5 Transformation

The numbers in a sample are assigned to objects by a set of rules. For instance, in the United States the fuel efficiency of automobiles is commonly measured in miles per gallon. When the United States switches to the metric system, kilometers per liter will be used instead. Because one mile equals 1.61 kilometers and one gallon equals 3.79 liters, one mile per gallon equals 1.61/3.79 or .43 kilometer per liter. So if a car gets 22 miles per gallon, it gets $22 \times .43$ or 9.5 kilometers per liter. What has been done is take a number and transform it. *Transformation* is the process by which numbers in a sample are altered by some mathematical operation. So if a sample of numbers were measured in terms of miles per gallon, and there was a need to remeasure fuel efficiency in terms of kilometers per liter, the values of the numbers could be systematically changed by multiplying them by .43. Numbers are not immutable. The rules that are used to assign numbers to objects can be changed, and possible transformation of the numbers should be considered.

Transformations of data are more common than might be realized. Most of us received a gross Scholastic Aptitude Test (SAT) score, say 580 in verbal and 560 in math. But attached to each is a percentile rank, say 76% and 68%. A percentile rank is a transformation of the raw SAT score. Also, when a person compares his or her test score to a friend's test score, the person might subtract his or her own score from the friend's score. The person then knows how much better or worse is the friend's score. Another commonly used transformation is to rank order the sample of numbers. Typically the times from a running race are rank ordered from fastest to slowest.

A *transformation* takes the original numbers and performs a mathematical operation on them. A transformation can be represented by

some mathematical operation on the original sample of numbers
= transformed sample of numbers

The mathematical operation can be as simple as addition or as complex as some trigonometric function, but the principle remains the same.

Transformations have three major purposes: to increase *interpretability,* *comparability*, and *symmetry*. Numbers can be confusing, and so if a transformation makes the meaning of the numbers any clearer, they should be transformed. For instance, if it is known that a consultant made $15,000 in 20 weeks, it is more informative to know the consultant's rate of pay per hour. It is $15,000 divided by the number of hours worked. The number of hours is 20 $\times$ 5 $\times$ 8, or 800, and so the hourly rate is 15,000/800, which is $18.75. By transforming the number from dollars per 20 weeks to dollars per hour, it is clearer how much the person made. Transformations are also used to make samples of numbers more comparable than they would be otherwise. One reason for converting to the metric system is to make numbers in the United States comparable with those in the rest of the world.

Finally, transformations are used to make distributions more symmetric than they would be otherwise. Symmetry in a distribution is desired because it is assumed by many inferential statistical techniques. Moreover, experience shows that asymmetric distributions sometimes are not at the interval level of measurement. Transformations that promote symmetry tend to increase the likelihood that the level of measurement is interval in nature.

Data transformations are a basic part of data analysis. There are four major types of transformations: no-stretch, one-stretch, two-stretch, and flat transformations. This classification system focuses on the effect of each type of transformation on the shape of the distribution. Each is considered in turn.

One example will be used to illustrate the various transformations. It is taken from Duncan and Fiske's (1977) study of two-person interactions. Two strangers interacted for a period of seven minutes, the last five of which were videotaped. In Table 5.1 are the number of smiles during the five minutes of the interaction. The numbers in the table were based on measurements from 22 women, graduate students at the University of Chicago, who interacted with 22 men. The mean number of smiles for the women is 9.55 and the standard deviation is 4.77.

TABLE 5.1 **Number of Smiles of 22 Females During Five Minutes of Interaction with a Male**

8	13
4	10
18	4
11	7
5	11
7	9
5	9
21	7
14	14
1	9
9	14

Data were gathered by Duncan and Fiske (1977).

No-Stretch Transformations

A no-stretch transformation is the simplest and most common type. Most of the time that this transformation is used, the researcher is not even aware of it. When quarts are converted to gallons or feet are converted to miles, a no-stretch transformation (sometimes called a linear transformation) is performed. A *no-stretch transformation* does not alter the basic shape of the distribution. Although the distribution is stretched, it is uniformly stretched.

