



# German Stata Conference 2025

*The 22nd German Stata Conference takes place on **28 March** at the University of Hamburg. There will also be an optional workshop on **27 March**.*

Do you use Stata for your day-to-day work? Have you developed skills, or would you like to share your results and findings with others in the user community? Do you simply want to meet fellow researchers and StataCorp developers? The Stata Conference is an international event that provides Stata users worldwide with the opportunity to exchange ideas, experiences, and information on new applications of the software. The academic program of the meeting is being organized by Christian Brzinsky-Fay (University of Hamburg), Johannes Giesecke (Humboldt University Berlin), and Ulrich Kohler (University of Potsdam).

The conference language will be English because of the international nature of the meeting and the participation of non-German guest speakers. The logistics of the conference are being organized by DPC Software GmbH, distributor of Stata in several countries, including Germany, the Netherlands, Austria, the Czech Republic and Hungary (<http://www.dpc-software.de>).

## KeyNote Speaker

Professor Frauke Kreuter

Professor Frauke Kreuter is the Professor of Statistics and Data Science in Social Sciences and the Humanities at the Ludwig-Maximilians-University of Munich, Germany; Co-director of the Social Data Science Center (SoDa), and a faculty member in the Joint Program in Survey Methodology (JPSM) at the University of Maryland, USA.

## Stata Corp. Speaker

Di Liu

Di Liu is a Principal Econometrician in the econometric development team at StataCorp LLC. He is the primary developer of some Stata features, including heterogeneous DID, instrumental variable quantile regression, treatment effects estimation using lasso, lasso for prediction, lasso for inference, spatial autoregressive models, heckpoisson, and betareg. He also published research articles in Canadian Journal of Economics, Econometrics Reviews, Empirical Economics, Econometrics and Statistics, and the Stata Journal. Di has a PhD degree in economics from Concordia University in Montreal, Canada; an engineer's degree in software engineering and statistics from Polytech'Lille in Lille, France; and master's and bachelor's degrees in computer science from Hohai University in Nanjing, China.

# Overview

**Conference:** March 28, 2025, 9:00 - 17:30

**Workshop:** March 27, 2025, 10:00 - 17:00

**Venue:** University of Hamburg Albrecht-Mendelssohn-Bartholdy-Hörsaal (Hörsaal K), Edmund-Siemers-Allee 1, 20146 Hamburg

**Fees:** Conference only: 45 EUR (students: 20 EUR) Workshop only: 75 EUR (students: 35 EUR)  
Conference and workshop: 100 EUR (students: 50 EUR)

## Program

**8:30 - 9:15** Registration

**9:15 - 9:30** Welcome Christian Brzinsky-Fay

**9:30 - 10:30** Co-Creating with AI: The Role of LLMs as Intelligent Data Science Agents, Frauke Kreuter

**10:30 - 10:45** Coffee

**10:45 - 11:15** Into the Multiverse: Conducting and Visualizing Multiverse Analysis in Stata, Daniel Krämer

**11:15 - 11:45** Pairwise comparisons of means with unequal variances in Stata, Daniel Klein and Felix Bittmann

**11:45 - 12:15** `_gunitchg`: An egen-function for unit conversion, Ulrich Kohler

**12:15 - 13:15** Lunch Break

**13:15 - 14:30** Heterogeneous difference in differences, Di Liu

**14:30 - 15:00** StataNow and Beyond: How to select the best license model for your research and organization, Raoul Dittrich

**15:00 - 15:30** Coffee

**15:30 - 16:00** The Oaxaca-Blinder decomposition in Stata: an update, Ben Jann

**16:00 - 16:30** Recent Developments in Discrete-Time Multistate Estimation in Stata, Daniel C. Schneider

**16:30 - 17:30** Open panel discussion with Stata developers

**17:30** End of meeting



Please register for the conference and/or the workshop using the form on

<https://dpc-software.de/2025-german-stata-conference/>

For questions concerning enrolment, you can approach Tim Prenzel by E-mail at [Tim.Prenzel@dpc-software.de](mailto:Tim.Prenzel@dpc-software.de)

# Abstracts

## 9:30 - 10:30 Co-Creating with AI: The Role of LLMs as Intelligent Data Science Agents

**Frauke Kreuter (LMU München)**

*Abstract:* As AI advances, large language models (LLMs) are shifting from passive tools to active agents that collaborate with experts to co-create knowledge and artifacts. In this talk, we explore the role of LLMs as intelligent agents in data science workflows – partners that not only automate tasks but also enhance decision-making by understanding core data science principles, identifying cognitive biases, and nudging experts toward more robust conclusions.

