

**** Program for the Advanced Quantitative Methods Summer School, 2020 ****

The Quantitative Methods Hub at the Department of Education, University of Oxford, offers its annual Advanced Quantitative Methods Summer School, this year consisting of seven different courses during running Tuesday to Friday during two calendar weeks (in Oxford parlance these are called week 3 and 4 of Trinity Term; 11-22/5/2020). The courses require an 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 this portal:

<https://www.oxforduniversitystores.co.uk/conferences-and-events/department-of-education/events/advanced-quantitative-methods-summer-school-2020>

The courses will take place at the Department of Education at 15 Norham Gardens, OX2 9AZ, Oxford. Please bring along your own lap-top as we will not be in a computer class.

The program is:

First week

Tue 12/5	10:00-15:30	Thees Spreckelsen	Intro to R
Wed 13/5	10:00-15:30	Thees Spreckelsen	Data management, graphing and documenting in R
Thu 14/5	10:00-15:30	Josh McGane	Intro to Rasch modelling
Fri 15/5	10:00-15:30	Kit Double	Multilevel modelling

Second week

Tue 19/5	10:00-15:30	Yves Rosseel	Structural Equation Modelling (SEM) in R lavaan, part 1
Wed 20/5	10:00-15:30	Yves Rosseel	SEM in R lavaan, part 2
Thu 21/5	10:00-15:30	Lars Malmberg	Longitudinal SEM
Fri 22/5	10:00-15:30	Lars Malmberg	Multilevel and intraindividual SEM

***** First week (Week 3 Trinity Term) *****

Day1: Introduction to R for the Analysis of Educational Data

Instructor: Thees F Spreckelsen

Time: Tuesday 12 May 2020, 10:00 to 15.30,

Location: Seminar Room D (bring you own laptop), Department of Education,
University of Oxford, 15 Norham Gardens, OX2 9AZ, Oxford

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 datamanagement 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: Participants should consider the reasons why they want to learn R in order to get the most out of the course.

Contents:

10:00 – 10:45 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 Lunch

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: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£80 / £320 / £600 for the week). Handouts, lunch and refreshments included in price.

Day 2. Data management, graphing and documenting in R

Instructor: Thees F Spreckelsen

Time: Wednesday 13 May 2020, 10:00 to 15:30,

Location: Seminar Room D (bring your own laptop), Department of Education,
University of Oxford, 15 Norham Gardens, OX2 9AZ, Oxford

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 ggplot, 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, workflows

12:30 – 13:00 Lunch

13:00 – 14:30 Visualisation using ggplot

14:30 – 14:45 Break

14:45 – 15:30 RMarkdown.

Cost: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£80 / £320 / £600 for the week). Handouts, lunch and refreshments included in price.

Day 3. Introduction to Item Response Theory and Rasch modelling

Instructor: Joshua McGrane

Time: Thursday 14 May 2020, 10.00 to 15.30,

Location: Seminar Room D (bring your own laptop), Department of Education,
University of Oxford, 15 Norham Gardens, OX2 9AZ, Oxford

Over the past decades, Item Response Theory has become increasingly used as a validation tool for assessments and questionnaires in a range of fields, including education, psychology and health sciences. This one-day workshop will introduce Item Response Theory for assessment and survey data, with a particular focus upon Rasch modelling. Firstly, an overview of the conceptual and philosophical foundations of Item Response Theory will be provided, including the key differences between the models and their applications. The workshop will then focus upon Rasch modelling in more detail, including the different models for test and survey data, the information these models provide regarding the reliability and validity of assessments, and in particular, how the Rasch model is a powerful tool for diagnosing anomalies in assessment and survey data. Participants will then be introduced to Rasch modelling in R using the RStudio graphical user interface.

The workshop will be both theoretical and hands-on, and participants are encouraged to bring their own data for the practise session in the afternoon. An example dataset will be provided for those who do not bring their own data.

Course prerequisites: It is assumed that participants will have a background in basic statistical methods and familiarity with statistical software.

Contents:

10:00 - 10:15 Welcome, agenda

10:15 - 11:00 Foundations of psychometrics and Item Response Theory: basic concepts, different IRT models, paradigmatic differences

11:00 - 11:15 Break

11:15 - 12:30 Introduction to Rasch modelling: dichotomous and polytomous models, model estimation, tests of model fit

12:30 - 13:00 Lunch

13:00 - 13:45 Rasch modelling using R

13:45 - 14:00 Break

14:00 - 15:30 Practise session using R

Cost: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£80 / £320 / £600 for the week). Handouts, lunch and refreshments included in price.

Day4. Multilevel Modelling

Instructor: Kit Double

Time: Friday 15 May 2019, 10:00 to 15:30,

Location: Seminar Room D (bring you own laptop), Department of Education,
University of Oxford, 15 Norham Gardens, OX2 9AZ, Oxford

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 (<https://www.r-project.org/>) and Rstudio (<https://www.rstudio.com/>) on their own laptop prior to the course.

Contents:

10:00 - 11:30 Overview of MLM – Models Types and Research Questions

11:30 - 11:45 Break

11:45 - 12:30 Growth curve models and MLM resources 12:30 - 13:00 Lunch

13:00 - 14:30 Addressing research questions with MLM in R

14:30 - 14:45 Break

14:45 - 15:30 Growth Curve MLM in R

Cost: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£80 / £320 / £600 for the week). Handouts, lunch and refreshments included in price.

