche, CNR-IRCrES

Optimal Policy Learning (OPL) has emerged as a powerful framework for designing data-driven treatment allocation rules that maximize social welfare while accounting for treatment effect heterogeneity. In many real-world policy applications, however, decision makers face operational constraints such as limited budgets, minimum coverage requirements, or combinations of both, making unconstrained policy learning impractical. This presentation introduces a comprehensive *Stata* implementation for constrained Optimal Policy Learning that enables researchers and practitioners to estimate, evaluate, and compare treatment assignment policies under realistic resource limitations. The framework integrates modern causal machine learning methods for estimating Conditional Average Treatment Effects (CATEs) with efficient optimization algorithms that solve budget-constrained, coverage-constrained, and joint constrained allocation problems. We discuss the underlying theoretical framework, the computational implementation, and the associated *Stata* commands, illustrating their use through empirical examples and simulation evidence. The proposed tools provide an accessible and reproducible environment for translating heterogeneous treatment effect estimates into implementable policy recommendations, thereby bridging the gap between causal inference, machine learning, and evidence-based policy design within the *Stata* ecosystem.

How to deal with confounders in survival curves; a short (useful) review • Rino Bellocco, Università degli Studi di Milano-Bicocca

In observational studies, where the end-point is time to event, often we must deal with the issue of confounding which can potentially bias the association between the exposure and the outcome of interest. In my presentation, I will present an overview of the methods proposed for controlling for confounders, from regular simple *Stata* options to more advanced procedures.

16:40-16:55 Pausa caffè

16:55-17:25 SESSIONE VII - EXPLOITING THE POTENTIAL OF *STATA* 19, III

Stress Testing Guaranteed Minimum Income Schemes: A Counterfactual Microsimulation Framework Using EU-SILC • Ewa Aksman, University of Warsaw

This paper develops a counterfactual microsimulation framework for evaluating the fiscal consequences of Guaranteed Minimum Income (GMI) schemes under adverse labour supply responses. The framework combines a theoretical cost-effectiveness measure based on the Lorenz curve and the Gini coefficient with an empirical implementation using EU-SILC microdata in *Stata*. It enables a systematic comparison of alternative GMI architectures while quantifying their fiscal implications under identical behavioural assumptions.

The analysis considers an extreme counterfactual scenario in which all households eligible for GMI—defined as those with disposable incomes below 60% of the relevant median disposable income—withdraw completely from the labour market. The scenario is intended as a fiscal stress test rather than a behavioural forecast. Two institutional arrangements are compared: (i) decentralized national GMI schemes, where eligibility is defined relative to national median income, and (ii) a centralized EU-wide GMI system, where eligibility is defined relative to the EU-wide median income.

The proposed framework is implemented in *Stata* through a reproducible computational workflow that automates the identification of eligible households, construction of counterfactual income distributions, computation of population-weighted cost-effectiveness indicators, and decomposition of the cost-effectiveness ratio into contributions stemming from between-country and within-country redistributive effects. The implementation provides a flexible computational environment for evaluating alternative GMI designs under different eligibility thresholds and behavioural assumptions and can be readily adapted to other comparative redistribution analyses based on EU-SILC microdata.

The empirical application shows that complete labour market withdrawal increases total GMI expenditures by 89% under national schemes and by 104% under a centralized EU-wide scheme relative to baseline estimates. Although beneficiary coverage remains unchanged and the redistributive impact of both systems is preserved, fiscal efficiency deteriorates substantially. The proposed framework illustrates how *Stata* can be used to integrate theoretical distributional analysis with large-scale EU-SILC microsimulation, providing a transparent and reproducible approach for stress-testing alternative social protection policies.