We discuss how an LLM, equipped with statistical reasoning, ethical AI considerations, and an awareness of human cognitive pitfalls, can challenge assumptions, suggest alternative methodologies, and improve model interpretability. From guiding feature selection to questioning spurious correlations, these AI agents act as reflective collaborators rather than mere calculators.

We will examine case studies where LLMs have meaningfully influenced analytical processes, highlight challenges in aligning AI nudges with human intent, and explore the future of AI-augmented data science, generally and while using Stata.

This talk is primarily conceptual and designed to inspire but also to rethink our relationship with AI – not as a tool, but as a co-creator in the pursuit of knowledge.

## 10:45 - 11:15 Into the Multiverse: Conducting and Visualizing Multiverse Analysis in Stata

**Daniel Krähmer (LMU Munich, Department for Sociology)**

*Abstract:* Multiverse analysis is becoming an important tool in the methodological repertoire of social scientists. The idea behind the method—variously referred to as “multiverse analysis,” “multimodel analysis,” “specification curve analysis,” or “vibration of effects”—is straightforward: since there are many credible ways of formulating an analysis, and any single statistical estimate may suffer from selective reporting, multiverse analysis explores all reasonable specifications, contrasting authors’ preferred estimate with a range of possible estimates. Instead of luring readers into a dark corner of the “garden of forking paths,” multiverse analysis provides a bird’s-eye view of the maze of researcher decisions and the resulting range of defensible findings. While multiverse analysis holds significant promise for quantitative empirical research, it poses conceptual, computational, and practical challenges. This talk provides a primer on implementing multiverse analysis in Stata. It highlights the strengths and limitations of existing multiverse tools (e.g., `mrobust`, `multivrs`) and introduces a new plot type designed to visualize multiverse results effectively. By addressing key challenges in conducting and visualizing multiverse results, the talk seeks to encourage Stata users to adopt multiverse analysis and unlock its potential for robust and transparent research.



## 11:15 - 11:45 Pairwise comparisons of means with unequal variances in Stata

**Daniel Klein (DZHW) and Felix Bittman (LIFBI)**

*Abstract:* Researchers often want to mitigate the increased risk of type I errors that arises from multiple pairwise comparisons of means. Stata provides seven methods to adjust the corresponding confidence intervals and p-values. However, four of these methods assume equal sample sizes, variances, or both, and none explicitly addresses unequal variances, which might pose limitations on applied research. In this presentation, we briefly review how the implemented methods modify the significance level or obtain critical values from alternative distributions to adjust for multiple comparisons. We then discuss three methods that explicitly account for unequal variances by making additional adjustments to standard errors and degrees of freedom. Finally, we (re-)introduce the pwmc command in Stata, which implements these three methods, and compare their performance using a Monte Carlo simulation.

