



Introduction

Syllabus

Assessment

Why more
Methods?

Quantitative Social Research II

Workshop 1: Introduction

Jose Pina-Sánchez



Building on SLSP3065

Introduction

Syllabus

Assessment

Why more
Methods?



Building on SLSP3065

Introduction

Syllabus

Assessment

Why more
Methods?

- Lots of similarities between the two modules
 - lecture + workshop
 - using R



Building on SLSP3065

Introduction

Syllabus

Assessment

Why more
Methods?

- Lots of similarities between the two modules
 - lecture + workshop
 - using R
- SLSP3065 is more important
 - covers the essentials of data analysis using R
 - the steep part of the learning curve
 - a process where each step builds upon the previous
 - to the point where you can undertake your own data analysis independently



Building on SLSP3065

Introduction

Syllabus

Assessment

Why more
Methods?

- LAW3287 seeks to broaden your data analysis toolbox
 - every workshop will cover a different method
 - we will meander, workshops do not build upon each other
 - SLSP3065 is like climbing a mountain, LAW3287 is like surveying the landscape from up there



Building on SLSP3065

Introduction

Syllabus

Assessment

Why more
Methods?

- LAW3287 seeks to broaden your data analysis toolbox
 - every workshop will cover a different method
 - we will meander, workshops do not build upon each other
 - SLSP3065 is like climbing a mountain, LAW3287 is like surveying the landscape from up there
- These techniques will allow you to understand and undertake cutting-edge quantitative social research
 - you will be able to interpret a bigger share of the literature
 - you will be able to tackle a wider range of research questions
 - and carry out more rigorous analysis to obtain more robust findings



Syllabus

Introduction

Syllabus

Assessment

Why more
Methods?

- Ten workshops ahead of us
 - we will learn one (sometimes two) new methods in each of them
 - today's workshop is just a recap of regression modelling
- Most methods are just derivations of the linear regression model
 - the linear model can be rarely applied using real world data
 - we will learn to circumvent some of their restrictive assumptions
 - will allow us to answer more interesting/complex questions



Syllabus

Introduction

Syllabus

Assessment

Why more
Methods?

- Ten workshops ahead of us
 - we will learn one (sometimes two) new methods in each of them
 - today's workshop is just a recap of regression modelling
- Most methods are just derivations of the linear regression model
 - the linear model can be rarely applied using real world data
 - we will learn to circumvent some of their restrictive assumptions
 - will allow us to answer more interesting/complex questions
- We will practice two important and transversal skills
 - model building strategies and cleaning up data
 - crucial 'know how' rarely taught, always assumed
 - in this module we will learn how the 'sausage is made'
 - we will use real data to replicate and expand studies published in academics journals



Assumptions

- Question: Can you remember some of the key assumptions invoked when estimating a linear model?

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_k X_k + e_i$$

what are the assumptions that will allow us to obtain an unbiased estimate of the effect of X on Y ?



Assumptions

- Question: Can you remember some of the key assumptions invoked when estimating a linear model?

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_k X_k + e_i$$

what are the assumptions that will allow us to obtain an unbiased estimate of the effect of X on Y ?

- normality: residuals are normally distributed
- homoskedasticity: the variance of the residuals is constant
- independence: residuals are independent of each other
- no multicollinearity: explanatory variables are not highly correlated



Assumptions

- Question: Can you remember some of the key assumptions invoked when estimating a linear model?

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_k X_k + e_i$$

what are the assumptions that will allow us to obtain an unbiased estimate of the effect of X on Y ?

- normality: residuals are normally distributed
- homoskedasticity: the variance of the residuals is constant
- independence: residuals are independent of each other
- no multicollinearity: explanatory variables are not highly correlated
- perfectly measured variables
- no missing data (other than missing at random)
- no unobserved confounders: we control for all common causes of X_1 and Y
- no reverse causality: Y does not cause X_1
- linearity: the effect of X_1 on Y is the same across the range of X_1



Assumptions

- Question: Can you remember some of the key assumptions invoked when estimating a linear model?