This transformation is called a no-stretch transformation for the following reason: The effect of this type of transformation is not to alter the basic shape of the distribution. So, if the distribution is symmetric it remains symmetric. If there is a positive skew it remains. If the distribution is flat it remains flat. Many of the statistics that are discussed in later chapters do not change after a no-stretch transformation. For instance, the correlation coefficient and t test do not change.

One type of no-stretch transformation is one in which all the numbers in the sample are multiplied or divided by the same number. If the score is denoted as X and the number that the score is multiplied by is denoted as k, the transformed score is kX. For example, if X is a score whose unit of measurement is gallons and the researcher wishes to convert to quarts, then k is set to four because four quarts equal one gallon. Alternatively each score can be divided by a constant: X divided by k. For instance, if X is a score whose unit of measurement is feet and the researcher wishes to convert to miles, then k is 1/5280 or .0001894 because one mile equals 5280 feet. As another example, the data in Table 5.1 could be divided by five. Because the numbers in the table refer to the number of smiles in five minutes, dividing by five would yield smiles per minute.

Another type of no-stretch transformation is the adding of the same number to all the scores. If the numbers are denoted as X and the number added to all the scores is denoted as m, the transformed score is $X + m$. For instance, if X is a student's score on a statistics test and m is a five-point bonus that the teacher gives each student, then the transformed score is the previous score plus five bonus points. Alternatively, a constant may be subtracted from each score: $X - m$. For instance, if X is annual income earned by workers at a factory and m is $-10{,}000$, then the transformed score is a worker's earnings over $10,000. As another example, the value of 9.55 could be subtracted from the numbers in Table 5.1. Because 9.55 is the mean, a score minus the mean measures a given person's number of smiles relative to the mean. So a negative score would indicate that a person smiles less than average and a positive score indicates more smiling.

These two types of no-stretch transformations can be combined. That is, each score can be multiplied (or divided) by a number and then have another number added (or subtracted) to the score. If the score is denoted as X, the multiplier as k, and the score to be added as m, the transformed score is $m +$

kX. The temperature conversion from Celsius to Fahrenheit is of this form. The conversion for 20 degrees Celsius into Fahrenheit is $(1.8)(20) + 32$, which equals 68 degrees Fahrenheit. Thus, $k = 1.8$ and $m = 32$ for the conversion from Celsius to Fahrenheit.

Measures of central tendency and variability are altered in a no-stretch transformation. The three measures of central tendency are changed in the following way. If M is used to denote the measure of central tendency, then the new measure of location is $kM + m$, where k is the term that is multiplied and m is added. So if the mean is 10.4 and k and m are 2.0 and 1.0, respectively, then the transformed mean is $(2.0)(10.4) + 1.0 = 21.8$.

The variability is not affected by adding or subtracting a constant to the scores. For example, if ten is added to all the scores, the variability does not change. If scores are multiplied by k, however, the range, interquartile range, and the standard deviation of the transformed scores equal the measure of variability of the untransformed scores multiplied by k. The variance of the transformed scores equals the variance of the original scores multiplied by k^2.

There are four basic reasons for employing a no-stretch transformation: change in unit of measurement, change in scale limits, reversal of scale, and standardization.

Change in the Unit of Measurement

This purpose has already been discussed. Converting from inches to feet or from pounds to grams are examples of transformation to change the unit of measurement. This transformation is so common that is not even viewed as a transformation.