## 11:45 - 12:15 \_gunitchg: An egen-function for unit conversion

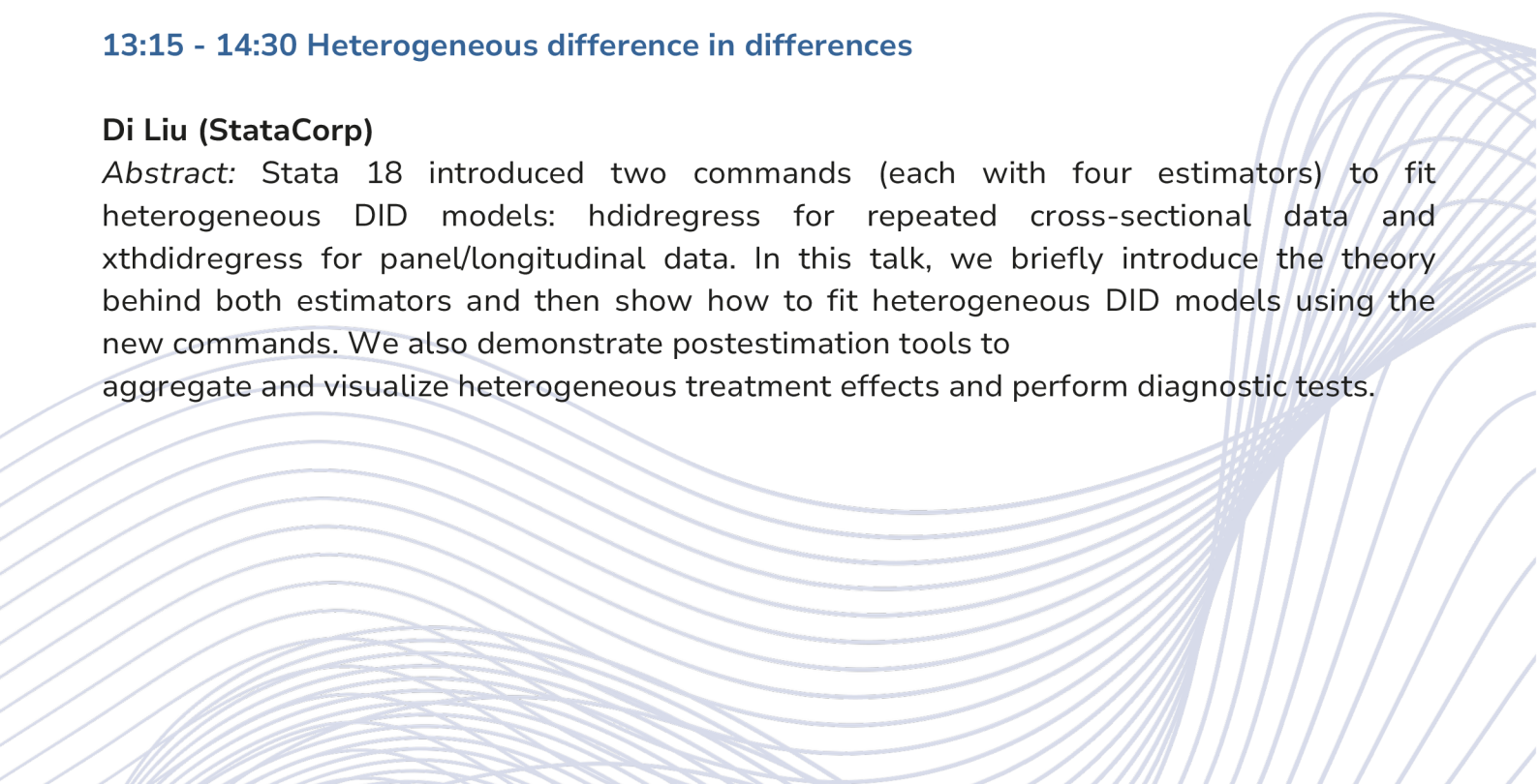
**Ulrich Kohler (University of Potsdam, Faculty for Economic and Social Sciences)**

*Abstract:* This talk presents an egen-function to convert units of measurements for length, areas, volumes, angles, masses, temperatures and currency. The function allows both, to convert many non-SI units (e.g., inch, furlong, sunradius) to SI-units (from pico to peta) or directly from a non-SI unit to another non-SI unit. Currencies are converted by calling the European Central Bank through an API. The conversion rate can be selected on a daily base or by an average of a specified period. German users may be relieved to realize that the function allows converting areas also into units of “Saarland”.

## 13:15 - 14:30 Heterogeneous difference in differences

**Di Liu (StataCorp)**

*Abstract:* Stata 18 introduced two commands (each with four estimators) to fit heterogeneous DID models: `hdidregress` for repeated cross-sectional data and `xthdidregress` for panel/longitudinal data. In this talk, we briefly introduce the theory behind both estimators and then show how to fit heterogeneous DID models using the new commands. We also demonstrate postestimation tools to aggregate and visualize heterogeneous treatment effects and perform diagnostic tests.



