



Module Title: Advanced Methods of Sociological Data Science

Module Code: SOU44071

Module Name: Advanced Methods of Sociological Data Science

- **ECTS Weighting: 10**
- **Semester/Term Taught: Michaelmas Term**
- **Contact Hours: 22**
- **Module Personnel: Dr Yekaterina Chzhen**

Module Learning Aims and Outcomes:

On successful completion students will be equipped to:

- acquire critical methodological and practical skills in data science and data analytics which form the foundation for an academic career in research but also a career in the data-driven digital economy.
- be proficient in theory-driven and model-based analysis of quantitative data.
- be able to analyse linear but also categorical and limited dependent outcome variables in a regression context.
- understand and competently apply the concept of causal mediation and moderation analysis.
- be able to apply longitudinal methods to improve causal inference.
- be proficient in using the professional statistical software Stata to competently analyse data.
- be competent in the effective visualisation and presentation of statistical results.
- be competent in practical aspects of data management.

Module Content:

The goal of this fourth-year module is to equip students with methodical and technical skills in advanced methods of quantitative data analysis. It directly builds and expands upon the foundation in data analysis that had been laid by the third-year module 'Researching Society'. Primarily addressed are students who are interested to advance their quantitative research competence and seek a career in social science research or the rapidly growing market of data science. For students who are pursuing a piece of quantitative research in their Sociology dissertations and are attending the dissertation module, the advanced methods module provides an excellent complement. The module will dedicate to selected but practically important facets of quantitative data analysis. For example, the module will treat: testing causal hypotheses through mediation analysis, the analysis of conditional processes through moderation analysis (modelling of interaction terms), statistical models for categorical dependent variables which are very common in sociology and require special treatment (e.g., logistic, multinomial, Poisson regression), and an introduction to the world of longitudinal data analysis (e.g., repeated measurements, panel data).



models, and growth curve analysis). Data visualisation is key in interpreting and understanding results from (complex) statistical analyses as well as communicating those results to academic or non-academic audiences. Hence, the module will introduce good practices in presenting statistical findings through tables and figures. Finally, the module will also tap into challenges related to the everyday use of complex survey data (accounting for survey design such as stratification, clustering, and unequal selection) and the analysis of incomplete data (missing data). The module consists of a lecture (1 hour per week) and a lab session (1 hour per week). The module will discuss the methodological backgrounds of advanced methods (lectures) but will also teach the needed practical skills to apply methods using Stata (lab sessions), a professional data science software, with actual social science datasets. Students will replicate examples from Undergraduate: Sociology | Module Previews 2024/25 existing social science research that applied advanced methods to address specific research problems.

Recommended Reading List:

Specific readings will be made available on blackboard at begin of the module. A list of general indicative readings are presented below. Readings marked by asterisk (*) are recommended for beginners who do not yet meet the recommended requirements (see explanation above):

- Agresti, A., & Finlay, B. (2014). *Statistical Methods for the Social Sciences*. Essex: Pearson Education.

- *Bittmann, F. (2019). *Stata. A really short introduction*. Berlin: Walter de Gruyter.

- *Daniels, L., & Minot, N. (2020). *An Introduction to Statistics and Data Analysis Using Stata®. From Research Design to Final Report*. Thousand Oaks: SAGE.

- Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York: The Guilford Press.

- Healy, K., & Moody, J. (2014). Data Visualization in Sociology. *Annual Review of Sociology*, 40(1), 105–128.

- Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). *Applied Survey Data Analysis*. Boca Raton: Taylor & Francis.

Jaccard, J., & Turrisi, R. (2003). *Interaction Effects in Multiple Regression*. Thousand Oaks: SAGE Publications.

- Jaccard, J. (2001). *Interaction Effects in Logistic Regression*. Thousand Oaks: SAGE Publications.

- Long, J. S., & Freese, J. (2014). *Regression Models for Categorical Dependent Variables Using Stata*. Texas: Stata Press.



- Singer, J. D. & J. B. Willet (2003): Applied Longitudinal Data Analysis. Modeling Change and Event Occurrence. Oxford University Press, New York. (Part I on studying change)

Module Pre-Requisite: The module is at advanced level and therefore cannot reiterate basics of quantitative research methods. When joining the module, students ideally possess a basic understanding of quantitative data analysis (e.g., comparing means between groups, contingency table analysis, linear regression analysis) and at least rudimentary Stata skills, which has been taught in Researching Society (third year). Preferably students have heard Introduction to Social Research (in their second year) and Researching Society (in their third year) or comparable modules covering the basics of quantitative research, especially, the basics of quantitative data analysis, and ideally Stata. **Students who do not meet these requirements are still very welcome to join the module** but are expected to catch up quickly through auto-didactical learning of the basics in quantitative data analysis and Stata (using additional material provided by the lecturer and in the indicative reading list above).

Module Co-Requisite:

Assessment Details: 100% Practical assignments (2 x 50%).

Module Website: