

1 Introduction to LMEs

In modern statistical analysis, data sets have very complex structures, such as clustered data, repeated data and hierarchical data (henceforth referred to as grouped data).

Repeated data considers various observations periodically taken from the same subjects. ‘Before and after’ measurements, as used in paired t tests, are a well known example of repeated measurements. Clustered data is simply the grouping of observations according to common characteristics. For example, an study of pupils of a school would account for the fact that they are grouped according to their classes.

Hierarchical structures organize data into a tree-like structure, i.e. groups within groups. Using the previous example, the pupils would be categorized according to their years (i.e the parent group) and then their classes (i.e the child group). This can be extended again to multiple schools, where each school would be the parent group of each year.

An important feature of such data sets is that there is correlation between observations within each of the groups (?). Observations in different groups may be independent, but any assumption that these observations within the same group are independent is inappropriate. Consequently ? states that there is two sources of variations to be considered, ‘within groups’ and ‘between groups’.