SMI606 Introduction to Quantitative Research: Module Outline

Dr. Calum Webb, Sheffield Methods Institute

2024-09-20

This module has been designed to help you develop your ability to conduct quantitative data analysis in the social sciences. The emphasis is not only on technical competence but also on understanding the principles behind the methods, as well as being able to competently interpret your results. We will be using real data with countless examples from across the social sciences (e.g., politics, economics, psychology, sociology, criminology, etc.) to learn about descriptive, exploratory, and inferential statistics. In doing so, we will cover topics such as data distributions, measurement, causality, correlation, regression, cluster analysis, spatial analysis, sampling and populations, and probability theory, using the R statistical programming language.

Learning Outcomes: What will you learn?

Make sure you revisit these learning outcomes occasionally and take stock of what outcomes you feel you are reaching, which ones you are struggling with, and which ones you might need more support with.

By the end of the module, you should be able to demonstrate that you:

1. Understand the nature and assumptions that underpin quantitative research and analysis in social research
2. Can produce appropriate descriptive statistics for different variable types, as well as data visualisations, to understand data distributions
3. Can produce statistics to quantify, describe, and make inferences about relationships between variables, including generalisation to populations and claims of causality
4. Can create statistical models for assessing the relationship between continuous and binary categorical dependent variables and multiple independent variables
5. Can use other quantitative methods for exploring clustering and grouping among observations
6. Can use `R` to show patterns spatial patterns and concentration

Teaching: How will I be taught and what do I need to do each week?

I would recommend setting up recurring events in your Google calendar at the start of term so you know when and where each session is.

		Hours
Lectures & Activities	Each week will include a lecture and learning activities designed to provide an accessible introduction to the topic being covered. Activities might include gentle introductions to putting these concepts into practice in `R`. There may be recommended reading for each week.	~10
Data Lab Sessions	After each lecture we will work on a supported exercise in `R` in a group environment, learning how to apply the quantitative analysis methods we have learned in `R` with `Rstudio`.	~20

Table 2: Week by week outline

Week	Topic	Recommended Reading
1	What is quantitative social science?	Powell, T. C. (2019). 'Can Quantitative Research Solve Social Problems?' & Mehmetoglu & Mittner (2022) Chapter 1, 2 & 3, Applied Statistics Using R
2	Types of quantification	Mehmetoglu & Mittner (2022) Chapter 4, Applied Statistics Using R
3	Relationships between variables	Mehmetoglu & Mittner (2022) Chapter 5, Applied Statistics Using R
4	Inference	Mehmetoglu & Mittner (2022) Chapter 6, Applied Statistics Using R
5	Causality	Goldthorpe, J. H. 'Causation, Statistics and Sociology'
6	Bivariate Linear Regression	Mehmetoglu & Mittner (2022) Chapter 7, Applied Statistics Using R
7	Reading Week	Catch up reading week
8	Multiple Linear Regression	Mehmetoglu & Mittner (2022) Chapter 8 & 9, Applied Statistics Using R
9	Logistic Regression	Mehmetoglu & Mittner (2022) Chapter 11, Applied Statistics Using R
10	Cluster Analysis	UC Business Analytics 'R' Programming Guide 'k-means Cluster Analysis' ([Available free online](https://uc-r.github.io/kmeans_clustering)) 'Hierarchical Cluster Analysis' ([Available free online](https://uc-r.github.io/hc_clustering))
11	Spatial Analysis	Imai, K. Chapter 5.3 'Quantitative Social Science: An introduction (Available online through the library)

Course outline: What will be covered each week?

Rather than giving 'homework' style tasks after each session that some people might be used to, you will often be expected to complete some preparation before each session: whether this is reading an article, completing a task, or watching a pre-recorded video. This is why it's important for you to always check in advance what preparation needs to be completed.

Below are details on which topics will be covered in each week of the course, along with some suggested reading on the topic. As the content becomes more complicated, I have chosen complementary readings where the first reading will cover the more theoretical aspects of the analytical method and the second reading will reinforce this theory and apply it in practice using `R`.