Change of Scale Limits

Imagine a set of scores with a possible range of from 100 to 500. It may be desirable to change the upper limit of 500 to 10 and the lower limit from 100 to 1. Thus, the transformed measure would range from 1 to 10. A little notation can help:

UL	upper limit of the original sample
LL	lower limit of the original sample
TUL	transformed upper limit of the transformed sample
TLL	transformed lower limit of the transformed sample

So, UL after transformation the becomes TUL and LL becomes TLL. To change the limits, one computes for each score X

$$(X - \text{LL})\frac{\text{TUL} - \text{TLL}}{\text{UL} - \text{LL}} + \text{TLL}$$

for the above example

$$\begin{aligned} \text{UL} &= 500 \\ \text{LL} &= 100 \\ \text{TUL} &= 10 \\ \text{TLL} &= 1 \end{aligned}$$

The appropriate transformation is then

$$(X - 100)\frac{10 - 1}{500 - 100} + 1$$

or

$$(X - 100)\frac{9}{400} + 1$$

For example, if X is 300, then transformed X equals 5.5.

Reversal

Sometimes scales are oriented in the "wrong direction." Generally larger numbers indicate more of some quantity. If this is not the case, the scale needs to be reversed. For instance, a variable may be the rating of a political leader on a ten-point scale—that is, a scale from one to ten. On some questions a response of ten is a favorable response toward the political leader and on others a one is a favorable response. It may be desirable to reverse the questions in which one is a favorable response. That is, make a response of one a ten and a ten a one. This transformation of the response X is accomplished in this example by $11 - X$. In general, the transformation for reversal is LL + UL $- X$, where LL is the lower limit, UL is the upper limit and X is the score to be transformed. So, the score to be transformed is subtracted from the sum of the lower and upper limits.

Standard Scores

The most often used no-stretch transformation in statistical work is one that is said to standardize the scores. To *standardize* a set of numbers, the mean is subtracted from each score, and this difference is divided by the standard deviation. The transformed score for person i is denoted by Z_i and is given by

$$Z_i = \frac{X_i - \bar{X}}{s}$$

where $\bar{X}$ is the mean of the X variable and s is the standard deviation. This score is called a *Z score* or a *standard score*. The effect of this transformation is to make the mean of the transformed scores equal zero and to make the standard deviation equal one. Researchers employ this transformation because a standard score (or Z score) tells them how far each person is from the mean in standard deviation units. For instance, a Z score of -2 indicates that the person scores two standard deviations below the mean. When the Z score transformation is used, the scores are said to be *standardized.*

For the smile numbers in Table 5.1, the mean is 9.55 and the standard deviation is 4.77. The formula for the Z score of variable X for person i is

$$Z_i = \frac{X_i - 9.55}{4.77}$$

So if $X = 14$, then $Z = (14 - 9.55)/4.77 = .933$, which is about one standard deviation above the mean. A positive Z score indicates that the score is above the mean, and a negative Z score indicates a score below the mean.

One-Stretch Transformations

A no-stretch transformation does not alter the basic character or the shape of the distribution. So if the shape of the distribution needs to be changed, a no-stretch transformation does not do the job. For instance, if there is a large positive skew (the scores trail off in the positive direction), it may be desirable to transform the numbers to remove that skew. One way to remove it would be to stretch the lower numbers. A method is needed to stretch one end of the distribution, and so the transformation is called *one-stretch*. The presence of positive skew is quite common when scores have a lower bound of zero and no upper bound. Examples of such variables are income, reaction time, age, and number of cars on a freeway. The three major one-stretch transformations are square root, logarithm, and reciprocal.

Square Root

The square root transformation, or $\sqrt{X}$, is relatively simple. A square root transformation is simply the square root of every number in the sample. One should employ this transformation only if all the numbers are positive. It is generally appropriate when X is a count. Examples of counts are the number of bar presses by laboratory rats or the number of cars passing through an intersection. This transformation has become easier to perform now because most hand-held calculators have a square root key.

