Quantitative Social Research II LAW3287 – 20 credits



School of Law Level 3 Semester 2 - 2024/2025

Module Convenors

Prof. Jose Pina-Sánchez Email: j.pinasanchez@leeds.ac.uk Office: Liberty Building, G.38 Support hours: Tuesdays and Wednesdays 11.00 to 12.00

Prof. Dan BirksEmail: d.birks@leeds.ac.ukOffice: Liberty Building, 2.04Support hours: Mondays 10.00 to 11.00 (online) and Thursdays 15.00 to 16.00 (online)

Dr. Toby Davies Email: t.davies@leeds.ac.uk Office: Liberty Building, 2.15 Support hours: Tuesdays and Wednesdays 12.00 to 13.00

Teaching Assistant

Ms. Jade Parker Email: ss20jep@leeds.ac.uk Office hours: Mondays 12.00 to 13.00 and Thursdays 11.00 to 12.00

Module Summary

This module deepens in the content covered in SLSP3065 'Quantitative Social Research I'. Its goal is to familiarise students with a wider range of quantitative techniques for the empirical analysis of social phenomena. Special emphasis is placed on understanding the pitfalls of 'naive' data analyses. As such, students undertaken this module will be equipped with the skills to engage critically with the quantitative literature in the Social Sciences. The focus of the module is eminently applied. The teaching and learning activities rely on the exploration of widely used datasets covering topics of interest to Criminology, Criminal Justice, Sociology, and Social Policy.

Learning Outcomes

- Identify the appropriate methods to analyse different types of data, and to explore different types of research questions.
- Evaluate the strengths and limitations of those methods when employed either in their own research or in published research.
- Evidence ability to prepare, analyse and interpret quantitative data to explore substantive research questions.
- Effectively present and report findings from quantitative data analysis using graphs and tables in a clear way.

Learning Activities

The module is composed of eleven workshops, each of them divided into a one hour lecture and a one hour computer practical. To make the most of our limited contact time we follow a 'flipped-lecture' approach. This means that you are required to read the slides for the lecture and undertake the exercises included in the practical in advance of every workshop. Each practical is organised around one or two exercises, following a given dataset and set of instructions where text and R script is combined. In these practical exercises you are asked to engage with research questions stemming from the Criminology or Sociology literature. Near full written guidance (specifying the R commands to undertake the required analysis) will be offered for exercises demonstrating novel procedures. When dealing with procedures that have already covered in previous workshops, you will be requested to undertake the required analyses yourself, for which some light guidance will be offered in the form of prompts and hints. A valid - not necessarily the right - solution to the unguided part of the practical will be made available in Minerva at the end of each week.

Timetable

Week	Topic	Tutor	Workshop
14	Introduction	Jose	West Teaching Lab Cluster (G.29)
15	Selecting Explanatory Variables	Jose	West Teaching Lab Cluster (G.29)
16	Path Analysis and the Causal Framework	Jose	West Teaching Lab Cluster (G.29)
17	Nonlinear Effects	Jose	West Teaching Lab Cluster (G.29)
18	Time-Series	Jose	West Teaching Lab Cluster (G.29)
19	Data Reduction Techniques	Toby	West Teaching Lab Cluster (G.29)
20	Missing Data	Jose	West Teaching Lab Cluster (G.29)
21	Hierarchical Data	Jose	West Teaching Lab Cluster (G.29)
22	Longitudinal Data	Jose	West Teaching Lab Cluster (G.29)
23	Crime Mapping	Dan	West Teaching Lab Cluster (G.29)
24	Agent-Based Modelling	Dan	West Teaching Lab Cluster (G.29)

Week 14: Introduction

Introduction to the module. Discussion of the content to be covered and the module's assessment. Contrast with 'Quantitative Social Research I'; what will continue and what will change. Review of the assumptions invoked in regression modelling, linking them to the different sessions of the module. The practical will involve a refresher of 'Quantitative Social Research I', where we will review how to estimate and interpret results from linear and logit models. In addition, to avoid problems of incompatibility during the running of the upcoming practicals, we will proceed to install all the R libraries that we will use in each session, and make sure that we have installed the latest version of R, which in January 2025 is version 4.4.2, 'Pile of Leaves'.

