XVIII Italian Stata Users Conference
Florence, 9th-10th May 2024

PROGRAM | 9th MAY

8.30-9.00 Registration

9.00-9.45 SESSION I - INVITED SPEAKER SESSION

Structural Equation Modelling with Partial Least Squares using Stata • Sergio Venturini, Univeristà Cattolica del Sacro Cuore Cremona and Mehmet Mehmetoglu, Norwegian University of Science and Technology

Structural equation modeling (SEM) is a multivariate statistical framework that can model both observed and unobserved (latent) variables through complex relationships. In this talk we present **plssem**, a user-contributed Stata package for partial least-squares SEM, an approach to SEM that has attracted a lot of interest in the last 20 years from an increasing number of researchers and practitioners from many fields such as marketing, information systems, economics, psychology, and others. After introducing the topic to the audience, the talk will illustrate the current architecture of the package and its main features.

9.45-11.30 SESSION II - COMMUNITY CONTRIBUTED, I

Optimal Policy Learning using Stata • Giovanni Cerulli, National Research Council, IRcRES Roma

This presentation introduces the Stata package **opl** for optimal policy learning, facilitating ex-ante policy impact evaluation within the Stata environment. Despite theoretical progress, practical implementations of policy learning algorithms are still poor within popular statistical software. To address this limitation, the package implements three popular policy learning algorithms in Stata (threshold-based, linear-combination, and fixed-depth decision tree), and provides practical demonstrations of them using a real database. Also, I present a policy scenario development proposing a menu strategy, particularly useful when selection variables are affected by welfare monotonicity. Overall, the package contributes to bridging the gap between theoretical advancements and practical applications of policy learning.

Using Marginal Effects for Interpretation in Item Response Theory and in Tests of Differential Item Functioning: Introducing Stata Commands **irt_me** and **irt_dif** • Trenton D. Mize, Purdue University, Indiana

The field of categorical data analysis has largely shifted from the limitations of coefficient interpretations to the more flexible and powerful possibilities afforded by marginal effects, spurred by Stata's widespread implementation of the **margins** command. Despite item response theory being the latent variable corollary of categorical data analysis, a similar transformation in interpretation tools and practices has yet to emerge. I propose using tests of marginal effects for interpretation in item response theory models, demonstrating the advantages to this strategy over focusing on coefficients. Further, I show how to solve several issues when translating the idea of marginal effects to a latent variable model. A new command, **irt_me**, automates the estimation of marginal effects after any item response theory model (irt) in Stata, including models with binary, ordinal, nominal, and count items (or a mix).

Differential item functioning is a method of detecting item bias that has traditionally relied on tests of interaction terms in the item response theory coefficients. However, it is well-established in the categorical data analysis realm that coefficients are inappropriate for tests of interaction: tests of the equality of marginal effects are instead the recommended approach. A new command, **irt_dif**, provides tests of differential item functioning by testing the equality of marginal effects from item response theory models fit across separate groups.

10.45-11.00 Coffee break

Too much or too little? New tools for the CCE Estimator • Jan Ditzen, Free University of Bozen

This talk will cover new developments in the literature of the CCE (Common Correlated Effects) and their implementation into Stata. First, I will discuss regularized CCE. CCE is known to be sensitive to the selection of the number of cross-section averages. **rCCE** overcomes the problem by regularizing the cross-section averages. Secondly, I will discuss the test for the rank condition based on DeVos, Everaert and Sarafidis (2024, Econometrics Reviews). If the rank condition fails, CCE will be inconsistent and therefore testing the condition is key for any empirical application. Finally, I will the selection of cross-section averages using the information criteria from Karabiyik, Urbain, Westerlund (2019, Journal of Applied Econometrics) and Margaritella and Westerlund (2023, The Econometrics Journal).

11.30-12.30 SESSION III - EXPLOITING THE POTENTIAL OF STATA 18, I

Causal mediation analysis with Stata • Joerg Luedicke, Senior Social Scientist and Software Developer, StataCorp

Causal inference is an essential goal in many research areas and aims at identifying and quantifying causal effects. By decomposing causal effects into direct and indirect effects, causal mediation provides further insight into underlying mechanisms through which causal effects operate. This talk presents the basic theoretical framework for causal mediation analysis and discusses a variety of examples using Stata's **mediate** command. Examples will include linear and generalized linear models using a variety of outcome and mediator variables as well as different types of treatments.

12.30-13.30 Lunch

13.30-14.30 SESSION IV - STATA TIPS AND TRICKS

NNLS: Non-negative least squares using Stata • Giovanni Cerulli, National Research Council, IRcRES Roma

The **NNLS** command enables users to carry out "Non-Negative Least Squares" using Stata calling Python in the background. A simple application of the **NNLS** Stata command on real data will be provided.

htmltab2stata: Converting html tables into a Stata dataset • Jan Ditzen, Free University of Bozen

htmltab2stata parses html code from websites. It detects tables enclosed with the html <table>environment and transforms the table into a Stata dataset.

Implementing Groupwise - Heteroskedasticity - Robust Variance - Covariance Estimators in Fixed-Effects Panel Data Regression with Stata • Giovanni Bruno, Bocconi University, Milano

Stock and Watson (2008) prove that the plain White heteroskedasticity-robust VCE is generally inconsistent for fixed T, $N \to \infty$ in fixed-effect panel data regression. Bruno (2024) proves that the aforementioned VCE is (fixed T, $N \to \infty$) consistent under groupwise heteroskedasticity (GH), that is when the conditional variance of the idiosyncratic error is time-invariant, but can vary across individuals. As is well known, the **vce(robust)** option of **xtreg** in Stata implements the cluster-robust VCE, not the White VCE. In this paper I show simple Stata procedures to implement the White VCE and a second GH-robust VCE in fixed-effects panel data regression. Monte Carlo experiments prove that both VCEs, under GH, have good finite-sample properties, compared to the bias-adjusted VCE by Stock and Watson and the cluster-robust VCE.

14.30-16.00 SESSION V - EXPLOITING THE POTENTIAL OF STATA 18, II

Bayesian model averaging • Meghan Cain, Assistant Director of Educational Services, StataCorp

Are you unsure which predictors to include in your model? Rather than choosing one model, aggregate results across all candidate models to account for model uncertainty with Bayesian model averaging (BMA). Which predictors are important given the observed data? Which models are more plausible? How do predictors relate to each other across different models? BMA can answer these questions and many more.

Stata 18 introduced the **bma** suite of commands to perform BMA in linear regression models. In this talk, you will learn how to explore influential models, make inferences, and obtain better predictions with BMA. I will demonstrate the utility of BMA for any researcher—Bayesian, frequentist, and everyone in between! No prior knowledge of the Bayesian framework is required.

Text mining in Economics and Health Economics using Stata • Carlo Drago, University Niccolò Cusano, Roma

Within the more relevant data science topic, text mining is an important and active research area that offers various ways to extract information and insights from text data. Its continued use and improvement could drive innovation in several areas and improve our ability to interpret, evaluate, and utilize the vast amounts of unstructured text produced in the digital age. Extracting insightful information from text data through text mining in healthcare and business holds great promise. Text mining in business can provide insightful information by analyzing large amounts of text data, including research papers, news, and financial reports. It can help analyze market sentiment, identify emerging trends, and more accurately predict economic indicators by economists. For example, economists can find terms or phrases that reflect investment behavior and sentiment changes by applying text-mining methods to financial news. Text mining can provide essential insights into health economics by examining various textual data, including patient surveys, clinical trials, medical records, and health policy. Researchers and policymakers can use it to understand healthcare utilization patterns better, identify the variables that influence patient outcomes and evaluate the effectiveness of different healthcare treatments. Text mining can examine electronic health data and identify trends in disease incidence, treatment effectiveness and healthcare utilization. In this presentation I will illustrate the instruments currently available in Stata to facilitate a number of text-mining methods.

16.00-16.15 Coffee break

16.15-17.30 SESSION VI - COMMUNITY CONTRIBUTED, II

geoplot: A new command to draw maps • Ben Jann, Institute of Sociology, University of Bern

geoplot is a new command for drawing maps from shape files and other datasets. Multiple layers of elements such as regions, borders, lakes, roads, labels, and symbols can be freely combined and the look of elements (e.g. color) can be varied depending on the values of variables. Compared to previous solutions in Stata, **geoplot** provides more user convenience, more functionality, and more flexibility. In this talk I will introduce the basic components of the command and illustrate its use with examples.

Fitting spatial autoregressive logit and probit models using Stata: The **spatbinary** command • Daniele Spinelli, University of Milano-Bicocca, Milano

Spatial regressions can be estimated in Stata using the **spregress**, **spxtregress**, and **spivregress** commands. These commands allow users to fit spatial autoregressive models in cross-sectional and panel data. They are designed to estimate regressions with continuous dependent variables. The **spatbinary** command now allows Stata users to fit spatial logit and probit models, important models in applied econometrics.

