CHAPTER 16

THE GARBAGE CAN

(Version 3, 14 March 2013)

Page

Every book on methods must end with a series of miscellaneous items that do not fit easily into one of the more structured chapters. In this final chapter I discuss the statistical ideas of transformations and repeatability, the problems of estimating trend lines, how to measure temporal variability, and two new computer-intensive methods that hold much promise for ecological statistics.

16.1 TRANSFORMATIONS

A transformation is a change of numerical scale. When you convert inches to centimeters you are applying a transformation. Most transformations we are familiar with are of this type, and are called *linear* transformations. With all linear transformations, the standard statistics can be readily converted. Given a multiplicative transformation of the type $X_T = c\,X_0$, it follows that:

$$\bar{x}_T = c\,\bar{x}_0 \qquad (16.1)$$

where $\bar{x}_T$ = Transformed mean (e.g. cm)
$\bar{x}_0$ = Original mean (e.g. inches)
c = Constant of conversion (e.g. 2.54 to convert inches to centimeters)

Similarly 95% confidence limits can be converted with equation (16.1). To convert standard deviations or standard errors:

$$S_T = c\,S_0 \qquad (16.2)$$

where S_T = Transformed standard error or standard deviation
S_0 = Original standard error or standard deviation
c = Constant of conversion

If a fixed amount is added to or subtracted from each observation, the mean will be increased or decreased by that amount, but the standard deviation and standard error are not changed at all. These simple transformations are not usually the type that statisticians write about.

Parametric methods of statistical analysis assume that the variable being measured has a normal distribution. Ecological variables often have skewed distributions which violate this assumption. In addition to this normality assumption, more complex parametric methods such as the analysis of variance assume that all groups have the same variance and that any treatment effects are additive. All statistics books discuss these assumptions (e.g. Sokal and Rohlf 2012, Chap. 13 or Zar 2010, Chap. 13).

There are four solutions to the problems of non-normality, heterogeneous variances, and non-additivity. The first solution is to use non-parametric methods of data analysis. There is a large literature on non-parametric methods which can be applied to much ecological data (Siegel and Castellan 1988, Tukey 1977, Chambers *et al.* 1983). Non-parametric methods rely on ranks rather than absolute values and as such they lose the arithmetic precision many ecologists desire. Nonparametric methods have been waning in popularity in recent years (Underwood 1997, Day and Quinn 1989, p. 448).

The second solution is to transform the scale of measurement so that the statistical demands of parametric analyses are approximately satisfied. The

advantage of using this solution is that the whole array of powerful methods developed for parametric statistics can be employed on the transformed data. If you choose to use transformations on your data, you have to decide on exactly what transformation to use. There are two ways to do this: (1) use one of the four standard transformations discussed in every statistics book; or (2) choose a general transformation that can be tailor-made to your specific data. The most widely-used general transformation is the *Box-Cox transformation*.

The third solution is to ignore the problem and to argue that parametric methods like ANOVA are robust violations of their assumptions. This is not recommended, although some statisticians argue that if you have large sample sizes (n > 30 in all groups) you need to worry less about these assumptions. In most cases I would not recommend this solution because one of the other three is preferable.

The fourth solution is to use computer intensive methods of data analysis. These are the most recent tools available to ecologists who must deal with data which are not normally distributed, and they are now available widely in statistics packages for desktop computers. Randomization tests are one simple type of computer intensive methods (Sokal and Rohlf 2012 p. 803). Students are referred to Noreen (1990) and Manly (1991), specialized books which discuss these procedures in detail.

16.1.1 Standard Transformations

The four most commonly used statistical transformations are the *logarithmic* transformation, the *square-root* transformation, the *arcsine* transformation, and the *reciprocal* transformation. These are discussed in more detail by Hoyle (1973), Thöni (1967), and Mead (1988, Chapter 11).

Logarithmic Transformation

This transformation is commonly used in ecological data in which percentage changes or multiplicative effects occur. The original data are replaced by:

$$X' = \log(X) \quad \text{or}$$
$$X' = \log(X + 1) \quad \text{(if data contains 0 values)} \tag{16.3}$$

where X' = Transformed value of data
X = Original value of data

and logarithms may be to base 10 or base *e* (which differ only by a constant multiplier).

Asymmetrical or skewed distributions are common in ecological data. Figure 16.1 illustrates two examples. Positively skewed distributions have a tail of data pointing to the right, and negatively skewed distributions have a tail pointing to the left. A number of measures of skewness are available (Sokal and Rohlf 1995, p. 114) in standard statistical books and packages.