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_k X_k + e_i$$

what are the assumptions that will allow us to obtain an unbiased estimate of the effect of X on Y ?

- normality: residuals are normally distributed
- homoskedasticity: the variance of the residuals is constant
- independence: residuals are independent of each other
- no multicollinearity: explanatory variables are not highly correlated
- perfectly measured variables
- no missing data (other than missing at random)
- no unobserved confounders: we control for all common causes of X_1 and Y
- no reverse causality: Y does not cause X_1
- linearity: the effect of X_1 on Y is the same across the range of X_1
- Others: Stable Unit Treatment Value Assumption (SUTVA), no interference between units, consistency, etc.



Introduction

Syllabus

Assessment

Why more
Methods?

- 1 Introduction
- 2 Selecting explanatory variables
- 3 Path analysis and the causal framework
- 4 Non-linear effects
- 5 Time-series
- 6 Data reduction techniques
- 7 Missing data
- 8 Hierarchical data
- 9 Longitudinal data
- 10 Crime mapping
- 11 Agent-based modelling



Syllabus in Detail: Part 1

Introduction

Syllabus

Assessment

Why more
Methods?

① Introduction



Syllabus in Detail: Part 1

Introduction

Syllabus

Assessment

Why more
Methods?

① Introduction

② Selecting explanatory variables

Modelling strategies to predict \neq to explain

Stepwise regression

How accurately can we predict a custodial sentence?

Are aggravating factors more important than mitigating factors?



Syllabus in Detail: Part 1

Introduction

Syllabus

Assessment

Why more
Methods?

① Introduction

② Selecting explanatory variables

Modelling strategies to predict \neq to explain

Stepwise regression

How accurately can we predict a custodial sentence?

Are aggravating factors more important than mitigating factors?

③ Path analysis and the causal framework

Identify the different roles of variables in causal processes

Path analysis (mediating effects)

What is the role of police legitimacy in criminal offending?

What factors explain the gender pay gap?



Syllabus in Detail: Part 1

Introduction

Syllabus

Assessment

Why more
Methods?

① Introduction

② Selecting explanatory variables

Modelling strategies to predict \neq to explain

Stepwise regression

How accurately can we predict a custodial sentence?

Are aggravating factors more important than mitigating factors?

③ Path analysis and the causal framework

Identify the different roles of variables in causal processes

Path analysis (mediating effects)

What is the role of police legitimacy in criminal offending?

What factors explain the gender pay gap?

④ Non-linear effects

How can we model non-linear effects?

Polynomial regression and LOWESS curves

What is the effect of age on income?

Syllabus in Detail: Part 1

Introduction

Syllabus

Assessment

Why more
Methods?

① Introduction

② Selecting explanatory variables

Modelling strategies to predict \neq to explain

Stepwise regression

How accurately can we predict a custodial sentence?

Are aggravating factors more important than mitigating factors?

③ Path analysis and the causal framework

Identify the different roles of variables in causal processes

Path analysis (mediating effects)

What is the role of police legitimacy in criminal offending?

What factors explain the gender pay gap?

④ Non-linear effects

How can we model non-linear effects?

Polynomial regression and LOWESS curves

What is the effect of age on income?

⑤ Time-series

Analysing changes in time of aggregated indices

Autoregressive moving average (ARIMA) models

Have the sentencing guidelines made sentencing more severe?

Syllabus in Detail: Part 2

Introduction

Syllabus

Assessment

Why more
Methods?

- ⑥ Data reduction techniques
 - Can we find commonalities between cases / variables?
 - Cluster analysis and principal components analysis
- ⑦ Missing data
 - What are the implications of missing data, measurement error, or non-probability sampling methods?
 - Probability weights and imputation methods
 - Can we adjust for non-response in the Crown Court Sentencing Survey?*
 - Can we make our own surveys more representative?*
- ⑧ Hierarchical data
 - What can we do when cases are not independent?
 - Fixed effects, multilevel modelling and the sandwich estimator
 - Is the variability in self-reported happiness between countries larger than the variability between individuals?*
- ⑨ Longitudinal data
 - How can we model repeated observations across time?
 - Growth curve models and cross-lagged panel models
 - Do judges become more severe as they progress in their careers?*
- ⑩ Crime mapping
- ⑪ Agent-based modelling



The Teaching Team

Introduction

Syllabus

Assessment

Why more

Methods?