For comprehensive details of the suggested reading for each week, you should **use the reading and resource list**.

Reading and Learning Resources

You can find a detailed guide to recommended reading and independent learning activities in the Reading and Resources List page, which can be used in conjunction with the library Resources List.

Conventionally, a course like this would have one set methods textbook. However, because `R` is a statistical programming language used across multiple disciplines with a very large number of additional libraries that are continually being developed, I would encourage you to experiment with different learning materials to find what works best for you. One author's textbook chapter on data manipulation might, for example, not work as well for you as another's. Further, you should feel free to use online resources if this helps you learn — because `R` is so widely used, you will find a lot of existing support.

With that said, there one textbook I would strongly recommend that you buy or access from the library throughout the module:

- Mehmetoglu & Mittner (2022) 'Applied Statistics Using R: A guide for the social sciences'. (Around £36 from SAGE)

There is also a very short and accessible text I would recommend reading through to familiarise yourself with statistical and probability theory:

- Rowntree, D. (1981). 'Statistics without tears: An introduction for non-mathematicians'. (Around £10 also available for free on Archive.org)

In the reading and resource list you will find the following:

- **Recommended methods textbooks**, where the content has been checked to ensure it aligns to the learning outcomes of this module.
- A list of **online resources and activities designed to help people learn `R`**.
- **Academic reading material** on statistical and social science theory.

Responsibilities and Independent Learning

This is a 15-credit module; therefore, it is expected it will require 150 study hours per student (including the above teaching hours). As a rough guideline this means you should expect to devote around 10 hours of study time per week to this module during a typical 15-week semester.

Everyone will be starting this module with different kinds of experience: some people might have never studied any kind of social statistics before and many will have never had to write code. I am unlikely to set you 8-10 hours of directed independent learning each week; it is up to you to identify your learning needs and use the resources provided to meet them. I am happy to help you with this process and provide suggestions if you find yourself stuck for things to do each week.

To ensure your independent learning is productive, you should **regularly think about what parts of the module are challenging you**. Examples of this might be:

- Reading and manipulating data in `R`.
- Interpreting the output of statistical tests.
- Creating graphs.

You should then **use the reading and resources list to plan some independent learning activities for that week**.

Assessment

This module will be assessed through two written data analysis projects using R to ensure that you have achieved the aims and learning outcomes of this module.

The first 1,000-word limit submission counts for 30% of your final module grade and is due in December; the second 2,000-word limit submission counts for 70% of your grade and is due at the end of January. More details about these assessments will be made available as the module progresses. The assessments will be posted on Blackboard well in advance of the relevant deadlines.

Word limits

Please note that word limits will be enforced as per the student handbook. This means that you should not exceed the word limit. You will not be penalised for writing under the word limit, but should bear in mind that if you write substantially less than the limit you are unlikely to have included enough detail to demonstrate you have achieved all of the requirements for the assignment. Bibliographies (if necessary) are not included in word counts, but in-text citations are. You may choose whatever referencing format you wish (e.g., APA, Chicago, etc.), but please just be consistent throughout your submission.

Tables and graphs are **not** included in the word limit, though they should be used appropriately (there must be a good reason if long sections of text are included in a table). There are no limits on the number of tables or graphs you can use, but the content of all graphs and tables should be discussed and pertinent parts should be *at least* briefly described in the main text.

`R` code, comments, and output is also **not** included in the word limit, however, all pertinent information should always be in the main body of your assigment.[1] Comments should be restricted to explaining the purpose of the code, and not used for communicating or interpreting the results.

Submission

Coursework must be submitted online through Blackboard/Turnitin and by no later than 12:00pm (noon) on the day of the deadline. Any unauthorised late submissions after midday on the day of the deadline will incur a penalty of up to 100%, as per the student handbook. Marked coursework will generally be returned within 3 working weeks. The pass mark for this module is 50% overall. Any change to assessment arrangements will be announced in Blackboard.