The logarithmic transformation will convert a positively skewed frequency distribution into a more nearly symmetrical one. Data of this type may be described by the *log-normal* distribution. Note that the logarithmic transformation is very strong in correcting positive skew, much stronger than the square root transformation. Figure 16.2 illustrates the impact of a logarithmic transformation on a positively skewed frequency distribution.

In all work with parametric statistics, you use the transformed (log) values, so that means, standard errors and confidence limits are all expressed in logarithm units. If you wish to have the final values expressed in the original measurement units, remember the following three rules (which apply to all transformations):

1. Never convert *variances*, *standard deviations*, or *standard errors* back to the original measurement scale. They have no statistical meaning on the original scale of measurement.

2. Convert *means* and *confidence limits* back to the original scale by the inverse transformation. For example, if the transformation is log (X+1), the mean expressed in original units will be:

$$\bar{X} = \left[\text{antilog}\left(\bar{X}'\right)\right] - 1 = 10^{\bar{X}'} - 1 \qquad (16.4)$$

In the particular case of the logarithmic transformation, the mean expressed in original units is equivalent to the geometric mean of the original data. In some cases it may be desirable to estimate the arithmetic mean of the original data (Thöni 1967,

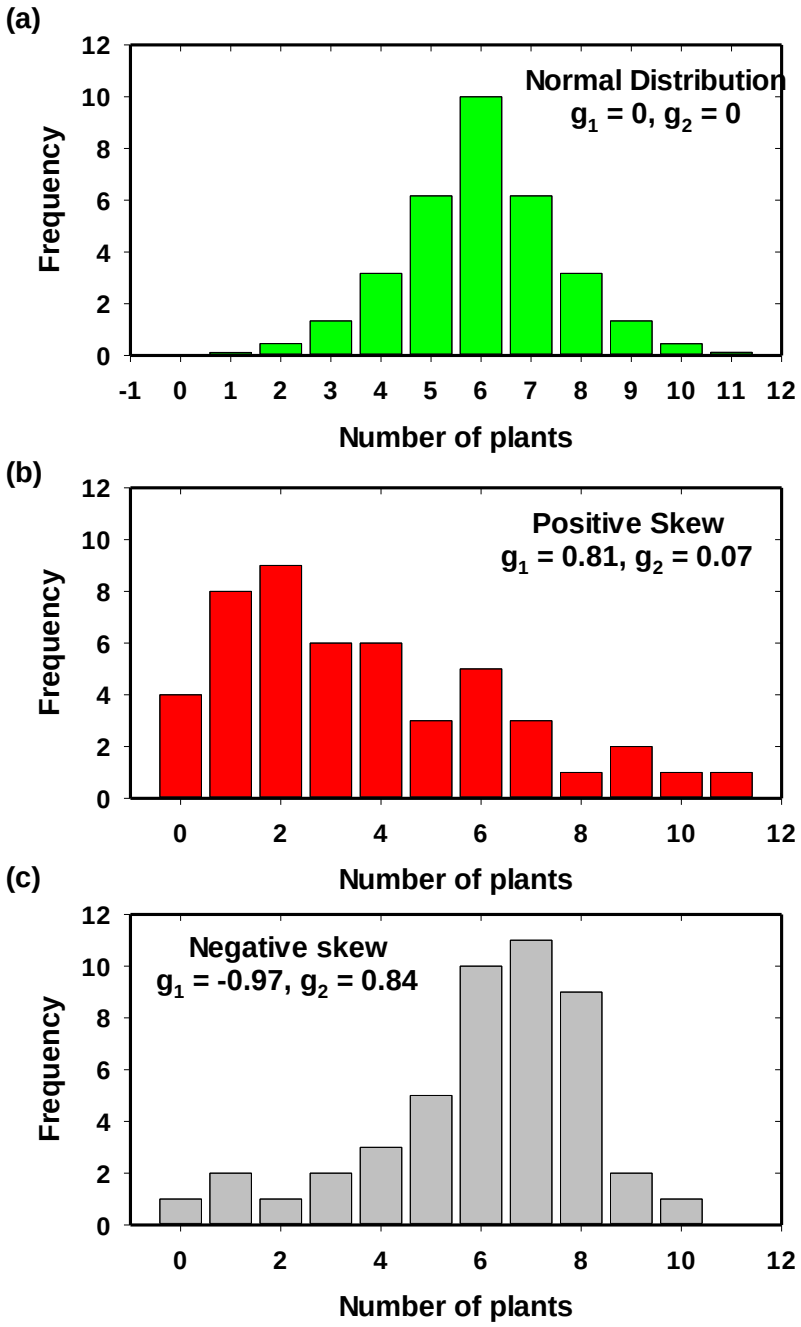

Figure 16.1 Illustration of positive and negative skew in observed frequency distributions. (a) The normal distribution, assumed in all of parametric statistics. (b) Positive skew illustrated by plant counts that fit a clumped distribution. (c) Negative skew illustrated by counts of highly clumped species. Skewed distributions can be normalized by standard transformations like the square root or the logarithmic transformation.

