



Course Code & Title	LISS242 Data Management & Manipulation		
Convenor(s)	Dr Ludovico Carrino (Ludovico.carrino@kcl.ac.uk)		
Institution	King's College London	Department	LISS DTP, Department of Global Health and Social Medicine
Academic Year	2021-22	Term	Summer
Number of Sessions	5	Length of Session(s)	5-6 hours
Day, Date		Start : End	Room Location
Monday 23 May 2022 Tuesday 24 May 2022 Wednesday 25 May 2022 Thursday 26 May 2022 Friday 27 May 2022		10:00-15:00	KINGS BLDG K4.33 (SCR) – King's College London Strand Campus *MEETS IN PERSON*
Enrolment Link:	https://bit.ly/LISS242 You will be prompted to log into SkillsForge		

Course Description:

This course will develop researchers' knowledge and use of large social science datasets. Researchers will be able to understand how complex social science large scale datasets are structured and critically engage with the implications of this. They will develop the skills to manipulate, recode and compute variables and learn how to deal with missing data. They will also learn how to combine datasets, and aggregate and disaggregate data from different files in a relational database, as well as how to transform the structure of datasets from long form to short form and vice versa. Researchers will gain practical experience and skills of how to manipulate complex datasets to answer questions of importance in the context of health and social research.

- This course is suitable for students with basic statistical knowledge. Prior to entry students must be able to demonstrate:
- They understand the concepts behind basic descriptive and inferential statistics for social science research, and how to interpret these statistics.
- They understand the principles of correlation, t-tests, chi-square and regression in social science research, and how to interpret these tests.
- Competence in a basic level of data manipulation using STATA to prepare social science datasets for statistical analysis such as designation of missing data, transformations of data, and conditional recode and compute commands.
- Competence in the use of statistical software STATA to interrogate social science datasets using descriptive statistics and commonly used statistical tests.



London Interdisciplinary Social Science Doctoral Training Partnership Advanced Research Methods in Social Sciences

Course Outline:

On completion of this course, students will:

- Be able to source large social science datasets and associated documentation
- Understand how complex social science large scale datasets are structured and the implications of this
- Be able to manipulate, recode and compute variables and understand how missing data can be dealt with
- Learn how to combine datasets, and aggregate and disaggregate data from different files in a relational database
- Learn how to transform the structure of datasets from long form to short form and vice versa
- Learn complex computation of derived variables including do loops
- Have practical experience of the manipulation of complex datasets to answer questions of importance in the context of health and social research

Eligibility:

As prerequisites to registering on this module, doctoral students MUST already be able to:

- Understand the concepts behind basic descriptive and inferential statistics for social science research, and how to interpret these statistics.
- Understand the principles of correlation, t-tests, chi-square and regression in social science research, and how to interpret these tests.
- Demonstrate competence in a basic level of data manipulation using STATA to prepare social science datasets for statistical analysis such as designation of missing data, transformations of data, and conditional recode and compute commands.
- Demonstrate competence in the use of statistical software STATA to interrogate social science datasets using descriptive statistics and commonly used statistical tests.

Pre course preparation:

See eligibility criteria above!

Number of students:

10