Reassessment

In the case of students who fail the assessment of this module, repeat assessment will be by a written, essay examination. Further instructions will be communicated to those who fail the module. Resit candidates must consult Blackboard for further information, up until the time of the reassessment, not merely in semester time. Resit marks will be capped at 50%.

Use of Unfair Means in the Assessment Process

Because coursework, unlike examinations, is not invigilated, the University lays down general rules so that everyone is clear about what is acceptable practice. These rules are set out formally in the regulations for non-invigilated examinations in Part I of the University Calendar. Further information can be found here: Unfair Means – Advice to Students.

What constitutes unfair means? The basic principle underlying the preparation of any piece of academic work is that the work submitted must be your own work. Plagiarism, submitting bought or commissioned work, double submission (or self-plagiarism), collusion and fabrication of results are not allowed because they violate this principle. Rules about these forms of cheating apply to all assessed and non-assessed work. More details can be found in the student handbook.

What happens if I use unfair means? Any form of unfair means is treated as a serious academic offence and action may be taken under the Discipline Regulations. Where unfair means is found to have been used, the University may impose penalties ranging from awarding a grade of zero for the assignment through to expulsion from the University in extremely serious cases.

Detection of Unfair Means The University subscribes to a national plagiarism detection service called Turnitin which helps academic staff identify the original source of material submitted by students. It is also a resource, which can help tutors to advise students on ways of improving their referencing techniques. Your work will be checked by this service and results will be assessed by the marker.

Generative AI and Unfair Means Usage of generative AI constitutes unfair means when content that was created using generative AI technologies is presented as your own work. This means when generative AI has been used to create a part, or the entirety, of an assessment submission, either,

[1] You can use the Rstudio `wordcountaddin` Add In to get an accurate word count of your assignment if it is written in Rmarkdown, which you are encouraged to use (requires installation with `devtools`: `devtools::install_github("https://github.com/benmarwick/wordcountaddin")`).

a) Without proper or adequate acknowledgement, description, and/or evidence in the event that the use of generative AI is appropriate for the module learning objective, or
b) Where generative AI has been used to gain an unfair advantage in demonstrating that learning objectives have been met or in meeting the marking criteria for an assessment, unless some application of generative AI as an assistive or enhancive technology is explicitly stated as a learning outcome in the module handbook.

In the first instance, a good rule to follow is: 'if you copied output from a generative AI model directly into your assignment, this probably constitutes unfair means'. In addition, if you know there is output from a generative AI tool that you included in your assessment but did not acknowledge, describe, and evidence this in your assessment cover sheet, this would be considered unfair means.

In the second instance, as a general rule, if your use of generative AI is replacing, rather than reinforcing, your learning, this probably constitutes unfair means.

Generative AI outputs should not be used as sources for assessment and you should never cite anything from a generative AI tool. This is because content generated by AI tools is not reliable and is usually non-recoverable and non-reproducible at a later date, so it cannot be retrieved from a link or citation.

Illness, Coursework Extensions and Extenuating Circumstances

SMI's policy on coursework extensions and extenuating circumstances can be found in the SMI Postgraduate Handbook. If you wish to apply for a coursework extension, you should contact SMI in the first instance (smi@sheffield.ac.uk), in advance of the coursework deadline, and include the relevant Extenuating Circumstances Form plus medical certification if appropriate. Your application will then be considered, and the outcome communicated to you as soon as possible.

Please note that module leaders cannot grant extensions.

Feedback

Feedback on the assessments will generally be provided within three working weeks of the submission date. However, you will also receive regular feedback on non-assessed activities to help you improve your ability, understanding, and work prior to assessment.

Module Evaluation

We always welcome your comments and feedback as the module progresses, and informal evaluation of this kind will take place throughout the module. Students will also be invited to complete an anonymised module evaluation form towards the end of the module. These will be reviewed by SMI's Director of Learning and Teaching and the Module leader. A summary will then be considered by SMI's Staff-Student Committee and the SMI Learning & Teaching Committee. Any changes made to the module as an outcome of student evaluation will be communicated to new/future cohorts of students and in the following year's module Blackboard pages.