

*** Program for the Advanced Quantitative Methods Summer School, 2022 ****
*** 7 March 2022 ****

The Quantitative Methods Hub at the Department of Education, University of Oxford, offers its Advanced Quantitative Methods Summer School, this year consisting of eight different courses running during three calendar weeks in May (in Oxford parlance these are called weeks 2, 3 and 4 of Trinity Term; 5-20/5/2022). The courses require a basic understanding of multiple regression modelling or other multivariate techniques. Students and staff are welcome to attend one, some, or all days, by signing up using the link to the online shop.

We offer the courses in mixed format, in weeks 2 and 3 online, and in week 4 face-to-face with possibility to follow online. These include a mixture of pre-recorded introductions, materials, readings, and demonstrations, and synchronous lectures, workshops and questions & answer sessions. Face-to-face sessions will take place in Seminar Room D in the Department of Education (15 Norham Gardens, OX2 9AZ), and will be available online via Teams. The teaching formats, duration and times will vary from course to course (please see details for each course below). For the face-to-face teaching in Week 4 we reserve the right to go online should too few participants choose the face-to-face option, or if new covid-restrictions so demand.

We will use a variety of software, including R-studio, various R-modules, and Mplus (free demo-version, but also the full version) during the courses.

Our program is:

Oxford Week 2

| | | |
|------------------|-------------------|--------------------------------------|
| Thu 5/5 (online) | Thees Spreckelsen | Getting up and running in R |
| Fri 6/5 (online) | Thees Spreckelsen | Data management and documenting in R |

Oxford Week 3

| | | |
|--------------------------|---|--------------------------------------|
| Mon 9/5 (online) | Rens van de Schoot and Laura Laura Hofstee | AI-supported systematic reviews |
| Tue-Wed 10-11/5 (online) | Kit Double | Multilevel modelling (parts 1 and 2) |
| Fri 13/5 (online) | Luning Sun | Intro to IRT modelling |

Oxford Week 4

| | | |
|------------------------------------|---------------|--|
| Mon 16/5 (face-to-face and online) | Lars Malmberg | Intro Structural Equation Models (SEM) |
| Wed 18/5 (face-to-face and online) | Lars Malmberg | Longitudinal SEM |
| Fri 20/5 (face-to-face and online) | Lars Malmberg | Multilevel and intraindividual SEM |

Cost per course: £25 for OU and Grand Union students (or £60 for a week), £100 for staff and external students (or £250 for a week), and £200 for professionals (£450 for a week).

***** First week (Week 2 Trinity Term) *****

Course 1: Getting up and running in R

Instructor: [Thees F Spreckelsen](#)

Materials: online materials made available in advance of the session

Synchronous session: Thurs 5 May 2022, 10:00 to 15.30, UK-time, via Zoom

This course will introduce R for statistical analysis using the RStudio graphical user interface. The morning session will provide an introduction to the R language, the RStudio interface, and basic functions for data analysis. Applied examples will demonstrate R/RStudio broad functionalities. Participants will learn to work with objects, including how to create, save, and inspect objects, particularly data frame objects. They will create a simulated data frame and perform basic analysis functions with this object. In the afternoon, in a code-along session, participants will work with existing educational data from large-scale assessment studies. They will learn to read data into R and will apply data analysis functions presented in the morning session to real educational data. Participants will do basic data management steps, calculate descriptive statistics and perform regression analysis.

Aims: To enable participants to use the key elements of the statistical programming language R. Specifically using syntax to handle and analyse data. These building block will allow participants to use R for more complex statistical techniques.

Course prerequisites: It is assumed that participants will have a background in basic statistical methods up to, and including, regression analysis. Familiarity with syntax language from other statistical packages (e.g. Stata, SPSS) is desirable.

Preparation: 1) Participants should consider the reasons why they want to learn R in order to get the most out of the course. 2) Participants are asked to install R & RStudio: [follow this video tutorial to see how to install](#)

Contents:

10:00 – 12:00 (online) Introducing, R & RStudio software, basics of programming.

(11:30– 11:40 Break)

11:00 - 12:30 Base R: Data, Analysis, and custom packages.

(12:30 - 13:00 Break)

13:30 – 14:45 Working with educational data: Files, Subsets and Variables.

(14:45 – 15:00 Break)

15:00 – 15:30 Basic analyses: Regressions, Formulae, and Functions.)

Cost per course: £25 for OU and Grand Union students (or £60 for a week), £100 for staff and external students (or £250 for a week), and £200 for professionals (£450 for a week).