Week 15: Selecting Explanatory Variables

Discussion of prediction and explanation as two different research goals that require different modelling strategies. Presentation of stepwise regression, and the variance inflation factor as a tool to detect multicollinearity. Practical using the Crown Court Sentencing Survey to be merged with area information from the Census and court characteristics from gov.uk reports: How accurately can we predict custodial sentences? Which carry a heavier weight in determining custodial sentences, aggravating or mitigating factors?

Recommended reading: Ruczinsky. Chapter 10 'Variable selection'

Week 16: Path Analysis and the Causal Framework

Introduction to DAGs as a tool to conceptualise causal effects. Presentation of confounding, collider and mediating variables. In the practical we will use data from an offenders' survey, to explore the question: Is the effect of procedural justice on offending mediated by perceptions of police legitimacy? In the second exercise we will explore the gender pay gap using the Labour Force Survey: Is there evidence of gender discrimination in the labour market?

Recommended reading 1: Cinelli, C., Forney, A., & Pearl, J. (2020). 'A Crash Course in Good and Bad Controls'

Recommended reading 2: VanderWeele, T. J., & Staudt, N. (2011). Causal diagrams for empirical legal research: A methodology for identifying causation, avoiding bias and interpreting results. Law, Probability & Risk, 10(4), 329-354.

Week 17: Non-Linear Effects

Revision of the assumption of linearity. Introduction to polynomial regression and lowess curves. Practical using academic salaries and the Labour Force Survey: What is the effect of age on salaries?

Recommended reading: Hanck et al. (2019). Chapter 8 'Nonlinear Regression Functions'

Week 18: Time-Series

Decomposition of the elements of a time-series (trend, seasonality, cycle and noise). Introduction to autoregressive and moving averages models and structural change tests. Practical using the MoJ Criminal Justice Statistics. Did the assault sentencing guidelines increased sentence severity?

Recommended reading: Hanck et al. (2019) Chapter 14 'Introduction to Time Series Regression and Forecasting'

Week 19: Data Reduction Techniques

Introduction to principal component analysis (summarising variables) and cluster analysis (summarising observations). Practical using violent crime rate data for US states, exploring several research questions. Which types of crime tend to co-vary? Which types have the greatest value when trying to discriminate between states? Application of clustering to identify a 'typology' of states, which could be useful if trying to find similar states.

Week 20: Missing Data

Introduction to the different forms of missing data and measurement error. Discussion of their implications to the validity and reliability of estimations. Introduction to adjustments based on weighting and imputation methods. Application of imputation methods to adjust for missing data problems in the Crown Court Sentencing Survey and explore the effect of offence seriousness on sentence length. Design of our own poststratification weights to enhance the representativity of the Leeds City Council Parks Survey, and find out what is the proportion of people who seldom use parks in Leeds.

Recommended reading: Yansaneh (2003) 'Construction and Use of Sample Weights'

Week 21: Hierarchical Data

Introduction to hierarchical data, implications of the violation of the assumption of independence. Distinction between treating such violation as a nuisance (sandwich estimator) or as substantively interesting information (through fixed effects and multilevel modelling). Practical based on the European Social Survey: What is the association between household income and social trust? Does this association vary across countries?

Recommended reading: Goldstein (1995) 'Chapter 2' Recommended online course: Brunton-Smith (2019) 'NCRM - Multilevel Modelling'

Week 22: Longitudinal Data

Introduction to retrospective and panel data survey designs. Distinction between cohort and age effects. Discussion of reverse causality. Introduction to growth curve models, time-to-event models and autoregressive models. Practical using the judicial data from the Czech Republic: Do judges sentence more harshly as they progress in their career?

Recommended reading 1: Steele (2008) 'Multilevel Models for Longitudinal Data' Recommended reading 2: Selig & Little (2012) 'Autoregressive and Cross-Lagged Panel Analysis for Longitudinal Data'

Week 23: Crime Mapping

Weeks 23 and 24 divert somewhat from the previous sessions - which mainly revolve around different forms of regression modelling - to give you an introduction to two different quantitative analytical techniques increasingly applied in both criminological social research. The techniques seen in these last two sessions won't be considered for Assessment.

