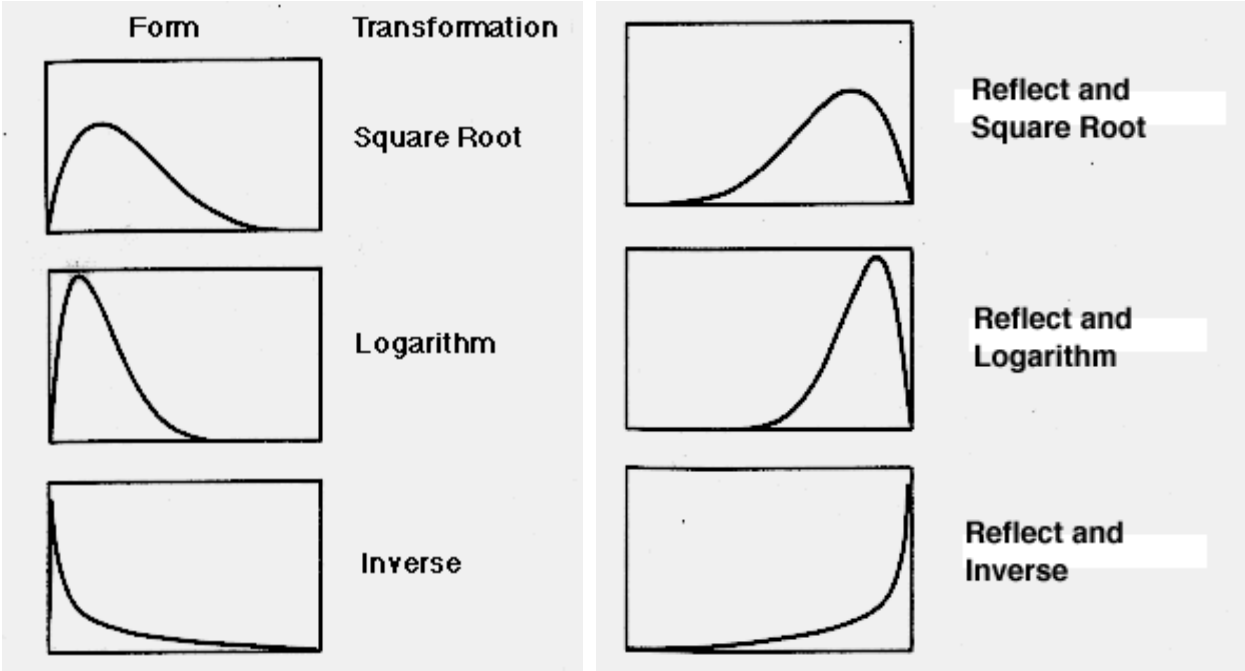
Data Transformations –

The most frequent reason that researchers transform their data is to make the distribution of the data "normal", and thus fufill one of the assumptions of conducting a parametric means comparison. Other reasons include more informative graphs of the data, better outlier identification (or getting outliers in line) and increasing the sensitivity of statistical tests. Indeed, parametric tests are often preferred over non-parametric tests for many types of data sets. (See Seth Roberts' article "Transform your data", Nutrition (2008) 24:492-494 at: http://sethroberts.net/articles/2008%20Transform%20your%20data.pdf)

Transformations involved applying a mathematical function (e.g. squaring the data) to each data point. A transformation is needed when the data is excessively skewed positively or negatively. The figure below suggests that type of transformation that can be applied depending upon the degree of skewness.

Positively skewed data | Negatively skewed data



Data that is negatively skewed requires a reflected transformation. This means that each data point must be reflected, and then transformed. To reflect a variable, create a new variable where the original value of the variable is subtracted from a constant. The constant is calculated by adding 1 to the largest value of the original variable. (E.g. (Largest value n_L +1) – (original value n_x). Next, transform the reflected data set.

Here are some short comments on the most common data transformations.

Logarithms : Growth rates are often exponential and log transforms will often normalize them. Log transforms are particularly appropriate if the variance increases with the mean.

Reciprocal : If a log transform does not normalize your data you could try a reciprocal (1/x) transformation. This is often used for enzyme reaction rate data.

Square root : This transform is often of value when the data are counts, e.g. blood cells on a haemocytometer or woodlice in a garden. Carrying out a square root transform will convert data with a Poisson distribution to a normal distribution.

Arcsine : This transformation is also known as the angular transformation and is especially useful for percentages and proportions which are not normally distributed.

As suggested by Tabachnick and Fidell (2007) and Howell (2007), the following guidelines should be used when transforming data.

If your data distribution is…	Try this transformation method.
Moderately positive skewness	Square-Root NEWX = SQRT(X)
Substantially positive skewness	Logarithmic (Log 10) NEWX = LG10(X)
Substantially positive skewness (with zero values)	Logarithmic (Log 10) NEWX = LG10(X + C)
Moderately negative skewness	Square-Root NEWX = SQRT(K – X)
Substantially negative skewness	Logarithmic (Log 10) NEWX = LG10(K – X)

C = a constant added to each score so that the smallest score is 1.
K = a constant from which

Howell, D. C. (2007). *Statistical methods for psychology* (6th ed.). Belmont, CA: Thomson Wadsworth.
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn and Bacon.