Course 2. Data management, graphing and documenting in R

Instructor: [Thees F Spreckelsen](#)

Materials: on-line materials made available in advance of the session

Zoom-sessions: Fri 6 May 2022, 10:00 to 15.30

Quantitative research is often taught with a focus on analyses techniques. However, data management, visualisation and “documenting what one has done”, are central to doing quantitative research, as well as often taken the most time. This course will introduce R’s functionalities for manipulating and visualising data, as well as to RMarkdown, which provides a very easy way to report results and document all steps from data input to the analyses in a transparent and reproducible way.

The first part of the course will cover: data import, ways to handle different variable types (strings, time), data-cleaning (new variables/recoding, missing values), how to do repeated operations, and workflows for data management. The second part will introduce the extremely versatile graphing package `ggplot2`, look at ways of plotting coefficients, as well as simple ways to generate interactive visualizations. In the third part participants will learn how to generate reports, presentations and webpages almost automatically from within R using RMarkdown, and to customize key elements of their RMarkdown documents.

The content of the course will finally be brought together by going from a raw dataset to a fully documented report.

Aims: To enable participants to efficiently handle and manipulate data, do powerful visualizations and providing the skills to generate graphs for any possible scenario. To make participants confident users of R’s facilities to generate reports, presentations and webpages.

Course prerequisites:

- Participants should have either past experience of working with quantitative data, or know a dataset they like to work with, or intend to collect quantitative data.
- It is assumed that participants will have a background in *basic statistical methods* up to, and including, regression analysis.
- Participants should have *used R/RStudio* previously, there will only be a minimal introduction.

Self-check: What do the following functions do?

```
“df[1,2]”, ”lm(z ~ x1+x2, data= df)”, ”select(df, gender)”
```

Contents:

10:00 – 10:15 Introducing, R & RStudio refresher

10:15 – 11:00 Data management: Files, Formatting, Variables and “best-practices”.

11:00 – 11:15 Break

11:15 – 12:30 String and time variables, repeated operations

12:30 – 13:00 Break

13:00 – 14:30 Visualisation using `ggplot`

14:30 – 14:45 Break

14:45 – 15:30 RMarkdown.

Cost per course: £25 for OU and Grand Union students (or £60 for a week), £100 for staff and external students (or £250 for a week), and £200 for professionals (£450 for a week).

***** Second week (Week 3 of Trinity Term) *****

Course 3: Automated systematic reviews

Instructors: Rens van de Schoot and Laura Hofstee

Pre-recorded videos and on-line materials made available in advance of the session

Synchronous session: Mon 9 May, 08:00 – 12:00, UK-time

The pre-conference workshop on ‘Automated systematic reviews’ is taught by [Rens van de Schoot](#), co-author of the famous [Multilevel Book](#) of Joop Hox, his work on [Bayesian methods](#), and his recent project [ASReview](#).

More and more researchers rely upon Systematic Reviews to synthesize state of the art in a particular scientific field. However, the scientific output of the world doubles every nine years. In this tsunami of new (multilevel) knowledge, there is not enough time to read everything – resulting in costly, abandoned, or error-prone work. Using the latest methods from the field of Artificial Intelligence (AI), you can reduce the number of papers to screen up to 95%! This means you can find relevant multilevel papers much quicker. Also, you can provide a systematic overview in the introduction section of your multilevel paper.

During the synchronous session, Rens will introduce you to the open-source software ASReview which enables you to screen a large number of multilevel papers to find the relevant ones needed for your introduction section. We will work in a small group so that you can optimally benefit from the workshop. After the workshop, you will be able to use the AI-aided tool to create a systematic overview for your next paper.

Note that this workshop will be taught online in GatherTown. Participants will receive a log-in after registration. More information on ASReview can be found in the paper published in [Nature Machine Intelligence](#), the website [asreview.ai](#), or the [Github organization](#).

This workshop is from 10:00 till 14:00 (UK-time).

Contents

- Main ideas of systematic reviews
- Introduction to open-source software ASReview
- Small-group work in breakout rooms
- Q&A session

Cost per course: £25 for OU and Grand Union students (or £60 for a week), £100 for staff and external students (or £250 for a week), and £200 for professionals (£450 for a week).

Course 4. Multilevel Modelling

Instructor: [Kit Double](#)

Pre-recorded videos and on-line materials made available in advance of the session

Synchronous introduction session (Part 1): Tue 10 May, 9:00 – 11:00, UK-time

Synchronous follow-up session (Part 2): Wed 11 May, 9:00 – 11:00, UK-time

This course will introduce multilevel modelling (MLM) and provide an analytical framework for the study of research questions with MLM in R. During the morning session, lectures will introduce the rationale for MLM, provide applied examples in educational research and introduce a framework for the analysis of substantive research questions using MLM.