Logarithm

The logarithm, or log, is the most commonly used one-stretch transformation, but it is the most complicated numerically. (See Chapter 1 for a review of logarithms.) For instance, income is regularly subjected to a logarithmic transformation. The logarithm of X is the number that satisfies the equation $X = 10^Y$. The term Y is said to be a common logarithm, base 10, of X. It is much more common in scientific work to use natural logarithms, which have e (approximately equal to 2.718) as the base. So Y is determined from

$$X = e^Y$$

To convert from base 10 to base e, multiply the base 10 logarithm by 2.303. It should be noted that $\log(1) = 0$ regardless of the base. Also the logarithm of zero and negative numbers is undefined regardless of the base. Thus, this transformation is only feasible when the numbers are positive. If zero is a possible value, it is necessary to compute $X + .5$ or $X + 1.0$ and then compute the logarithm. That is, .5 or 1.0 is added to all the numbers and then the logarithms are computed.

Reciprocal

The least commonly employed one-stretch transformation is the reciprocal transformation. The reciprocal of X is defined as $1/X$. In words, one is divided by the score. It is particularly useful when X is time. Say the number of minutes it takes to run a mile is measured. The reciprocal would measure how many miles or fractions thereof were run in a minute, or the rate at which one runs. Thus, the reciprocal of time equals rate, and the reciprocal of rate equals time. Unlike the square root and logarithm transformations, the reciprocal transformation reverses the order of scores, and thus the largest score becomes the smallest.

Of the one-stretch transformations, the square root stretches the least and the reciprocal stretches the most. To measure how much stretch there is in a one-stretch transformation, the amount of stretch in the lower numbers is compared with the amount of stretch in the higher numbers. To do this, the lower numbers are 1 and 5 and the higher numbers are 11 and 15. The stretch index is

$$\frac{\text{transformation}(5) - \text{transformation}(1)}{\text{transformation}(15) - \text{transformation}(11)}$$

so for instance, for reciprocal, the value is

$$\frac{1/5 - 1/1}{1/15 - 1/11}$$

Using this stretch index, the following values are obtained.

No transformation	1.00
Square root	2.22
Logarithm	5.19
Reciprocal	33.00

Although the square root exhibits some stretch, it stretches the least. The reciprocal transformation dramatically stretches the numbers.

One-stretch transformations are used primarily to remove positive skew.[1] To simplify the presentation, positive skew is indicated by the median being smaller than the mean. In practice, to determine the presence of positive skew a more detailed analysis is required. In particular, the histogram would have to be created and examined. One clue that a positively skewed distribution will be aided by a one-stretch transformation is the *coefficient of variation,* which equals $s/\bar{X}$, the ratio of the standard deviation to the mean. If the coefficient of variation is greater than .25, a one-stretch transformation is probably needed to remove skew.

To see how one-stretch transformations affect the shape of a distribution consider the following sample.

$$1, 4, 4, 9, 9, 9, 16, 16, 25.$$

This sample is positively skewed because the median is 9 and the mean is 10.33. However, after a square root transformation, the distribution becomes perfectly symmetric, with the mean and the median equal; that is,

$$1, 2, 2, 3, 3, 3, 4, 4, 5.$$

However, the logarithm transformation of the original numbers results in the following numbers:

$$0, 1.39, 1.39, 2.20, 2.20, 2.20, 2.77, 2.77, 3.22$$

and the skew has been overcorrected, as is indicated by the fact that the median (2.20) is now greater than the mean (2.02). There is no longer a positive skew but a slightly negative one. Thus researchers must be careful to avoid overcorrecting the skew.

In Table 5.2 are the numbers from Table 5.1 and their one-stretch transformed scores. There is a slight positive skew in the original scores, making the mean larger than the median. Applying the square root transformation makes the mean and median nearly equal. However, the log transformation overcorrects for skew and the reciprocal even more so. (Recall that the reciprocal transformation reverses the scores and so the relative size of the mean and median is reversed.) The square root transformation seems the most appropriate here.

