Click on 'Bookmarks' in the Left-Hand menu and then click on the required section.

8.6 Data Transformations

8.6.1 Introduction

One limitation on the use of many parametric analytical techniques is the requirement that the data is drawn from a population with a **normal** (or nearly normal) distribution.

For a number of **non-normal** distributions it is possible to **transform** the data by applying a mathematical transformation to each data value. The resulting, transformed, data may be more closely related to a normal distribution.

For example a positively skewed distribution with a long positive tail may be transformed into a near normal distribution by taking the logarithm of the data values.

Note however, that any **uncertainty ranges** will also be transformed and are unlikely to remain symmetrical about any best-estimate value.

8.6.2 Logarithmic Transformation

Each data value, x, is replaced the value, x', where:

$$x' = \log(x + c)$$

where c is a constant depending on the range of the original data.

Different values of c will be appropriate for distributions with different mean value, and should be chosen such that all values, $(x + c)$, are positive.

Logarithmic transformations can also be used for data where the standard deviation of sample sets is proportional to the mean value of the set. For example, this can be used for such data to satisfy the ANOVA requirement that the variances of sample sets should be similar.

8.6.3 Square Root Transformation

Each data value, x, is replaced the value, x', where:

$$x' = \sqrt{x}$$

This transformation is useful for distributions where the sample variance is proportional to the sample mean, e.g. with **Poisson distributions**.

8.6.4 ArcSine Square Root Transformation

This transformation is applied for proportions, *p*, between the value of 0 and 1, where the data is likely to follow a binomial distribution.
For sample sizes, *n*, and proportions, *p*, where $np(1-p) >> 1$ (see p239), the binomial distribution already approximates to a normal distribution.
For samples, with high or low values of *p*, that do not satisfy the above requirement, it is possible to use the transformation:

$$p' = \sin^{-1}(\sqrt{p}) = \arcsin(\sqrt{p})$$