Participants will learn how to set up, estimate, and interpret multilevel models in R. In the afternoon, participants will perform MLM on existing educational data, including specifying different model types, interpreting output and producing visual representations of MLM designs. The last part of the course will be dedicated to analysing growth curve models in R using an example dataset.

Course prerequisites: Participants need to understand the basics of multiple regression, or other relevant multivariate statistics. Ideally participants will have attended the ‘Introduction to R for the Analysis of Educational Data’ or have a basic knowledge of how to use R. Students should [install R](#) and [Rstudio](#) on their own laptop prior to the course.

Contents

- Overview of MLM – Models Types and Research Questions
- Running MLM in R
- Addressing research questions with MLM in R
- Growth Curve MLM in R

Cost per course: £25 for OU and Grand Union students (or £60 for a week), £100 for staff and external students (or £250 for a week), and £200 for professionals (£450 for a week).

Course 5. Introduction to IRT modelling

Instructor: [Luning Sun](#)

Materials will be made available in advance via dropbox

Synchronous zoom session: Fri 13 May 2022, 10:00 to 15.30, UK-time, via zoom

Contents

Item Response Theory (IRT) refers to a family of mathematical models that attempt to explain the relationship between the properties of items, individuals' responses to these items and the underlying trait being measured. IRT assumes that the latent construct and items of a measure are organised in an unobservable continuum. Its main purpose focuses on establishing the individual's position on that continuum. Often regarded as superior to Classical Test Theory, IRT has been widely applied to the design, analysis and scoring of tests, questionnaires and similar instruments measuring latent constructs. In this workshop, we will first learn the basic concept of IRT and different IRT models. In the second part, we will implement IRT modelling using R. As a prerequisite, participants will need working knowledge of R.

Contents

10:00-10:30 Classical Test Theory vs Item Response Theory

10:30-11:00 Different IRT models and item parameters

11:00-11:15 Morning break

11:15-12:00 IRT in R (ltm package)

12:00-12:30 IRT in R (mirt package)

12:30-13:00 Lunch break

13:00-13:40 IRT assumptions

13:40-14:15 DIF

14:15-14:30 Afternoon break

14:30-15:15 Polytomous IRT models

15:15-15:30 Q&A

Cost per course: £25 for OU and Grand Union students (or £60 for a week), £100 for staff and external students (or £250 for a week), and £200 for professionals (£450 for a week).

***** Third week (Week 4 of Trinity Term) *****

Course 6. Introduction to Structural Equation Modelling

Instructor: [Lars Malmberg](#),

Pre-recorded videos and materials made available in advance of the session

Face-to-face session: Mon 16 May 2022, 09:00 to 13:00, Seminar Room D, UK-time, (with possibility to attend online via Teams, broadcasted from Sem Room D)

The concept of a latent construct is central in the social sciences. A latent construct is a non-directly observed phenomenon (e.g., attitude, socioeconomic status) that we can model using manifest (observed) variables (e.g., survey and questionnaire responses, observation scores), by partitioning out residual (i.e., uniqueness, error variance). The structural equation model (SEM) is divided into two parts. In the *measurement part* of the model, we can inspect whether manifest variables measure the constructs they are intended to measure. This model is called confirmatory factor analysis (CFA) which allows the researcher to test whether an a priori model fits data, and whether this also holds across multiple groups. If measurement is satisfactory, the relationships between constructs can be estimated in the *structural part* of the SEM. Complex relationships between manifest variables and/or latent constructs can be tested in path-models not possible to specify in the multiple regression framework. During the course we will cover worked examples relevant for educational, psychological and social sciences.

Course Pre-requirements

Participants need to understand the basics of multiple regression, or other relevant multivariate statistics.

Contents

Video-clip 1 Basic concepts, models and measurement. From multiple regression to path-models using manifest variables.

Video-clip 2 Observed (manifest) variables and unobserved (latent) constructs. Specification of measurement models for testing quality of measurement, using continuous and dichotomous manifest variables. Goodness-of-fit indices.