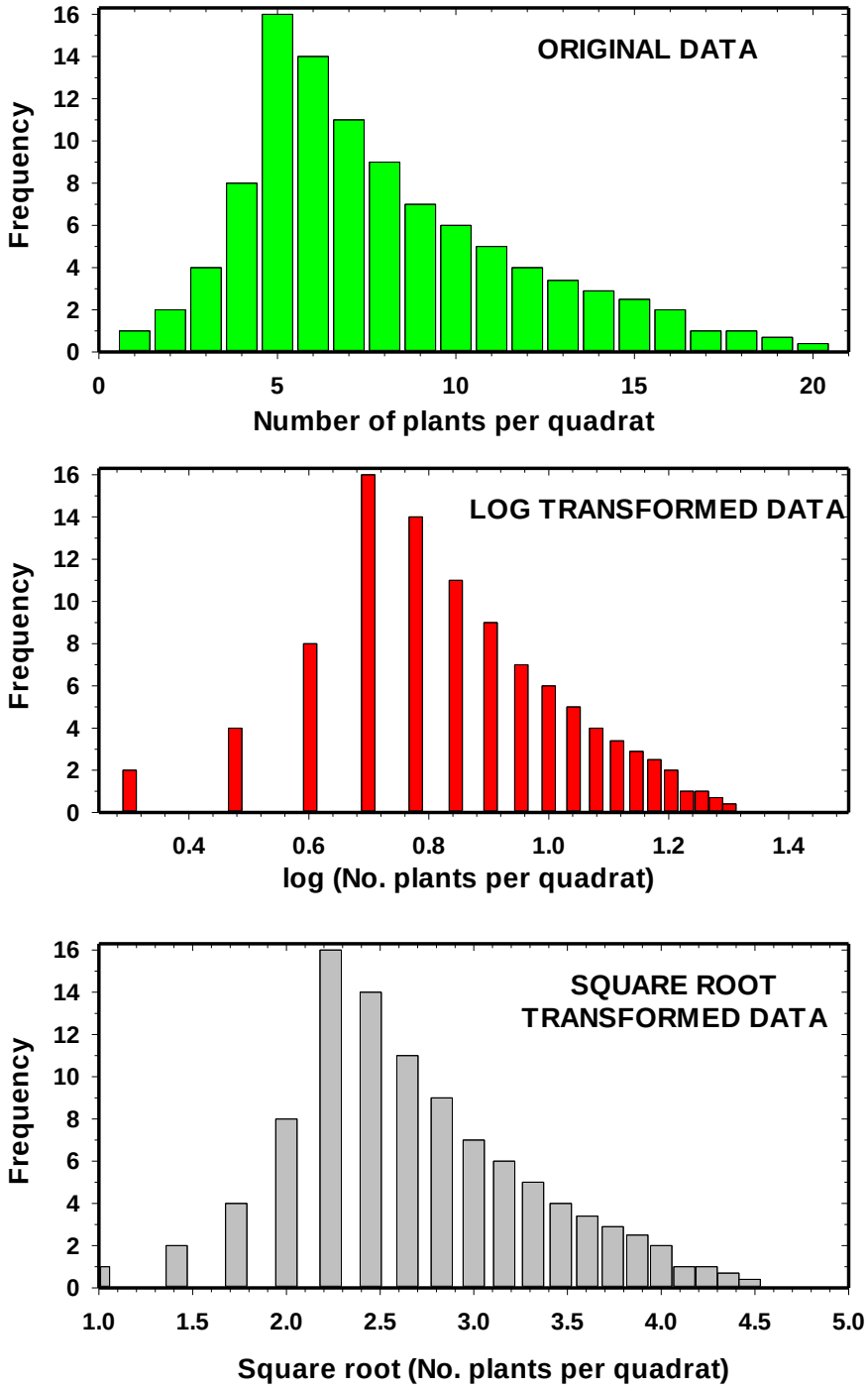

Figure 16.2 Illustration of the impact of a logarithmic transformation and a square root transformation on a positively skewed frequency distribution. Both these transformations make the original data more nearly normal in shape but the square root transformation is slightly better in this example.