The mode and median have an advantage over the mean when a one-stretch transformation is used. Assume that the numbers are each logged. The mean of the logs does not ordinarily equal the log of the mean of the untransformed

[1]One-stretch transformations have two other purposes. These transformations tend to remove heterogeneity of variance and nonlinearity. Heterogeneity of variance means that the variance changes for different samples. It is not uncommon for the standard deviation of a sample to be related to the mean of the sample. Typically, samples with larger means will have larger standard deviations. This is an example of heterogeneity of variance. One-stretch transformations generally bring about equal standard deviations. This topic is discussed in Chapter 13. Finally, in Chapter 7 the effect of one-stretch transformations on making relationships linear is considered.

TABLE 5.2 **One-Stretch Transformation**

	Raw Score	Square Root	Logarithm	Reciprocal
	8	2.828	2.079	.1250
	4	2.000	1.386	.2500
	18	4.243	2.890	.0556
	11	3.317	2.398	.0909
	5	2.236	1.609	.2000
	7	2.646	1.946	.1429
	5	2.236	1.609	.2000
	21	4.583	3.045	.0476
	14	3.742	2.639	.0714
	1	1.000	0.000	1.0000
	9	3.000	2.197	.1111
	13	3.606	2.565	.0769
	10	3.162	2.303	.1000
	4	2.000	1.386	.2500
	7	2.646	1.946	.1429
	11	3.317	2.398	.0909
	9	3.000	2.197	.1111
	9	3.000	2.197	.1111
	7	2.646	1.946	.1429
	14	3.742	2.639	.0714
	9	3.000	2.197	.1111
	14	3.742	2.639	.0714
Mean	9.55	2.986	2.100	.1625
Median	9.00	3.000	2.197	.1111

scores. However, the log of the mode equals the mode of the logs. The same holds for the median when sample size is odd and is closely approximated when sample size is even. So if the data are transformed by a one-stretch transformation, the mean must be recomputed, but the mode and median can be computed from the mode or median of the original data.

Before discussing two-stretch transformations, there is one other one-stretch transformation. Occasionally it happens that scores have a negative skew. For instance, the scores

$$0, 6, 6, 8, 8, 10$$

show a slight negative skew (the median, 7, being greater than the mean, 6.33), which is removed by squaring the numbers, as follows:

$$0, 36, 36, 64, 64, 100$$

Thus, squaring the numbers can remove a negative skew. What squaring the data does is to stretch the larger numbers so that the sample has less of a negative skew.

Two-Stretch Transformations

A one-stretch transformation is generally useful when the numbers have a lower bound (usually zero) but no upper bound. With a lower bound, numbers tend to pile up near it, creating a positive skew. What a one-stretch transformation does is to remove this lower bound.

A two-stretch transformation is useful for samples that have both a lower and an upper bound. There is one major type of data that has both an upper and lower bound: a proportion. A proportion has a lower bound of zero and an upper bound of one. (The student might review the discussion of proportions, percentages, and odds in Chapter 1.) One can score no higher than one and no lower than zero. The purpose of a two-stretch transformation is to stretch both the numbers near one and those near zero. It thus stretches twice.

Many times a researcher has proportions and does not realize it. For instance, if the variable for a test is the number correct out of 40, the numbers 16, 25, 38 do not look like proportions. But when the number of correct items on the test is divided by 40, the result is a proportion: 16/40 = .40.

The numbers near one and zero need stretching because a small change is more difficult when a proportion is near zero or one. For instance, reducing the risk of surgical procedure from .02 to .01 is more impressive than reducing it from .55 to .54. Although in both cases the risk has been reduced by 1%, the odds of dying have been cut in half in the .02 to .01 case and have hardly changed at all in the .55 to .54 case. Small differences between proportions near zero and one can be viewed as larger than small differences near .5.

The three types of two-stretch transformations are arcsin, probit, and logit. All three of these transformations are rather mathematically complicated; tables are given in Appendix A.

Arcsin

The arcsin is the most commonly used transformation for proportions in psychology. The sine is a trigonometric function and the arcsin is its inverse. Fortunately, one need know nothing about trigonometry to apply this transformation. The effect of this transformation is to stretch the distribution's tails.