Week 23 will cover the following: Introduction to geographic information systems (GIS) and crime mapping. Types and layering of spatial data. Features of crime and other agency data. Anatomy of recorded crime data. The geographic cone of resolution. Visualising geospatial patterns of crime through thematic mapping. Identifying crime hotspots. The ecological fallacy. The modifiable areal unit problem.

Week 24: Agent-Based Modelling

Introduction to computational agent-based modelling (ABM). Comparing aggregate and individual based modelling approaches. What is a complex system and why might ABM be useful for modelling them. Example of modelling crime patterns. Characteristics of agents in ABM. Model calibration and validation. A summary of challenges associated with ABM. Why model?

Core Reading

The course is self-contained in the sense that the workshops will cover all the necessary material to do the assignment. Having said this, it is useful to consult books, on-line resources, academic articles within your discipline using similar methods, and the recommended readings associated to each session.

Useful sources on Statistics and Methodology:

- Huntington-Klein, N. (2021). The effect: An introduction to research design and causality. CRC Press. Available online
- Hanck, Ch., Arnold, M., Gerber, A. & Schmelzer, M. (2019) Introduction to Econometrics with R. Available online
- Diez, D., Barr, C. & Çetinkaya-Rundel, M (2016) OpenIntro Statistics (3d edition). Available online
- Field, A., Miles, J. & Field, Z. (2012) Discovering Statistics using R. London: Sage

• Gelman, A. & Hill, J. (2006) Data Analysis using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

Useful sources on R:

- Grolemund, G. & Wickham, H. (2016) R for Data Science . O'Reilly. Available online
- Quick-R. Available online
- Beckerman, A. & Petchey. O. (2012) Getting Started with R. Oxford University Press
- Kabacoff, R. (2015) R in Action (2nd edition). Manning
- Teetor, P. (2011) R Cookbook . O'Reilly

Assessment

Formative assessment: There are two types of formative assessments embedded on this course. A collective form of formative assessment is available at the end of every workshop, where additional exercises are set out for students to undertake at their own pace, with the answers to those exercises made available at the end of the week. An individualised formative assessment is due to take place on Week 19, as part of the workshop on 'Data reduction techniques'. Here, just like in all other workshops you will have a set of exercises to complete on your own, which answers will become available at the end of the week. However, in this particular workshop you are also encouraged to submit your answers to Minerva, so we can offer individualised feedback.

Summative assessment: 3,000 word written assignment, where a dataset is to be used to explore a specific research question applying one or more of the techniques covered from Week 14 to Week 22. You can chose your own research question and dataset, which could be around the topic that you are already working on for your dissertation, although it cannot be the exact same question for reasons of plagiarism. Alternatively, you can chose to explore the research question and dataset identified in the assessment instructions, available on Minerva under 'Assessment'. The assignment has a free format, you do not have to replicate the structure of the final assignment for 'Quantitative Social Research I'. If you want to, you could simply follow a suggested template, which you can find in the assessment instructions in Minerva.

Attendance and Preparation

Attendance at weekly workshops is compulsory. Workshops provide an opportunity to apply what you have learnt, ask questions and participate in discussions, all of them important aspects of university education, and essential activities to acquire data analysis skills.

None of the classes in this module will be recorded, but all materials relating to the module will be made available through Minerva. You are required to prepare for each workshop by reading the lecture slides and replicating the instructions of the practical. For those exercises where full guidance is offered you are only required to reproduce the code, try to assimilate the rationale of the approach and bring any questions that might arise so we can discuss them during the workshop. You are also highly encouraged to explore the parts of the exercise where only light guidance (not full code) is offered. Even if you cannot complete these harder parts it would be extremely useful that you give them a try, since that will force you to write down your own code rather than simply running the code provided, which is the best way to learn data analysis and prepare yourself for the final assignment.

Support

You can find the office hours for each member of the teaching team in page 1. Do use them if you have any questions, general or specific. You should be able to find us then either in our respective offices or online through Teams. In addition, you can also use the Library's <u>Statistics</u> skills service to book '1on1' appointments to discuss questions about R or statistics more generally.