Hoyle 1973). If you use the original data to estimate the mean, you usually obtain a very poor estimate (both inaccurate and with low precision) if sample size is relatively small (n < 100). An unbiased estimate of the mean in the original measurement units can be obtained from the approximate equation given by Finney (1941):

$$\hat{\bar{X}} = \left(e^{\bar{X}'+0.5s^2}\right)\left[1-\frac{s^2\left(s^2+2\right)}{4n}+\frac{s^4\left(3s^4+44s^2+84\right)}{96n^2}\right] \qquad (16.5)$$

where $\hat{\bar{X}}$ = Estimated arithmetic mean of original data
X' = Observed $\log_e$ -transformed mean of data
s^2 = Observed variance of $\log_e$ transformed data
n = Sample size

Finney (1941) suggested that a minimum sample size of 50 is needed to get a reasonable estimate of the arithmetic mean with this procedure.

3. *Never* compare means calculated from untransformed data with means calculated from any transformation, reconverted back to the original scale of measurement. They are not comparable means. All statistical comparisons between different groups must be done using one common transformation for all groups.

When logarithmic transformations are used in linear regression on both the X and Y variables (log-log plots), the estimation of Y from X is biased if antilogs are used without a correction. Baskerville (1972) and Sprugel (1983) give the correct unbiased formulae for estimating Y in the original data units.

There is considerable discussion in the statistical literature about the constant to be used in logarithmic transformations. There is nothing magic about log(X+1) and one could use log(X+0.5) or log(X+2). Berry (1987) discusses the statistical problem associated with choosing the value of this constant. Most ecologists ignore this problem and use log(X+1) but Berry (1987) points out that the value of the constant chosen may greatly affect the results of parametric statistical tests like ANOVA. Berry (1987) argues that we should choose the constant c that minimizes the sum of skewness and kurtosis:

$$G_c = \left|g_1(c)\right| + \left|g_2(c)\right| \qquad (16.6)$$

where G_c = Function to be minimized for a particular value of c
c = The constant to be added in a logarithmic transform
$g_1(c)$ = Estimate of skewness from a normal distribution for the chosen value of c
$g_2(c)$ = Estimate of kurtosis from a normal distribution for the chosen value of c

Figure 16.1 illustrates the concept of skewness in data. Kurtosis refers to the proportion of observations in the center of the frequency distribution in relation to the observations in the tails. Measures of both skewness (g_1) and kurtosis (g_2) are relative to the normal distribution in which g_1 and g_2 are zero. Methods to estimate skewness and kurtosis are given in most statistics books (Sokal and Rohlf 2012, p. 115, Zar 2010 p. 67):

$$g_1 = \frac{n\sum X^3 - 3\left(\sum X\right)\left(\sum X^2\right) + {2\left(\sum X\right)^3}/{n}}{(n-1)(n-2)s^3} \qquad (16.7)$$

where g_1 = Measure of skewness (= 0 for a normal distribution)
n = Sample size
$\sum X$ = Sum of all the observed X-values
$\sum X^2$ = Sum of all the observed X-values squared
s = Observed standard deviation

$$g_2 = \frac{(n+1)\left\{n\sum X^4 - 4\left(\sum X\right)\left(\sum X^3\right) + \left[{6\left(\sum X\right)^2\left(\sum X^2\right)}/{n}\right] - \left[{3\left(\sum X\right)^4}/{n^2}\right]\right\}}{(n-1)(n-2)(n-3)s^4} - \frac{3(n-1)^2}{(n-2)(n-3)} \qquad (16.8)$$

where g_2 = Measure of kurtosis (= 0 for a normal distribution)
$\sum X^3$ = Sum of all the observed X-values cubed
$\sum X^4$ = Sum of all the observed X-values to fourth power

The procedure is to compute G_c (equation 16.6) using the logarithmic transformation with a series of c values from (–) the minimum value observed in the raw data to 10 times the maximum value observed in the data (Berry 1987), as well as for the raw data with no transformation. This procedure is tedious and best done by computer. Program EXTRAS (Appendix 2, page 000) does these calculations. Plot the

resulting G_c values on the *y*-axis against the *c*-values (*x*-axis) as in Figure 16.3, and choose the value of *c* that minimizes G_C. Berry (1987) recommends using this value as the basis of the transformation:

$$X' = \log(X + c) \qquad (16.9)$$

where *c* = the constant that minimizes the function G_C defined in equation 16.6

In some cases *c* will be nearly 1.0 and the standard logarithmic transformation of equation (16.3) will be adequate. In other cases the G_c curve is relatively flat (as in Figure 16.3) and there is a broad range of possible *c* values that will be equally good. In these cases it is best to use the smallest value of *c* possible.

