*** Program for the Advanced Quantitative Methods Summer School, 2021 ****
*** version 22 March 2021 ****

The Quantitative Methods Hub at the Department of Education, University of Oxford, offers its Advanced Quantitative Methods Summer School, this year consisting of nine different on-line courses running during three calendar weeks in May (in Oxford parlance these are called weeks 2, 3 and 4 of Trinity Term; 4-21/5/2021). 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 this portal https://www.oxforduniversitystores.co.uk/conferences-and-events/department-of-education/events/advanced-quantitative-methods-summer-school-2021.

The courses will take place online, and include a mixture of pre-recorded lectures and demonstrations, synchronous lectures, workshops and questions & answer sessions. Synchronous session sessions will run in zoom. Links to materials and zoom-sessions will be sent to participants. You will use R-studio and various R-modules, and Mplus (Mplus demo) during the courses. The teaching format will vary slightly from course to course (please see details for each course below)

Our program is:

Oxford Week 2		
Tue 4/5	Thees Spreckelsen	Intro to R
Wed 5/5	Thees Spreckelsen	Data management and documenting in R
Thu 6/5	Thees Spreckelsen	Generalized Linear Models (GLM) in R
Oxford Week 3		
Mon 10/5	Luning Sun	Intro to IRT modelling
Wed 12/5	Kit Double	Multilevel modelling
Thu 13/5 - Fri 14/5	Prathiba Natesan Bately	Simulations in R
Oxford Week 4		
Mon 17/5	Lars Malmberg	Intro Structural Equation Models (SEM)
Wed 19/5	Lars Malmberg	Longitudinal SEM
Fri 21/5	Lars Malmberg	Multilevel and intraindividual SEM

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

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

Course 1: Introduction to R for the Analysis of Educational Data
Instructor: Thees F Spreckelsen
[https://www.gla.ac.uk/schools/socialpolitical/staff/theesspreckelsen/ **]**
Materials: on-line materials made available in advance of the session
Synchronous session: Tue 4 May 2021, 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 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: 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: this video tutorial shows how to: https://learnr-examples.shinyapps.io (start by clicking “continue”).

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 (£60 / £250 / £450 for the week).

Course 2. Data management, graphing and documenting in R
Instructor: Thees F Spreckelsen
[https://www.gla.ac.uk/schools/socialpolitical/staff/theesspreckelsen/]
Materials: on-line materials made available in advance of the session
Synchronous session: Wed 5 May 2021, 10:00 to 15.30, UK-time, via zoom

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: [check time-table and 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 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 (£60 / £250 / £450 for the week).

Course 3. Generalized Linear Models (GLM) in R
Instructor: Thees F Spreckelsen
[https://www.gla.ac.uk/schools/socialpolitical/staff/theesspreckelsen/]
Materials: on-line materials made available in advance of the session
Synchronous session: Thurs 6 May 2021, 10:00 to 16:00, UK-time, via zoom

Yes/No questions, Agreed-Disagree scales, and categories like ethnicity are essential to social science data. And social scientists ask what increases the probability for e.g. a YES!- answer. Using the European Social Survey this course will introduce participants to three common models for these outcomes: Logit, Ordered Logit, and Multinomial models. Each model will be introduced in turn with a focus on running these in R and interpreting the results. Following that alternatives, relating to violations of assumptions will be briefly discussed.

Aims: To enable participants to: 1) model binary, ordered, and categorical outcomes with generalized linear regressions using R. 2) interpret the regression results from these models; 3) select from alternative modelling approaches; 4) plan and conduct their own analyses.

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 *statistical methods* up to Ordinary Least Squares (OLS) regression analysis.
- Participants should have *used R/RStudio* previously, there will only be a minimal introduction.

Self-check:

- What does the following function do: "`lm(z ~ x1+x2, data= df)`"?
- What does this formula stand for: $y = b_0 + b_1 x_1 + e$

Format: The course is taught in four sessions via Zoom (see below). Each session starts with an *intuition & brief demonstration*, followed by a first *get-your-hands-dirty R exercise*, followed by a more *in-depth what is going on* part, and finally another *check-your understanding exercise.*

Contents: [check time-table and contents]
10:00 – 11:00 When OLS fails – modelling binary variables
11:00 – 11:15 Break
11:15 – 12:30 Interpreting results: Predicted probabilities & Visualizations.
12:30 – 13:00 Lunch
13:00 – 14:30 Modelling ordered and categorical outcomes
14:30 – 14:45 Break
14:45 – 16:00 When thing go wrong: Model checks, alternatives, and where to go next..

Cost: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£60 / £250 / £450 for the week).

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

Course 4. Introduction to IRT modelling
Instructor: Luning Sun [https://www.jbs.cam.ac.uk/faculty-research/research-teaching-staff/luning-sun/]
Materials will be made available in advance via dropbox
Synchronous session: Mon 10 May, 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.

Time-table

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: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£60 / £250 / £450 for the week).