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

Days 5 and 6. Structural Equation Modelling in R lavaan, parts 1 and 2

Instructor: Yves Rosseel,

Time: Tuesday 19 May and Wednesday 20 May 2020, 10.00 to 15.30,

Location: Seminar Room D (bring you own laptop), Department of Education,
University of Oxford, 15 Norham Gardens, OX2 9AZ, Oxford

Structural equation modeling (SEM) is a general statistical modeling technique to study the relationships among observed variables. It spans a wide range of multivariate methods including path analysis, mediation analysis, confirmatory factor analysis, growth curve modeling, and many more. SEM is often used to test theories or hypotheses that can be represented by a path diagram. In a path diagram, observed variables are depicted by boxes, while latent variables (hypothetical constructs measured by multiple indicators) are depicted by circles. Hypothesized (possibly causal) effects among these variables are represented by single-headed arrows. If you had ever found yourself drawing a path diagram in order to get a better overview of the complex interrelations among some key variables in your data, this course is for you.

The first day of the course provides an introduction to the theory and application of structural equation modeling. On the second day, we discuss several special topics that are often needed by applied users (handling missing data, nonnormal data, categorical data, longitudinal data, etc.). Hands-on sessions are included in order to ensure that all participants are able to perform the analyses using R package lavaan.

Course prerequisites: Participants should have a solid understanding of regression analysis and basic statistics (hypothesis testing, p-values, etc.). Some knowledge of exploratory factor analysis is recommended, but not required. Because lavaan is an R package, some experience with R (reading in a dataset, fitting a regression model) is recommended, but not required.

Contents:

Tuesday 19 May

10:00 - 11:15 A gentle introduction to SEM

11:15 - 11:30 Break

11:30 - 12:30 Model estimation and model evaluation

12:30 - 13:00 Lunch

13:00 - 14:15 Introduction to lavaan

14:15 - 14:30 Break

14:30 - 15:30 Practice session or work on your own data

Wednesday 20 May

10:00 - 11:15 Multiple groups and measurement invariance

11:15 - 11:30 Break

11:30 - 12:30 Missing data, nonnormal and categorical data

12:30 - 13:00 Lunch

13:00 - 14:15 Longitudinal and Multilevel SEM

14:15 - 14:30 Break

14:30 - 15:30 Practice session or work on your own data

Software: the R package lavaan (<http://lavaan.org>)

Cost per day : £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£80 / £320 / £600 for the week). Handouts, lunch and refreshments included in price.

Day 7. Structural Equation Modelling of longitudinal data

Instructor: Lars Malmberg

Time: Thursday 21 May 2019, 10.00 to 15.30,

Location: Seminar Room D (bring your own laptop), Department of Education,
University of Oxford, 15 Norham Gardens, OX2 9AZ, Oxford

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. We will cover worked examples relevant for educational, psychological and social sciences.

Pre-requirements:

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

10:00 - 11:00 Overview of SEM for longitudinal data. Repeated measures and autoregressive models with manifest and latent constructs.

11:00-11:15 Break

11:15 - 12:30 The reciprocal effects model. The latent change model.

12:30 - 13:00 Lunch

13:00 - 14:15 The latent growth model. Coding of time and error structures.

14:15 - 14:30 Break

14:30 - 15:30 Practice session or work on own data

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

Cost: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£80 / £320 / £600 for the week). Handouts, lunch and refreshments included in price.

Day 8. Structural Equation Modelling for multilevel and intraindividual data

Instructor: Lars Malmberg

Time: Friday 22 May 2019, 10.00 to 15.30,

Location: Seminar Room D (bring you own laptop), Department of Education,
University of Oxford, 15 Norham Gardens, OX2 9AZ, Oxford

In this follow-up of the introduction to SEM we focus on SEM for multilevel data. We start with an introduction to multilevel modelling, its notation, and model specification in the SEM framework. We will then consider two kinds of nested structures: students nested in institutions (e.g., students nested in schools), and time-points nested in persons.

Both cross-sectional and longitudinal data can be collected applying a nested structure (e.g., students in classrooms, parents in families, time-points in persons). Using SEM we can model repeated latent constructs over time, or across hierarchical levels net of measurement error. We will start by defining the multilevel model within the structural equation modelling framework. We will then introduce multilevel confirmatory factor analysis (MCFA) and full multilevel structural equation models (MSEM). In the worked examples we will use example data of both persons nested within organisations and time-points nested within persons.

During the course we will cover worked examples relevant for educational, psychological and social sciences. Participants need to understand the basics of multiple regression, other relevant multivariate statistics, and have some exposure to either multilevel regression or SEM.

Contents:

10:00 - 11:00 Introduction to MLM. Fixed effects and random effects. Comparison of MLM notation and visual depiction of MLM in the SEM framework.

11:00-11:15 Break

11:15 - 12:30 Factor structures across levels. Structural models at the within and between levels.

12:30 - 13:00 Lunch

13:00 - 14:15 Time-points nested in persons, parents nested in families

14:15 - 14:30 Break

14:30 - 15:30 Practice session or work on own data

Cost: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£80 / £320 / £600 for the week). Handouts, lunch and refreshments included in price.