

Section 1 - Introduction and overview

Longitudinal data analysis comprises a set of techniques used to explore variables that have been measured at different time points. This section will introduce some of the important fundamentals to such analysis and describe the types of analysis methods covered in this module.

What is longitudinal data analysis?

Longitudinal data analysis (LDA) comprises a set of techniques that can be used to analyse data where the same variable or variables have been measured at different time points. Throughout the life course there are a wide variety of subjects, including physical and mental health, and socio-economic and behavioural factors, with which we can get important insights by looking at them using repeated measures.

The utility of longitudinal studies and their differences to cross-sectional designs are described in the Learning Hub's Introduction to Longitudinal Studies (<https://learning.closer.ac.uk/introduction/types-of-longitudinal-research/>). LDA methods allow us to make use of the rich data collected by longitudinal studies and to explore the temporal relationships between measures collected across different life stages. There are a number of such methods, each suited to the analysis of different types and combinations of variables. Some variables are continuous (e.g. age) and others are categorical (e.g. a list of occupations). We call categorical variables with two levels 'dichotomous' (e.g. deceased or living) and where they are coded as 0 or 1, we can also call them 'binary'. This guide will teach you about different analytic approaches to exploring how certain types of outcomes are associated with potential explanatory factors.

Dissimilar outcomes can occur even among people who share the same characteristics. The term 'heterogeneity' is often used to refer to differences like these. Longitudinal data can help control for such differences by including a wide range of explanatory variables across the life course in statistical models. The problem of 'omitted variable bias' is also improved by longitudinal data analysis, but always remains, as there are connections between the outcome and explanatory variables that have not or could not be included as they are unmeasurable.

We will use an extract from the National Child Development Study (NCDS) CLOSER Training Dataset (CTD) to illustrate some of the different methods that can be used in analysing longitudinal data. The NCDS is a cohort study of people born in England, Scotland and Wales during a single week of 1958. In the NCDS, detailed information has been collected on participants from childhood, through adolescence into early adulthood and later life, allowing us to look at different outcomes and potential explanatory variables. Measurements that have been collected over time include assessments of physical health (e.g. Body Mass Index (BMI) measured at ages 7, 11, 16, 23, 33, 42 and 50), as well as a series of mental health (e.g. Malaise inventory), socio-economic position, and behavioural factors (e.g. smoking) measured at ages 23, 33, 42, and 50. These measures are examples of the variety of data available in the NCDS and other longitudinal studies.

Overview of this guide

In the following sections, we will present a variety of longitudinal data techniques you can apply to longitudinal data and repeated measures. First, we will explore and prepare the dataset before demonstrating how to apply general linear, logistic and multinomial regression approaches which are commonly used in the analysis of longitudinal study data. In future updates to this module, we will also illustrate how to transfer data to a format suitable for repeated measures analysis. We will also be adding guidance on techniques for analysing such repeated measures data, including multilevel regression, fixed effects, and latent growth models.

We will guide you through these methods as performed in the STATA statistical software package, and we will provide documented syntax to explain the steps involved. Guidance for other statistical software packages is forthcoming.