**Longitudinal Data**

Longitudinal data (also referred to as panel data) track the same sample at different points in time. As opposed to repeated cross-sectional data, which refers to data from the same survey given to different samples over time (also providing long term data). Longitudinal studies track the same people over time, so that observed differences are less likely to be the result of cultural generation differences etc. In the context of advertising, it is used to observe the changes that advertising has produced in the attitudes of a target audience. Generally, longitudinal data are typically collected via fewer measurements and over a larger number of subjects, in comparison with time series data. Regarding the latter, seasonality is typically of more interest than the former.

**Statistical modelling of longitudinal data:**

Statistical analysis requires methods that can properly account for the correlation between the intra-subject response measurements. Time-varying covariates should also be properly accounted for, as there may be complicated feedback between outcome and exposure, depending on the study.

Two popular approaches consist of *subject-specific* and *population-averaged* models. For the former, individual behaviour is modelled according to a stochastic time series model (taking autocorrelation into account), which is then averaged over all subjects. The individual model contains terms describing how the subject deviates across the sample. Conversely, in the population-averaged approach, the population average profiles are modelled directly. A stochastic time series model is applied to the sample mean behaviour. For linear models, the subject specific population average and the direct population average have the same form. This is not the case for nonlinear models. The most appropriate approach depends on the subject matter and objective of the analysis.

**In Rasch analysis (from rasch.org):**

When dealing with longitudinal data, constant item difficulties are often required over all time points. To ensure this, and to assess person ability change, constant *anchor* values are used to fix item difficulties over the time points. In defining the anchor values, the data is stacked in a way that each column responds to each item, and each row corresponds to each person at each time set. The item difficulties can be calculated from this stacked data, but the Rasch assumption of local independence is likely to be violated (due to the intra-person covariance). Another way to define anchor points is to assign the item difficulties for one of the time points. This may be appropriate, depending on the problem under consideration (e.g. in healthcare, when treatment decisions are made at the initial time). However, in other problems, the item difficulties at one time may not be suitable for the other time points. Alternatively, the anchors may be defined according to a random data set consisting of patients across each time-point, where each patient is only in the data set once and the time points are equally represented. This takes the approach that the data difficulties at each time are of equal importance.

In some cases, it may be of interest to assess how item difficulty changes over time, for persons that do not change. In this case, the data can be racked – each person is included once in the data set, with each item for each time point (the same item over each time can be thought of as distinct, yet linked items). In racked analyses, item difficulties are of greater interest than person ability. The same information regarding changes in item difficulty can also be assessed via an item Differential Item Functioning (DIF) on the person samples at the different times. In standard DIF, each item is investigated for signs of interaction between sample characteristics (e.g. age group, gender, ethnicity).

**In classical test theory:**

In classical test theory (CTT), the observed score for an individual is assumed to be equal to the true score plus an error of measurement. It is assumed that systematic effects between responses of the subjects/persons are solely due to their variation in ability. Other sources of variation (internal and external conditions) are assumed to be constant through rigorous standardization or be random in effect by nature. The standard deviation of the distribution of errors for each individual reveals the measurement error magnitude. Typically, the random error distribution is assumed to be the same for each individual. In CTT, every test is sample dependent, and item difficulty and ability score are therefore sample dependent. For longitudinal data, a test score (i.e. sum-score) is calculated for an individual at a certain time by summing over the total number of item responses. Sum-scores are population dependent and have an assumed constant standard error for all subjects (although some sum-scores are more informative than others – different subjects are likely to be measured with different precisions).

Note to self: I don’t think that there is any particularly specific treatment for longitudinal data in CTT. The sum-scores are simply analysed at each time point.