## 14:30 - 15:00 StataNow and Beyond: How to select the best license model for your research and organization

### **Raoul Dittrich (DPC Software GmbH)**

*Abstract:* The Stata license model is gradually changing from perpetual licenses to a pay-as-you-go model called StataNow. Whilst this gives researchers and users of Stata the advantage of always having access to the latest features of the software, the pay-as-you-go model requires different planning and budgeting for software. Many different options to license Stata exist, depending on edition, usage, organisation, and many other factors. This talk makes suggestions for finding a license option that meets the functional requirements, including multi-year models and covering of EUR/USD fluctuations.

## 15:30 - 16:00 The Oaxaca-Blinder decomposition in Stata: an update

### **Ben Jann (University of Bern)**

*Abstract:* In 2008, I published the Stata command `-oaxaca-`, which implements the popular Oaxaca-Blinder (OB) decomposition technique. This technique is used to analyze differences in outcomes between groups, such as the wage gap by gender or race. Over the years, both the functionality of Stata and the literature on decomposition methods have evolved, so that an update of the `-oaxaca-` command is now long overdue. In this talk I will present a revised version of `-oaxaca-` that uses modern Stata features such as factor-variable notation and supports additional decomposition variants that have been proposed in the literature (e.g., reweighted decompositions or decompositions based on recentered influence functions).

## 16:00- 16:30 Recent Developments in Discrete-Time Multistate Estimation in Stata

### **Daniel C. Schneider (Max Planck Institute for Demographic Research, Rostock)**

*Abstract:* Multistate life tables (MSLTs), or multistate survival models, have become a widely used analytical framework in the social and health sciences. These models can be cast in continuous or discrete time. The `-dtms-` Stata module (dtms stands for "discrete-time multistate"), which was presented at the German Stata Conference 2023 (Schneider 2023), implements the estimation of the discrete-time flavor of these models. This presentation first outlines discrete-time multistate estimation and then gives an overview of recent package enhancements. Among them are: External multinomial logistic estimation results, for example, from the interpolated Markov chain (IMaCh) executable (Brouard 2021), can be imported for further processing; difficulties with reloading saved dtms files across package versions have been resolved; the initial state distribution has been incorporated into the asymptotic analysis; new result type "evol" calculates the evolution of population fractions, along with the corresponding covariance matrix; estimation based on restricted transitions has been improved; transition probabilities can be based on time-varying covariate values; and several dtms trees can now be held in memory.

# Workshop

## Interaction between Text Writing and Statistical Analysis: Result Export and Dynamic Documents with Stata

Thursday, March 27 2025; 10am to 5pm

**Dr. Christian Brzinsky-Fay (University of Hamburg)**

*Abstract:* The workshop deals with the exchange of results between Stata and a Word processing program. In the first part, I will demonstrate the different options to customize and export tables from Stata into MS Word. We will learn how to create single tables using Stata's dtable and etable commands, and we will proceed to the more sophisticated use of the collect suite that is available since Stata 17.

In the second part, we will learn how to create dynamic or automated documents. These are documents containing particular commands (tags) that integrate up-to-date results (graphs, tables) into text documents, which avoids repeated and annoying copy-and-paste actions between Stata and MS Word. Using the dyndoc command, HTML or docx-files can be created. I will also explain how to customize Word or Excel files by using putdocx and putexcel.

The workshop addresses all students with a fundamental knowledge of Stata who aim to use Stata results in seminar papers or final theses.

### **Fees:**

Workshop only: 75 EUR (students 35 EUR)

Workshop and Meeting: 100 EUR (students 50 EUR)

### **Adress:**

University of Hamburg

Von-Melle-Park 9

20146 Hamburg

Please register for the conference and/or the workshop using the form on

<https://dpc-software.de/2025-german-stata-conference/>

For questions concerning enrollment you can approach Tim Prenzel by E-mail to

[Tim.Prenzel@dpc-software.de](mailto:Tim.Prenzel@dpc-software.de)



For detailed information on the full program and registration, please visit:

<https://www.stata.com/meeting/germany25/>

<https://dpc-software.de/2025-german-stata-conference/>