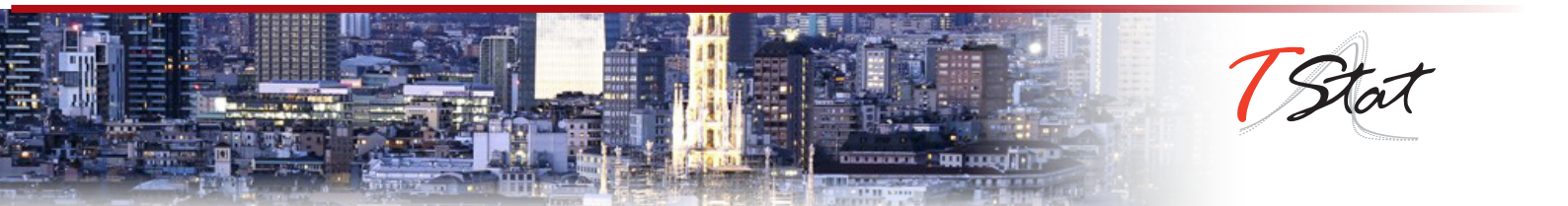
Geographic Distribution of Smoking and High-Risk Alcohol Consumption in Italy: A Quintile-Based Approach Using PASSI Surveillance Data • **Giovanni Capelli**, Federica Asta, Valentina Minardi, Benedetta Contoli e Maria Masocco, Istituto Superiore di Sanità

Geographical analysis is a powerful tool for public health surveillance because it allows the visualization of spatial patterns, territorial inequalities and population groups at greater risk. Maps provide an immediate and intuitive representation of epidemiological indicators, supporting the identification of geographic clusters and helping policy makers prioritize prevention strategies and resource allocation. Within the Italian PASSI surveillance system, referred to adult population aged 18-69 years, geographic visualization can enhance the interpretation of behavioral risk factors by highlighting regional differences that may not emerge from national averages alone.

This study presents a geographic analysis of two PASSI indicators, smoking status and high-risk alcohol consumption, using data from the 2024–2025 biennium. Regional estimates were calculated as weighted prevalences through the svy command in *Stata*, ensuring that the results account for the complex sampling design and are representative of the resident adult population. The indicators were then displayed through thematic maps at the regional level and local level realized through *Stata* maps visualization commands.

To improve the interpretability of territorial differences, prevalence estimates were classified using quintiles rather than fixed thresholds or comparisons with the national average. Quintile-based classification offers several advantages: it distributes regions more evenly across categories, enhances visual contrast, facilitates the identification of relative geographic gradients, and reduces the risk of masking meaningful variability when indicator distributions are skewed. Unlike classifications centered on a national benchmark, quintiles emphasize the relative position of each region within the overall distribution, providing a clearer picture of territorial inequalities.

The use of weighted prevalence estimates combined with quintile-based thematic mapping represents an effective approach for communicating PASSI surveillance data



and supporting evidence-based public health planning. Furthermore, this methodology can be easily extended to finer geographic levels, such as Local Health Authorities (LHAs) and municipalities, allowing the identification of local patterns and inequalities that may be hidden in regional-level analyses.

17:30-17:45 OPEN PANEL DISCUSSION WITH *STATA* DEVELOPERS • DI LIU AND KRISTIN MACDONALD, *STATACORP*

La sessione "*Open panel discussion with Stata Developers*" offre ai partecipanti la possibilità di interagire direttamente con la *StataCorp*: sarà possibile evidenziare problemi o limitazioni del software nonché suggerire eventuali miglioramenti o comandi che potrebbero essere inclusi in *Stata*.

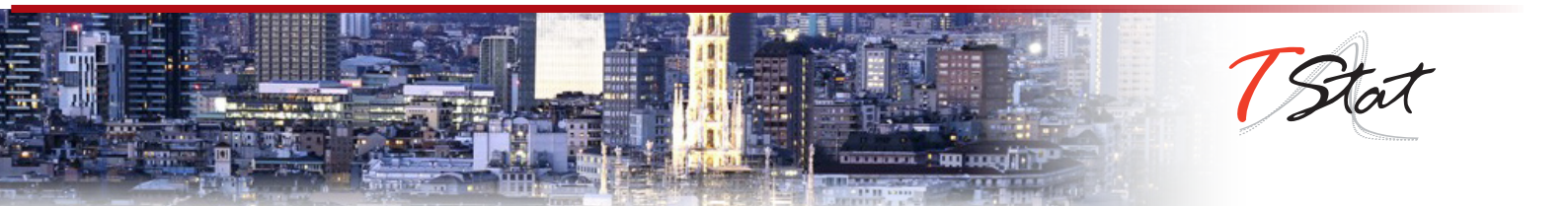
20.00 Cena Sociale *(opzionale)*

COMITATO SCIENTIFICO

Una-Louise BELL, TStat - TStat Training
Rino BELLOCCO, Università degli Studi di Milano-Bicocca
Giovanni CAPELLI, Istituto Superiore di Sanità
Giovanni CERULLI, IRCRES-CNR
Jan DITZEN, Libera Università di Bolzano
Maurizio PISATI, Università degli Studi di Milano-Bicocca