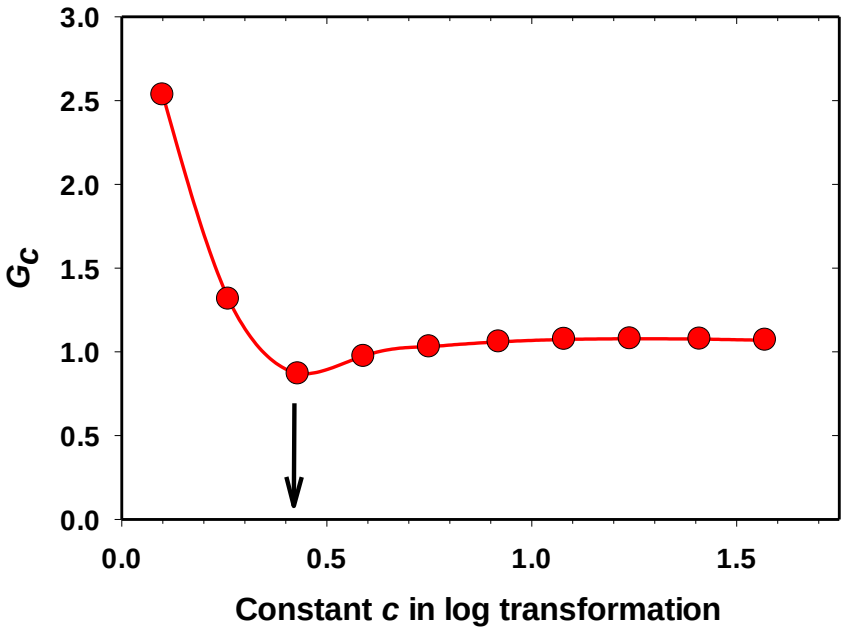

Figure 16.3 Estimation of the constant *c* to be used in logarithmic transforms of the type log (*X*+*c*) using the method of Berry (1987). Locate the minimum of the function G_c as defined in equation (16.6). Data from Box 4.4 on the abundance of black bean aphids was used for these calculations. There is a minimum in the G_c function around 0.43 (arrow) and the recommended logarithmic transform for these data would be log (*X*+0.43).

Square Root Transformation

This transformation is used when the variance is proportional to the mean, a common ecological situation (see Taylor's power law, Chapter 9, page 000). Any

ecological data of counts fitting a Poisson distribution should be transformed with square roots before parametric statistics are applied. The original data are replaced by:

$$X' = \sqrt{X + 0.5} \qquad (16.10)$$

This transform is preferable to the straight square root transform when the observed data are small numbers and include some zero values (Zar 1996, p. 279).

If you wish to obtain the mean and confidence limits in the original measurement units, you can reverse the transformation -

$$\bar{X} = \left(\bar{X}'\right)^2 - 0.5 \qquad (16.11)$$

This mean is slightly biased (Thöni 1967) and a more precise estimate is given by:

$$\bar{X} = \left(\bar{X}'\right)^2 - 0.5 + s^2\left(1 - \frac{1}{n}\right) \qquad (16.12)$$

where $\bar{X}$ = Mean expressed in original measurement units
$\bar{X}'$ = Mean obtained from square root transformation (eq. 16.10)
s^2 = Variance of square-root transformed data
n = Sample size

Several variants of the square root transformation have been suggested to reduce the relationship between the variance and the mean. Anscombe (1948) suggested that

$$X' = \sqrt{X + 3/8} \qquad (16.13)$$

was better than equation (16.10) for stabilizing the variance. Freeman and Tukey (1950) showed that when $X \leq 2$, a better transformation is:

$$X' = \sqrt{X} + \sqrt{X + 1} \qquad (16.14)$$

These variants of the square root transformation are rarely used in practice.

Arcsine Transformation:

Percentages and proportions form a binomial distribution when there are two categories or a multinomial distribution when there are several categories, rather than a normal distribution. Consequently parametric statistics should not be computed for percentages or proportions without a transformation. In cases where the percentages range from 30% to 70%, there is no need for a transformation, but

if any values are nearer to 0% or 100%, you should use an arcsine transformation. The term *arcsine* stands for the angle whose sine is a given value. Note that in mathematical jargon:

arcsine = inverse sine = $\sin^{-1}$

The recommended arcsine transformation is given by:

$$X' = \arcsin\sqrt{p} \qquad (16.15)$$

where X' = transformed value (measured in degrees)
p = observed proportion (range 0-1.0)

Transformed values may also be given in radians rather than degrees. The conversion factor is simply:

$$1 \text{ radian } = \frac{180}{\pi} \text{ degrees } = 57.2957795 \text{ degrees}$$