Probit

The probit transformation is relatively uncommon in psychology but is very common in other sciences. For instance, political scientists routinely transform the proportion of persons voting for a candidate to a probit. To understand this transformation fully requires knowledge of the normal distribution, which is explained in detail in Chapter 10. In that chapter it is explained how the values for the probit transformation are determined. For the moment, probit can be viewed as one of three possible transformations for percentage data.

Logit

The logit transformation is less commonly employed than either arcsin or probit. It is based on a very commonsense approach to probability. For instance, at the racetrack a horse is said to have odds of winning of two to one. These odds mean that for every three races, the horse would win twice. So a two to one odds mean a two out of three probability or a 67% chance of winning the race. For the logit, one first converts a proportion into an odds. The formula for an odds given a proportion p is

$$\frac{p}{1.0 - p}$$

The odds range from zero to positive infinity. The *logit* is the natural logarithm of the odds. In equation form the logit is

$$\ln\left[\frac{p}{1.0 - p}\right]$$

where ln is the natural logarithm (base e) and p is a proportion. Recall that e approximately equals 2.718. By convention a proportion of 1.00 is set to .9975 and .00 is set to .0025. (This same change is done for probit but is not necessary for arcsin.) It might be noted that a logit of zero corresponds to a proportion of .50.

The effect of the two-stretch transformation is to stretch numbers more as they move away from .50. That is, the smallest and the largest scores are stretched the most. The most stretching transformation is logit and the least is arcsin. If most of the values are between .25 and .75, the effect of these transformations is so small as to make them unnecessary.

Flat Transformations

A flat transformation turns the sample's distribution into a flat distribution regardless of the shape of the original distribution. There are two types of major flat transformations: rank order and percentile rank.

Rank Order

This transformation has already been implicitly employed when medians and interquartile ranges are computed. For this transformation the numbers are rank ordered from the smallest to the largest. (Occasionally, numbers are ranked from largest to smallest, as in the case of class rank.) For instance, the numbers

$$7,\ 8,\ 3,\ 9,\ 1,\ 11$$

are rank ordered

$$1,\ 3,\ 7,\ 8,\ 9,\ 11$$

Then the smallest value is assigned a one, the next smallest a two, and so on. Thus, for the above sample,

$$\begin{array}{cccccc} 1, & 3, & 7, & 8, & 9, & 11 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 2 & 3 & 4 & 5 & 6 \end{array}$$

The distribution is now flat. In the case of ties, the rule is to assign the average rank. So for the sample

$$1,\ 2,\ 2,\ 2,\ 2,\ 7,\ 8$$

there are four observations equal to 2. They have ranks 2 through 5 and so the average rank equals $(2 + 3 + 4 + 5)/4$ which equals 3.5. The ranks would then be

$$1,\ 3.5,\ 3.5,\ 3.5,\ 3.5,\ 6,\ 7$$

When there are ties, the rank-order transformation does not produce a perfectly flat distribution.

Percentile Rank

Scholastic Aptitude Test (SAT) scores as well as many personality tests are presented in terms of percentile rank. In this section a method is presented to convert a score from a sample into its percentile rank. A percentile rank is measured by the following formula:

$$100\left[\frac{R - .5}{n}\right]$$

where n is the sample size and R is the rank order of the score. So if $n = 20$ and an observation has a rank order of 9, the percentile rank is

$$100\,\frac{(9 - .5)}{20} = 42.5$$

There are two major purposes for employing flat transformations. First, rank orders and percentile ranks are often easier to interpret than raw scores. Second, some of the statistics to be discussed later in this book are based on the assumption that the level of measurement is at the interval level. The effect of the flat transformation is to remove distance information between observations and make the data ordinal. Once the data are at the ordinal level of measurement, statistical procedures are available that make fewer assumptions about the data. These procedures are discussed in Chapter 18.