Demonstration video-clips. Models in R-lavaan

Demonstration video-clips. Models in Mplus

Monday 16 May

09.00-10.30 Questions and answers session based on pre-recorded materials (questions can be posted in advance in the Q&A document)

10.30-10.45 Break

10.45-11.45 Work in small-groups or virtual breakout rooms

11.45-12.00 Break

12.00-13.00 Further questions and answer session, also about own data

Software: We will mainly use the [Mplus demo software](#). Some parallel code is available in [R \(Lavaan\)](#)

Cost per course: £25 for OU and Grand Union students (or £60 for a week), £100 for staff and external students (or £250 for a week), and £200 for professionals (£450 for a week).

Course 7. Structural Equation Modelling of longitudinal data

Instructor: [Lars Malmberg](#),

Pre-recorded videos and materials made available in advance of the session

Face-to-face session: Wed 18 May 2022, 09:00 to 13:00, Seminar Room D, UK-time, (with possibility to attend online via Teams, broadcasted from Sem Room D)

This follow-up of the two-day introduction to SEM is an advanced course in which we focus on SEM for longitudinal data. Prospective longitudinal data is usually collected over longer periods of time (e.g., term or year) while intensive longitudinal is gathered within shorter time-spans (e.g., numerous times a day). Using SEM we can model repeated latent constructs over time using autoregressive models (i.e., a construct at the concurrent time-point is regressed on a construct at a previous time-point). We can also specify latent change models using “phantom constructs”. When particular interest is in individual differences in change over time, we can model time explicitly in the latent growth curve model. The worked examples are based on educational longitudinal data, relevant for social sciences.

Pre-requirements:

Participants need to understand the basics of multiple regression, other relevant multivariate statistics, and have some exposure to either regression or SEM.

Contents

Video-clip 1: Introduction to longitudinal (repeated measures modelling)

Video-clip 2: Auto-regressive modelling

Video-clip 3: Growth modelling

Wednesday 18 May

09.00-10.30 Questions and answers session based on pre-recorded materials (questions can be posted in advance in the Q&A document)

10.30-10.45 Break

10.45-11.45 Work in small-groups or virtual breakout rooms

11.45-12.00 Break

12.00-13.00 Further questions and answer session, also about own data

Software: We will mainly use the [Mplus demo software](#). Some parallel code is available in [R \(Lavaan\)](#)

Cost per course: £25 for OU and Grand Union students (or £60 for a week), £100 for staff and external students (or £250 for a week), and £200 for professionals (£450 for a week).

Course 8. Structural Equation Modelling for intraindividual data

Instructor: [Lars Malmberg](#),

Pre-recorded videos and materials made available in advance of the session

Face-to-face session: Fri 20 May 2022, 09:00 to 13:00, Seminar Room D, UK-time, (with possibility to attend online via Teams, broadcasted from Sem Room D)

Multilevel structural equation modelling (MSEM) combines the best of two worlds, the multilevel model (MLM) and SEM. The multilevel SEM (MSEM) allows us to test structural validity in two or more hierarchical levels, and specify level-specific associations between level-specific predictors and outcomes. In multilevel structures we can investigate whether the metric at different levels of the hierarchical structure is the same, in order to draw equivalent conclusions about effects at the different levels. MSEM can be applied to different hierarchical data structures quite commonly found in educational research, e.g., students nested in classrooms, or time-points nested in students.

In the first session we introduce multilevel modelling using manifest indicators, comparison of notation in MLM and MSEM, and model specification in the SEM framework. We will then move on to the specify factor structures in two levels, considering students nested in teachers/schools.

In the second session we specify latent constructs for intraindividual data (time-points nested in students) and include level-specific predictors. We specify fixed and random effects models assuming “individuals as their own controls” type of models, in which the time perspective is not specified (Malmberg, 2020).

In the third session we apply Dynamic SEM assuming stationarity (no mean trends over time), specifying equidistant time-lags for lagged variables in diary data (a working-life diary for a year). These time-series like models can be specified using the Bayesian estimators, allowing us to investigate within-person variability.

Contents

Video-clip 1: Introduction to multilevel modelling in SEM, multilevel factor structures

Video-clip 2: Intraindividual SEM

Video-clip 3: Dynamic SEM

Friday 20 May

09.00-10.30 Questions and answers session based on pre-recorded materials (questions can be posted in advance in the Q&A document)

10.30-10.45 Break

10.45-11.45 Work in small-groups or virtual breakout rooms

11.45-12.00 Break

12.00-13.00 Further questions and answer session, also about own data

Software: We will mainly use the [Mplus demo software](#). Some parallel code is available in [R \(Lavaan\)](#)

Cost per course: £25 for OU and Grand Union students (or £60 for a week), £100 for staff and external students (or £250 for a week), and £200 for professionals (£450 for a week).