To convert arcsine transformed means back to the original scale of percentages or proportions, reverse the procedure:

$$\bar{p} = \left(\sin \bar{X}'\right)^2 \qquad (16.16)$$

where $\bar{p}$ = Mean proportion
X' = Mean of arcsine transformed values

Mean proportions obtained in this way are slightly biased and a better estimate from Quenouille (1950) is:

$$\bar{p}_c = \left(\sin \bar{X}'\right)^2 + 0.5\cos\left[\left(2\bar{p}\right)\left(1 - e^{-2s^2}\right)\right] \qquad (16.17)$$

where $\bar{p}_c$ = Corrected mean proportion
$\bar{p}$ = Mean proportion estimated from equation (16.16)
s^2 = Variance of the transformed values of X' (from equation 16.15)

If you have the raw data, you can use a better transformation suggested by Anscombe (1948):

$$X' = \sqrt{n + 0.5}\left(\arcsin\sqrt{\frac{X + 3/8}{n + 3/4}}\right) \qquad (16.18)$$

where X' = Transformed value
n = Sample size
X = Number of individuals with the attribute being measured

This transformation leads to a variable X' with expected variance 0.25 over all values of X and n. An alternative variant is suggested by Zar (1984, p. 240):

$$X' = 0.5\left(\arcsin\sqrt{\frac{X}{n+1}} + \arcsin\sqrt{\frac{X+1}{n+1}}\right) \qquad (16.19)$$

where all terms are as defined above.

This alternative may be slightly better than Anscombe's when most of the data are very large or very small proportions.

There is always a problem with binomial data because they are constrained to be between 0 and 1, and consequently no transformation can make binomial data truly normal in distribution. Binomial data can also be transformed by the logit transformation:

$$X' = \log_e\left(\frac{p}{q}\right) \qquad (16.20)$$

where X' = logit transform of observed proportion p
$q = 1-p$

Note that the logit transform is not defined for $p = 1$ or $p = 0$ and if your data include these values you may add a small constant to the numbers of individuals with the attribute as we did above in equation (16.18). The logit transformation will act to spread out the tails of the distribution and may help to normalize the distribution.

Reciprocal Transformation

Some ecological measurements of rates show a relationship between the standard deviation and the $(\text{mean})^2$. In these cases, the reciprocal transformation is applied to achieve a nearly normal distribution:

$$X' = \frac{1}{X} \qquad (16.21)$$

or if there are observed zero values, use:

$$X' = \frac{1}{X+1} \qquad (16.22)$$

Thöni (1967) discusses the reciprocal transformation in more detail.

16.1.2 Box-Cox Transformation

In much ecological work using parametric statistics, transformations are applied using rules of thumb or tradition without any particular justification. When there is no strong reason for preferring one transformation over another, it is useful to use a more general approach. Box and Cox (1964) developed a general procedure for finding out the best transformation to use on a set of data in order to achieve a normal distribution. Box and Cox (1964) used a family of power transformations of the form:

$$X' = \frac{X^{\lambda} - 1}{\lambda} \qquad \text{(when } \lambda \neq 0\text{)} \qquad (16.23)$$

or

$$X' = \log(X) \qquad \text{(when } \lambda = 0\text{)} \qquad (16.24)$$

This family of equations yields most of the standard transformations as special cases, depending on the value of λ. For example, if $\lambda = 1$, there is effectively no transformation (since $X' = X - 1$); when $\lambda = 0.5$, you get a square root transformation; and when $\lambda = -1$, you get a reciprocal transformation.

To use the Box-Cox transformation, choose the value of \ that maximizes the log-likelihood function:

$$L = -\frac{\nu}{2}\log_e s_T^2 + (\lambda - 1)\frac{\nu}{n}\sum(\log_e X) \qquad (16.25)$$

where L = Value of log-likelihood
ν = Number of degrees of freedom (n–1)
s_T^2 = Variance of transformed X-values (using equation 16.23)
λ = Provisional estimate of power transformation parameter
X = Original data values

This equation must be solved iteratively to find the value of λ that maximizes L. Since this is tedious, it is usually done by computer. Box 16.1 shows these calculations and Figure 16.4 illustrates the resulting plot of the log-likelihood function for a set of data.