- Jose Pina-Sánchez (module leader)
 - In charge of workshops 1 to 5 and 7 to 9
 - Office hours: Tuesdays and Wednesdays 11.00 to 12.00
 - j.pinasanchez@leeds.ac.uk
- Toby Davies
 - In charge of workshop 6
 - Office hours: Tuesdays and Wednesdays 12.00 to 13.00
 - t.davies@leeds.ac.uk
- Dan Birks
 - In charge of workshops 10 and 11
 - Office hours: Mondays 10.00 to 11.00, Thursdays 15.00 to 16.00
 - d.birks@leeds.ac.uk
- Jade Parker
 - Teaching assistant across all practicals
 - Office hours: Mondays 12.00 to 13.00, Thursdays 11.00 to 12.00
 - ss20jep@leeds.ac.uk



Formative Assessment, Support and Independent Work

Introduction

Syllabus

Assessment

Why more

Methods?

- Exercises
 - normally two exercises in each practical
 - the first offers almost full guidance, the second only a few hints
 - a solution will be uploaded at the end of each week
 - to make the most of the workshop you are required to replicate the instructions of the guided exercise
 - and to give a good try to the unguided exercise in advance of the workshop
 - individual written feedback will be available for the exercise in Workshop 6 (Data reduction techniques) through Minerva



Formative Assessment, Support and Independent Work

Introduction

Syllabus

Assessment

Why more
Methods?

- Recommended readings
 - focus on those topics you want to explore in more depth
- Support hours
 - use them if you have any questions general or specific
 - I am also available after each workshop
- Statistics skills at the Library
 - you can book 1 on 1 appointments to ask questions about R code or statistics more generally



Summative

Introduction

Syllabus

Assessment

Why more
Methods?

- A 3,000 words data analysis study similar to that from SLSP3065, but using one of the new methods covered here
 - you can choose your own method and research question
 - perfectly fine to employ data that you are using in your dissertation
 - on a different topic and using different data from your SLSP3065 assignment
 - if you are short of ideas you can study the following:
Does gender affect perceptions of trust in the legal system? If so, which are the specific (direct or indirect) mechanisms behind that relationship?
 - which can be explored using the European Social Survey
 - deadline end of May (specific date tbc)



Introduction

Syllabus

Assessment

Why more
Methods?

Why more Methods?

- Isn't what we covered in SLSP3065 enough?



Why more Methods?

Introduction

Syllabus

Assessment

Why more
Methods?

- Isn't what we covered in SLSP3065 enough?
 - yes and no
 - you now have a good understanding of significance testing and regression analysis (the main data analysis method)
 - That's more than what most Profs in Sociology and Criminology know
 - but social reality is far more complex, and data much messier, than what you have seen so far



Introduction

Syllabus

Assessment

Why more
Methods?

Why more Methods?

- Isn't what we covered in SLSP3065 enough?
 - yes and no
 - you now have a good understanding of significance testing and regression analysis (the main data analysis method)
 - That's more than what most Profs in Sociology and Criminology know
 - but social reality is far more complex, and data much messier, than what you have seen so far
- We need to keep pressing on, so you are able to...
 - explore new research questions
 - do so more rigorously
 - be able to read critically most of the new research in the social sciences



Employability

Introduction

Syllabus

Assessment

Why more
Methods?

[Insight & Innovation - Quantitative Team](#)

Surrey County Council

Placed on: 21-08-2017 Salary: £32,839 to £55,644 per annum

Closes

Sep

18

[Insight & Innovation - Qualitative Team](#)

Surrey County Council

Placed on: 21-08-2017 Salary: £28,591 to £48,395 per annum

Closes

Sep

18

Because Data Analysis is the Future (and the present)

Introduction

Syllabus

Assessment

Why more
Methods?