Two-Variable Transformations

All the transformations that have been considered involve a single sample of numbers. It is very common to have two samples of numbers from the same people. For instance, there could be scores on two quizzes for each person in a class. There are two major ways of combining information from two samples: the sum and the difference.

The simplest procedure is to sum the scores. This is sensible when the numbers measure the same trait. For instance, one can add the scores on the two quizzes. The resultant score should be more reliable than either score by itself.

A second alternative is to create a difference score. This is common for measuring change. If the earlier quiz score is subtracted from the more recent one, the resulting difference would be a measure of improvement. As a second example, studies on the effect of psychotherapy on adjustment typically have a baseline measure before psychotherapy. The outcome is the improvement from the baseline.

When working with two-variable transformations, the mean and variance have an advantage over other measures. The overall mean equals the mean of its components. For example, population change can be defined as the number of births minus the number of deaths plus the number of immigrants minus the number of emigrants. In equation form,

$$\text{population change} = \text{births} - \text{deaths} + \text{immigrants} - \text{emigrants}$$

The mean is the only measure of central tendency for which the mean number of births minus the mean number of deaths plus the mean number of immigrants minus the mean number of emigrants exactly equals the mean population change. The mean then equals the mean of its components.

The variance has a similar advantage over other measures of variability. The variance of the sum of two variables equals the sum of each variable's variance.[2] This fact is not true of either the interquartile range or the simple range.

[2]This fact holds only if the two variables are uncorrelated. When variables are correlated, the fact must be modified.

Summary

Numbers are not immutable. Researchers implicitly decide how to measure the objects of interest. There should be explicit consideration concerning how to use these numbers. In some cases it may be desirable, or even necessary, to transform the numbers.

Table 5.3 outlines the transformations that have been discussed in this chapter. The simplest type of transformation is a no-stretch transformation. This type of transformation involves adding or multiplying a number by each score in the sample. A *no-stretch transformation* preserves the basic shape of the distribution. It is used to change the units of measurement—for example, to switch from inches to centimeters. It is also used to express the score in terms of standard deviation units or Z scores. A Z or *standard score* is the individual score minus the mean, divided by the standard deviation.

The second type of transformation is called a *one-stretch transformation*. Its primary purpose is to remove positive skew from a distribution and so make the distribution symmetric. These transformations are the *square root, logarithm,* and *reciprocal*. The square root stretches the least and the reciprocal stretches the most.

The third type of transformation is a *two-stretch transformation*. This

TABLE 5.3 **Summary of Transformations**

No-Stretch	Two-Stretch
Types	Types (see Appendix A)
Adding a number to all scores	Arcsin
Multiplying scores by a number	Probit
Purposes	Logit
Change unit of measurement	Purpose
Change limits	Remove floor and ceiling
Reverse scale direction	Requirements
Standardization	Scores between 0 and 1
Requirements	
None	
One-Stretch	**Flat Transformation**
Types	Types
Square root	Rank order
Logarithm	Percentile rank
Reciprocal	Purposes
Purpose	Facilitate interpretation
Remove skew	Make data ordinal
Requirements	Requirements
Positive numbers	Few ties

transformation is appropriate for samples with a lower and an upper bound and is particularly appropriate for percentage data. These transformations are *arcsin, probit,* and *logit,* with arcsin making the smallest change and logit the largest.

The fourth type of transformation is a *flat transformation,* which turns the numbers into a flat distribution. The two transformations of this type are *rank order* and *percentile rank.* These transformations generally increase interpretability and allow for a relaxation of some assumptions for statistical tests.

At first reading, the topic of transformations may seem to be bewildering because some of the mathematical operations seem complex. However, because the one-stretch transformations can be performed with a calculator and two-stretch transformations are tabled in Appendix A, all the messy mathematical complications are circumvented. The other transformations involve only the simple operations of addition, multiplication, and rank ordering.