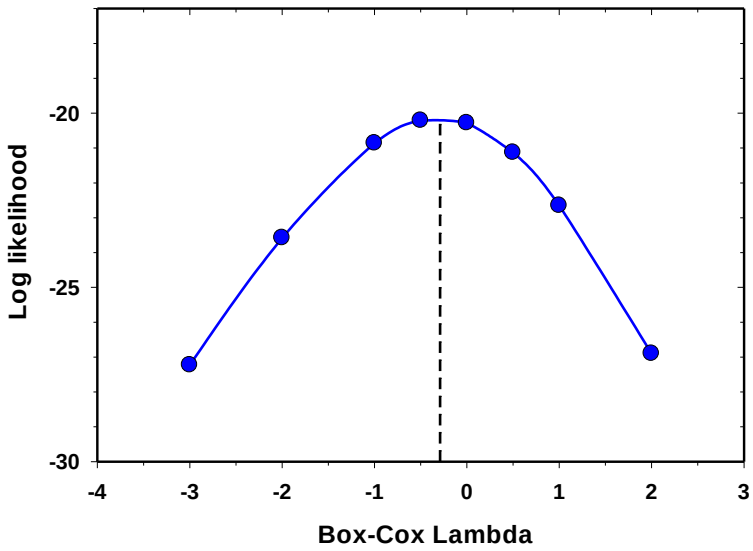

Figure 16.4 Log-likelihood function for the grassland plot data given in Box 16.1. Maximum likelihood occurs at λ = -0.29, and this exponent could be used in the Box-Cox transformation.

Box 16.1 CALCULATION OF BOX-COX TRANSFORMATION FOR CLIP-QUADRAT DATA

A series of grassland plots were clipped at the height of the growing season to estimate production for one season, with these results:

Plot no.	Dry weight of grass (g)
1	55
2	23
3	276
4	73
5	41
6	97

1. Choose a trial value of λ, say $\lambda = -2.0$.

2. Transform each data point using equation (16.23) and $\log_e (X)$:

X	$\log_e (X)$	$\frac{X^{\lambda}-1}{\lambda}$
55	4.00733	0.4998347
23	3.13549	0.4990548
276	5.62040	0.4999934
73	4.29046	0.4999062
41	3.71357	0.4997026

97	4.57471	0.4999469
Sum	25.34196	

3. Calculate the variance of the transformed weights (third column):

$$s_T^2 = \frac{\sum X^2 - \left(\sum X\right)^2 / n}{n-1} = 12.29013 \times 10^{-8}$$

4. Calculate the log-likelihood function (equation 16.25):

$$L = -\frac{\nu}{2}\log_e s_T^2 + \left[(\lambda - 1)\frac{\nu}{n}\sum(\log_e X)\right]$$

$$= -\frac{5}{2}\left[\log_e(0.0000001229)\right] + \left[(-2-1)\left(\frac{5}{6}\right)(25.34196)\right] = -23.58$$

5. Repeat (1) to (4) using different values of λ to get:

λ	*L*
–3	–27.23
–1	–20.86
–0.5	–20.21
0	–20.28
+0.5	–21.13
1	–22.65
2	-26.90

These values are plotted in Figure 16.4 Clearly, there is a maximum between λ = 0 and λ = –1.0. By further application of steps (1) to (4) you can show that the maximum likelihood occurs at λ = –0.29.

Program EXTRAS (Appendix 2, page 000) can do these calculations and estimates confidence limits for the best value of λ.

When the original data include zeros, equation (16.25) becomes insoluble because the log of 0 is negative infinity. For these cases, Box and Cox (1964) suggest adding a constant like 0.5 or 1.0 to each *X* value before doing this transformation. It is possible to use the log likelihood function to search for the best value of this constant as well as λ but in most cases (*X*+0.5) or (*X*+1.0) will be sufficient to correct for data that include zeros.

Box and Cox (1964) showed that confidence limits could be estimated for the λ parameter of the power transformation from the chi-squared distribution. If these confidence limits include $\lambda = 1.0$, the data may not need any transformation.

In practice, when one is collecting the same type of data from many areas or from many years, the Box-Cox procedure may be applied to several sets of data to see if a common value of λ may be estimated. This value of λ could then be used to specify the data transformation to be used in future data sets. As with other transformations, one should not change values of λ within one analysis or the transformed means will not be comparable.

The Box-Cox procedure is a very powerful tool for estimating the optimal transformation to use for ecological data of a wide variety. Its application is limited only by the amount of computation needed to use it. Program EXTRAS (Appendix 2, page 000) does these calculations to estimate λ and gives 95% confidence limits for λ. Rohlf (1995) provides additional programs that employ this transformation.