17.30-17.50 SESSION VII - APPLICATION STUDY USING STATA

Do Alternative Work Arrangements Substitute Standard Employment? Evidence from Worker-Level Data • Filippo Passerini, University of Bologna and LABORatorio R. Revelli, Bologna

This study analyses the impact of an Alternative Work Arrangement (AWA) called "voucher" on earnings of atypical workers and on their alternative income sources using Italian administrative data. Specifically, we investigate whether this form of very flexible work substitutes income from more standard labor contracts and welfare transfers related to employment insurance (sick and parental leave and unemployment benefits). We estimate cross-income elasticities using fixed effects and diff-in-diff specifications that correct for sample selection of individuals in the labor market. Results show that vouchers increase overall labor income, but they also substitute earnings derived from other labor contracts. We do not find relevant associations between vouchers and welfare transfers. The positive effect of vouchers on total income is smaller in specifications that correct for sample selection bias, and the substitution effect with other labor income sources is substantially larger. Overall, our findings show that AWAs tend to substitute standard employment, with small positive net effects on earnings, which are larger for intensive users of vouchers, and in geographic regions with a more sizable informal sector.

17.50-18.15 OPEN PANEL DISCUSSION WITH STATA DEVELOPERS • JOERG LUEDICKE AND MEGHAN CAIN, STATACORP

The "Open panel discussion with Stata Developers" session offers participants the opportunity to interact directly with StataCorp: providing participants with a forum in which to highlight any limitations encountered or suggest eventual improvements to the software.

20.00 *Conference Social Dinner (Optional)*

SCIENTIFIC COMMITTEE

Una-Louise Bell, TStat - TStat Training
Rino Bellocco, Università degli Studi di Milano-Bicocca
Giovanni Capelli, Istituto Superiore di Sanità
Giovanni Cerulli, IRCRES-CNR
Maurizio Pisati, Università degli Studi di Milano-Bicocca

CONTACT

Monica Gianni, TStat | TStat Training
Tel. +39 0864 210101 ext. 4

training@tstat.eu | formazione@tstat.it
www.tstattraining.eu | www.tstat.eu

TRAINING COURSE 10TH MAY 2024 | 9:00-16:45

INTRODUCTION TO PARTIAL LEAST SQUARES STRUCTURAL EQUATION MODELLING (PLS-SEM) USING STATA

PLS-SEM, also referred to as partial least squares path modelling, is an alternative approach to SEM which is being increasingly used in social sciences, psychology, business administration and marketing. PLS-SEM can be viewed as a component-based SEM alternative to the covariance-based structural equation modelling, which can be described as a factor-based SEM technique. As such, the PLS-SEM approach provides researchers with a multivariate statistical technique that can readily be used to estimate exploratory or/and complex SEM models. Although there are several stand-alone specialized PLS-SEM software packages available, this course introduces participants to the PLS-SEM methodology through the user-written Stata package, **plssem**, developed by the course instructors.

The course is of particular interest to researchers and professional working in social sciences, psychology, business administration, marketing and management. Due to its introductory nature, it is however, also accessible to individuals, regardless of their respective disciplines or fields, who need to acquire the requisite toolset to apply the PLS-SEM methodology to their own data. During the course, theoretical concepts are reinforced by applied case study examples, in which the course tutor discusses current research issues, highlighting potential pitfalls and the advantages of individual techniques.

COURSE OUTLINE

09:00-09:45	What is PLS-SEM?
10:00-10:45	PLS-SEM specification and estimation
10:45-11:00	*Coffee break*
11:00-11:45	Practical session: **pls-sem** commands and model specification
12:00-12:45	PLS-SEM assessment and interpretation models
12:45-14:00	Lunch
14:00-14:45	Practical session: advanced examples with assessment and interpretation
15:00-15:45	Mediation with PLS-SEM
16:00-16.45	Practical session: simple and complex mediation

Participants are expected to have previously followed a basic course in statistics. More specifically, a working knowledge of linear regression analysis is required. Previous exposure to Stata or other statistical software packages would also be an advantage.

The maximum number of participants admitted will be limited to 15.

LOCATION, DATE AND REGISTRATION

The 2024 edition of the Italian Stata Conference will take place in Florence on the 9th-10th May at Villa La Stella.

	Conference	Option 1 (1 night, Conference)	Conference plus Course	Option 2 (2 nights, Conference plus Course)
Full-time Students	€ 65	€ 145	€ 245	€ 405
Others	€ 95	€ 175	€ 375	€ 535

- All participation fees are subject to VAT (applied at the current Italian rate of 22%);
- Conference fees include: coffee breaks, lunch, course materials and a temporary licence of Stata for those participants attending the training course. We have also organized a limited number of single rooms (on a Bed and Breakfast basis) at Villa La Stella for those wishing also to stay in the complex. Accomodation will be given on a first come, first serve basis;
- Individuals interested in participating in the 2024 Italian Stata Conference should return their duly completed registration forms to the conference organizers by the 30.04.2024 at the latest.