The second aspect of transformation that leads to confusion is the decision of which transformation to use. Say a researcher wishes to employ a one-stretch transformation. How does he or she know whether to use square root, logarithm, or reciprocal? Which one is best? The researcher generally does not know this in advance. It may be necessary to try each out and determine which one makes the distribution more symmetric. Eventually, it should be clear which transformation, if any, is best.

Problems

1. Compute to three decimal points each of the following.

   a. ln 10
   b. square root of 39
   c. arcsin of .75
   d. 1/48
   e. probit of .66
   f. logit of .54
   g. ln 63
   h. square root of 75
   i. arcsin of .44
   j. logit of .17
   k. probit of .23
   l. logit of .50

2. Using the median and the mean for the following sample of numbers, which one-stretch transformation makes the distribution most symmetric?

1	4	5	3
9	3	1	4
3	15	2	7
4	8	18	6
8	7	9	3

3. Compute Z scores for the numbers in problem 2. The mean equals 6.0 and the standard deviation is 4.40.

4. For a sample with limits of zero and 100, state how to change the limits to one and ten.

5. For the following transformations give the new mean and standard deviation for Z.

   a. $Z = 10X + 3$, where $\bar{X} = 5$ and $s_x = 3$.
   b. $Z = Y/5 + 15$, where $\bar{Y} = 9$ and $s_y = 1$.
   c. $(X - 10)/15 = Z$, where $\bar{X} = 10$ and $s_x = 15$.
   d. $Z = 6X + 4$, where $\bar{X} = 5$ and $s_x = 2$.

6. For the following sample, compute percentile ranks for the scores 8, 12, and 17.

7	13	14	14
9	12	6	9
15	18	17	3
6	6	21	15
8	11	9	20

7. Suppose that a sample of numbers has a mean of 25 and a standard deviation of 3.20. Compute the means, standard deviations, and variances that would result from the following transformations.

   a. Add five to all scores.
   b. Multiply scores by three.
   c. Subtract 20 and then multiply scores by five.
   d. Add 100 and then multiply scores by two.

8. Chapter 1, problem 6, gives ratings of discomfort by subjects who were offered aid. The data are repeated below. The scale runs from one to ten, with higher scores indicating greater discomfort. Transform the eight numbers by reversing the scale so that the scale still runs from one to ten, and a higher score indicates more comfort.

   10, 1, 3, 5, 3, 7, 7, 3

9. a. For the data in the previous problem, transform the eight scores into Z scores.
   b. Transform the eight scores into percentile ranks.

10. Suppose that the possible points on a final exam range from five to 75.

Change the lower limit to zero and the upper limit to 100 and transform the following scores.

a. 55 b. 72 c. 68 d. 40

11. Vinsel, Brown, Altman, and Foss (1980) conducted a study in which one of the dependent measures was the type of wall decorations used by dormitory freshmen. The following table shows the average area covered by each type of decoration by male and female subjects.

	Area	
Category	*Males*	*Females*
A. Personal relationships	8.5	20.1
B. Abstract	66.1	80.7
C. Music/theater	24.4	20.3
D. Sports	77.4	21.1
E. Values	9.4	16.8
F. Reference items	18.7	5.8
G. Idiosyncratic	46.4	17.8
H. Entertainment/equipment	15.4	10.0

a. For the males, rank the categories from least area used to most area used.
b. Do the same for the females.
c. For which category does the ranking for the females and males differ the most?

12. Chapter 4, problem 7 gave the scores from an experiment on programmed instruction. Some of the data are repeated below.

Class Midpoint	*Frequency*
67	1
72	1
77	5
82	7
87	9
92	14
97	5

a. Apply a square root transform to the scores.
b. Perform a natural log transform on the scores.
c. Perform a reciprocal transform on the scores.
d. Compute the mean and mode of the original data. (Make sure to weight by frequency.) In which direction are the data skewed? Do the same for the three sets of transformed scores. How do the transformations affect the skew?