SEGRETERIA ORGANIZZATIVA

Monica GIANNI • Via Rettangolo, 12-14 • 67039 Sulmona (AQ)
T. +39 0864 210101 • www.tstat.it | www.tstattraining.eu
conference@tstat.it



TRAINING COURSE, 18TH SEPTEMBER 2026 | 9:00-17:00

MODERN DIFFERENCE-IN-DIFFERENCES: NEW PROBLEMS, NEW SOLUTIONS

COURSE OVERVIEW

Difference-in-differences has become the workhorse design of applied empirical research, yet the past decade has substantially revised what constitutes credible practice. Staggered treatment adoption, heterogeneous effects and violations of parallel trends have all been shown to compromise conventional two-way fixed effects estimation, prompting a rapid succession of new estimators and diagnostic tools.

This short course offers a structured route through this literature and its implementation in *Stata*. Beginning with the canonical 2×2 and event-study designs (parallel trends, the interpretation of ATT parameters as estimands, and inference). The course moves on to the identification problems that arise in practice: covariate adjustment and triple differences, compositional change, weak overlap, and the well-documented pathologies of differential timing. Participants are introduced to the leading modern alternatives, including the Goodman-Bacon decomposition, Callaway and Sant'Anna, Sun and Abraham, Borusyak et al. and Wooldridge, with attention to the distinction between imputation and weighting approaches. Later sessions address formal sensitivity analysis for bounding violations of parallel trends, extensions to continuous and dosage treatment frameworks, and a set of recommended checks illustrated through worked empirical examples.

COURSE LEADER

Scott Cunningham, Ben H. Williams Professor of Economics, Baylor University.

TARGET AUDIENCE

The course is designed for researchers who already use difference-in-differences and wish to bring their practice into line with current methodological standards, with an emphasis throughout on estimation, diagnostics and transparent reporting in *Stata*.

COURSE OUTLINE

SESSION 1: SIMPLE DID DESIGNS

1. The 2x2 design
2. 2xT event studies
3. Parallel trends
4. Interpreting ATT parameters as estimands
5. Clustering and standard errors

SESSION 2: VIOLATIONS OF PARALLEL TRENDS

Covariates, triple differences, compositional changes over time, addressing weak overlap.

SESSION 3: DIFFERENTIAL TIMING

Two-way fixed effects problems, Goodman-Bacon decomposition, Callaway and Sant'Anna, Sun and Abraham, Borusyak et al., Wooldridge, imputation vs. weighting estimators.

SESSION 4: SENSITIVITY ANALYSIS

Bounding parallel trends violations. Sensitivity of results across packages and languages.

SESSION 5: CONTINUOUS DID AND DOSAGE DESIGNS

Extending difference-in-differences to continuous treatments and dosage frameworks.

SESSION 6: A MODEST SUGGESTION OF CHECKS

Extending difference-in-differences to continuous treatments and dosage frameworks.

Maggiori dettagli riguardo il corso di formazione possono essere trovati online.

DATA, LUOGO E QUOTA DI ISCRIZIONE

L'edizione 2026 della Conferenza Italiana degli Utenti di *Stata* si terrà a Milano presso l'Hotel "*UNA Hotels Century Milano*", il 17-18 Settembre.

	Studenti e Dottorandi *Full-Time*	Altre Categorie
• Conferenza	€ 60.00	€ 110.00
• Conferenza / Corso di Formazione	€ 195.00	€ 420.00

1. I prezzi si intendono IVA 22% esclusa. L'aliquota IVA non sarà applicata per Enti Pubblici soggetti ad esenzione a norma dell'art. 14 c. 10 della L. 537/93 per la partecipazione a corsi di formazione dei propri dipendenti.
2. Nella quota di iscrizione sono inclusi: pause caffè, pranzo, materiale per il corso e una licenza temporanea *StataNow*™ per i partecipanti al corso di formazione.
3. Per ulteriori informazioni contattare la segreteria organizzativa, oppure inviare direttamente il modulo di registrazione entro il 10.09.2026.

TStat

www.tstat.it | www.tstat.eu