Notes on the use of Data Transformations[1]

Although the tradition is to assess normality by "eyeballing" the data, for example, examining a histogram of the distribution of scores, these are generally inadequate. It is difficult to assess normality by simply viewing a histogram. A better display of normality includes a Q-Q plot of the data – comparing the observed distribution with an expected distribution if the data were normally distributed. There are a number of statistical tests that can also assess the degree of departure from normality, including the Kolmorogov-Smirnov test, Goodman test, and others.

There are several transformations that can be employed to obtain normally distributed data for the purposes of securing stable and less biased estimates. Three common transformations include (1) square root transformation, (2) log transformation, and (3) inverse transformation. These are briefly described here. These transformations work with positively skewed data – they bring in the right tail. They can be applied to negatively skewed data once the distribution is "reflected", a process where each value is multiplied by –1 and a constant is added to push the distribution above 1.0 on the scale. There are strong statistical arguments for "anchoring" the distribution at 1.0, where the lowest observed value is "transformed" in such a way to make it 1.0.

These transformations work in a way that brings in the skewed tail of a non-normal distribution. It changes the distances between observed values to create a distribution that is closer to normal.

Square root transformation

A square root transformation takes the square root value of the observed data. This should only be used on data that have values greater than 1.0, since the square root of values greater than 1.0 become smaller while values between 0 and 1.0 become larger – and the square root of negative values is undefined.

Log transformation

There are any number of logarithmic transformations that can be employed. Common base values include 10 or 2 or the natural log with constant e (2.7182818) as the base. Taking the log of a value tends to pull in extreme values. Log transformation with a larger base pull extreme values in to a greater extent.

Inverse transformation

Taking the inverse of a value is simply computing 1/x. This makes very small numbers large and very large numbers very small. It will also reverse the order of the observations on the scale. Since this reorders observations, the data should probably be reflected prior to the transformation to maintain the original ordering of the values.

[1] Osborne, Jason (2002). Notes on the use of data transformations. *Practical Assessment, Research & Evaluation*, 8(6). Available online: http://ericae.net/pare/getvn.asp?v=8&n=6.