

1 Transformations

One of the simplest ways to use certain statistical procedures with data that are not normally distributed is to transform the data. The advantage of transformations is the increase in statistical power from using parametric statistics over nonparametric statistics. The disadvantage is the difficulty in interpretation that can sometimes accompany the transformation—it is much easier to think in terms of mg/dL than it is in terms of $\log(\text{mg/dL})$. Below we outline the more common transformations applied to data.

1.1 Transforming Percents, Proportions and Probabilities

The two most common methods for transforming percents, proportions, and probabilities are the *arcsine* transform and the *logit* transform. In both cases, percentages should first be changed to proportions by dividing the percentage by 100. Note that these transformations are applicable only to percentages that lie between 0 and 100. They should not be used in the case of “percent increase” which can give values greater than 100%.

When should these transformations be applied? The usual rule of thumb is that they should be used when there are a number of proportions close to 0 and/or close to 1. The transformations will “stretch out” proportions that are close to 0 and 1 and “compress” proportions close to 0.5.

1.1.1 Arcsine Transform

Sometimes called an *angular transformation*, the arcsine transform equals the inverse sine of the square root of the proportion or

$$Y = \arcsine \sqrt{p} = \sin^{-1} \sqrt{p}$$

where p is the proportion and Y is the result of the transformation. The result may be expressed either in degrees or radians. Table X.X gives a series of percentages along with the arcsine transforms of those percentages.

Table X.X. Examples of the arcsine and probit transformation for percents. A period (.) denotes that a value is undefined.

Percent	Arcsine Transform		Logit Transform
	Degrees	Radians	
0	0.000	0.000	.
5	12.921	.226	-2.944
10	18.435	.322	-2.197
20	26.565	.463	-1.386
30	33.211	.600	-0.847
40	39.232	.685	-0.405
50	45.000	.785	0
60	50.768	.886	0.405
70	56.789	.991	0.847
80	63.435	1.107	1.386
90	71.565	1.249	2.197
95	77.079	1.345	2.944
100	90.000	1.571	.

1.1.2 Logit Transform

A *logit* is defined as the *logarithm of the odds*. If p is the probability of an event, then $(1 - p)$ is the probability of not observing the event, and the odds of the event are $p/(1 - p)$. Hence, the logit is

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right).$$

The logit transform is most frequently used in logistic regression and for fitting linear models to categorical data (log-linear models). Note that the logit is undefined when $p = 0$ or $p = 1.0$. This is not a problem with either of the two above-named techniques because the logit transformation is applied to a *predicted* probability which can be shown to always be greater than 0 and less than 1.0. Table X.X also gives the logit transform for a series of percents.

1.2 Transforming Ratio Scales

Several different transformations are available for ratio scales.

1.2.1 Square Root Transformation

The square root transformation is simply $Y = \sqrt{X}$, although many statisticians recommend the transformation $Y = \sqrt{X + 0.5}$, especially when the variable has one or more 0s. It is often used for counts and for other measures where group means are correlated with within group variances. The square root is often encountered in biology because many biological variables—especially counts—follow a Poisson distribution

within groups. Because the mean of a Poisson variable equals the variance of the variable, group means will always be correlated with within-group variances in this case.

1.2.2 Log Transform

A metabolic product is often the result of several different steps, each of which may involve competitive binding. The typical mathematical process is essentially multiplicative, giving rise to a lognormal distribution. The easiest way to transform such data is to take the logarithm, giving $Y = \log(X)$. Because variable X may legitimately be 0, the transform $Y = \log(X + 1)$ is recommended over the simple logarithm. Any base for the logarithm can be used, but base 10 is often used because of interpretability—a difference of 1 unit in a $\log_{10}$ transform denotes a 10-fold increase (or decrease).

1.2.3 The Power Transform

A power transform is also called a *Box-Cox* transform after the two statisticians who developed it (Box and Cox, 1964). The mathematical equation for the transform is

$$Y = \frac{(X + c)^\lambda - 1}{\lambda}, \lambda \neq 0$$

$$Y = \log(X + c), \lambda = 0$$

Here, c is an arbitrary constant chosen so that all scores (i.e., $X + c$) are greater than 0. The value of λ used in this equation is the one that transforms the data closest to normality and must be found using computer algorithms.

1.2.4 “Normalizing” Transforms

We have seen how to find the area under the normal curve between negative infinity and a particular score. What we call “normalizing” transforms do just the opposite—they start with areas under the normal distribution and then find the particular score that corresponds to that area¹. The first step in these transformations is to calculate percentile scores for the variable that is to be transformed. The percentiles are then treated as areas under a normal curve and the appropriate score is then assigned. For example, let us take an observation that it is at the 5th percentile. The score in a standard normal distribution that divides the bottom 5% from the top 95% of the distribution is -1.645 . Hence, that observation is assigned the value of -1.645 . Despite their name, normalizing transformations do not always guarantee a normal distribution.

¹ There is no conventional term for these types of transforms. We call them normalizing because they are based on the mathematics behind the normal curve.