Course 5. Multilevel Modelling
Instructor: Kit Double [kitdouble.com]
Pre-recorded videos and on-line materials made available in advance of the session
Synchronous session: Wed 12 May, 9:00 – 11:00, UK-time, via zoom

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

- Overview of MLM – Models Types and Research Questions
- Running MLM in R
- Addressing research questions with MLM in R
- Growth Curve MLM in R
- A Q&A session will be held on 12 May at 9:00 (UK time)

Cost: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£60 / £250 / £450 for the week).

Course 6. Statistical Simulation: The best way to teach and learn statistics

Instructor: Prathiba Natesan Batley, https://www.brunel.ac.uk/people/prathiba-natesanbatley
Materials: on-line materials made available in advance of the session
Synchronous session: Thur 13 May **and** Fri 14 May 2021, 10:00 to 15.00 UK-time, via zoom

From quantum mechanics to financial planning, simulation is ubiquitous in today's computationally intensive world. Statistical simulation typically involves formulating a model (e.g. linear regression model), generating random data with specific statistical properties, estimating the parameters of that model with the 'simulated' data, and investigating how close the sample statistics are to population parameters. Since the population parameters are known, such comparisons help us understand the accuracy and optimal conditions for these models in real world applications. In essence, we can use simulations to understand the extent to which we can push the boundaries of statistical modelling. Personally, I have found simulation helpful in learning and teaching statistical concepts and in testing the performance of the newer models I build for specific real-world problems. This statistical simulation course will cover theory, planning, summarizing simulations, and estimating the adequacy of simulations. The course will be hands-on with some worked examples and discussions. R will be used.

Pre-requirements
Participants will need to know basics of univariate statistics and general linear models. Working knowledge of R is preferred although not required. The course will have a brief introduction to R.

Contents and time-table
Day 1
10:00-11:00: Introduction to simulation; Brief introduction to R
11:00-12:30: Generating random discrete and continuous data; Distribution types
12:30-12:45: Break
12:45-14:00: Playing with familiar concepts – sampling distribution
14:00-14:30: Determining simulation conditions
14:30-15:00: On the adequacy of simulation and Q&A

Day 2
10:00-10:30: On the adequacy of simulation continued…
10:30-11:30: Simulating General Linear Models
11:30-11:45: Break
11:45-12:45: Summarizing simulation/common diagnostics in simulation/Adequacy of simulation
12:45-13:45: Coverage rates and confidence intervals/Graphics for simulation
13:45-14:45: Simulating multivariate distributions and time-series models
14:45-15:00: Q&A

Cost: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£60 / £250 / £450 for the week).

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

Course 7. Introduction to Structural Equation Modelling
Instructor: Lars Malmberg, (http://www.education.ox.ac.uk/people/lars-erik-malmberg/)
Pre-recorded videos and materials made available in advance of the session
Synchronous session: Monday 17 May 9:00 to 12.00, UK-time, via zoom

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.

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 17 May
09.00-10.00 Questions and answers session based on pre-recorded materials (questions can be posted in advance in the Q&A document
10.00-11.00 Group-work in virtual breakout rooms
11.00-12.00 Further questions and answer session, also about own data

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

Software: the mplus demo (https://www.statmodel.com/demo.shtml)
the R package lavaan (http://lavaan.org)

Cost: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£60 / £250 / £450 for the week).

Course 8. Structural Equation Modelling of longitudinal data
Instructor: Lars Malmberg, (http://www.education.ox.ac.uk/people/lars-erik-malmberg/)
Pre-recorded videos and materials made available in advance of the session
Synchronous session: Wedn 19 May 9:00 to 12.00, UK-time, via zoom

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 19 May
09.00-10.00 Questions and answers session based on pre-recorded materials (questions can be posted in advance in the Q&A document
10.00-11.00 Group-work in virtual breakout rooms
11.00-12.00 Further questions and answer session, also about own data

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

Software: the mplus demo (https://www.statmodel.com/demo.shtml)
the R package lavaan (http://lavaan.org)

Cost: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£60 / £250 / £450 for the week).

Course 9. Structural Equation Modelling for intraindividual data
Instructor: Lars Malmberg, (http://www.education.ox.ac.uk/people/lars-erik-malmberg/)
Pre-recorded videos and materials made available in advance of the session
Synchronous session: Friday 21 May 9:00 to 12.00, UK-time, via zoom

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 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 dairy 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 19 May
09.00-10.00 Questions and answers session based on pre-recorded materials (questions can be posted in advance in the Q&A document
10.00-11.00 Group-work in virtual breakout rooms
11.00-12.00 Further questions and answer session, also about own data

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

Software: the mplus demo (https://www.statmodel.com/demo.shtml)
the R package lavaan (http://lavaan.org)

Reference
Malmberg, L-E. (2020). Intraindividual structural equation models for learning experiences. International Journal of Research & Methods in Education, 43 (4),413-430. 10.1080/1743727X.2020.1793939

Cost: £25 for OU and Grand Union students, £100 for staff and external students, and £200 for professionals (£60 / £250 / £450 for the week).