TABLE 16.1 REPEATABILITY OF AERIAL COUNTS OF MOOSE IN SIX AREAS OF HABITAT IN BRITISH COLUMBIA[a]

	Habitat Block					
	A	B	C	D	E	F
	16	8	27	43	4	14
	17	6	27	41	3	15
	15	8	24	44	4	13
	16	8	26	40	3	15
	18		25	41	4	14
	16		27	39		13
			26			15
No. of counts	6	4	7	6	5	7
$\sum X$	98	30	182	248	18	99
Mean count	16.3	7.5	26.0	41.3	3.6	14.1

[a] Several observers were used on sequential trips within the same 10-day period.

16.2 REPEATABILITY

Ecologists usually assume all the measurements they take are highly precise and thus repeatable. If individual *a* counts 5 eggs in a nest, it is reasonable to assume that individual *b* will also count 5 eggs in the same nest. For continuous variables like weight, this assumption is less easily made, and in cases where some observer skill is required to take the measurement, one should not automatically assume that measurements are repeatable. Consequently, one of the first questions an ecologist should ask is: *Are these measurements repeatable*?

Repeatability is a measure ranging from 0 to 1.0 that shows how similar repeated measurements are on the same individual items. It is also known as the *intraclass correlation coefficient* (Sokal and Rohlf, 1995, p. 213), and is calculated as follows. A series of individuals (or items) is measured repeatedly by the same individual or by several individuals. Table 16.1 gives an example from aerial census of moose. Data are cast in an ANOVA table for a one-way design:

ANALYSIS OF VARIANCE TABLE

Source	Degrees of freedom	Sum of squares	Mean square
Among groups	a -1	SS_A	MS_A
Within groups between measurements ('error')	$\sum(n_i - 1)$	SS_E	MS_E

where a = Total number of items being measured repeatedly
n_i = Number of repeated measurements made on item i

Formulas for calculating the sums of squares and mean squares in this table are given in every statistics book (e.g. Sokal and Rohlf 2012, p. 211; Zar 2010 p. 182). Box 16.2 works out one example. Repeatability is given by:

$$R = \frac{s_A^2}{s_E^2 + s_A^2} \qquad (16.26)$$

where R = Repeatability
s_A^2 = Variance among groups
s_E^2 = Variance within groups

R is thus the proportion of the variation in the data that occurs among groups. If measurements are perfectly repeatable, there will be zero variance within a group

and *R* will be 1.0. The two variance components are obtained directly from the ANOVA table above:

$$s_E^2 = \mathrm{MS}_E \tag{16.27}$$

$$s_A^2 = \frac{\mathrm{MS}_A - \mathrm{MS}_E}{n_0} \tag{16.28}$$

where $$n_0 = \frac{1}{a-1}\left(\sum n_i - \frac{\sum n_i^2}{\sum n_i}\right) \tag{16.29}$$

where a = Number of groups (items) being measured repeatedly
n_i = Number of repeated measurements made on item i

Lessells and Boag (1987) pointed out a recurring mistake in the literature on repeatabilities where incorrect values of *R* are obtained because of the mistake in confusing mean squares and variance components:

$$s_A^2 \neq \mathrm{MS}_A \tag{16.30}$$

Since repeatability can be used in quantitative genetics to give an upper estimate of the heritability of a trait, these mistakes are serious and potentially confusing. Lessells and Boag (1987) give an approximate method for estimating the correct repeatability value from published ANOVA tables. They point out that if:

$$\text{Published } R \text{ value} = \frac{F}{F+1} \tag{16.31}$$

where F = F-ratio from a one-way ANOVA table

then the published repeatability must be *wrong*! This serves as a useful check on the literature.

Confidence limits on repeatability can be calculated as follows (Becker 1984):

Lower confidence limit:

$$R_L = 1.0 - \frac{n_0 \mathrm{MS}_E F_{\alpha/2}}{\mathrm{MS}_A + \mathrm{MS}_E (n_0 - 1)(F)} \tag{16.32}$$

where $F_{\alpha/2}$ = value from F-table for $\alpha/2$ level of confidence (e.g. $F_{.025}$ for 95% confidence limits)

Upper confidence limit:

$$R_U = 1.0 - \frac{n_0 \, \mathrm{MS}_E \, F_{1-\alpha/2}}{\mathrm{MS}_A + \mathrm{MS}_E \left(n_0 - 1\right)\left(F\right)} \qquad (16.33)$$

where $F_{1-\alpha/2}$ = value from *F*-table for (1 - α/2) level of confidence (e.g. $F_{.975}$ for 95% confidence limits)

The *F*-table is entered with the degrees of freedom shown in the ANOVA table.

These confidence limits are not symmetric about *R*.

Box 16.2 gives an example of repeatability calculations. Program EXTRAS (Appendix 2, page 000) will do these calculations.

Box 16.2 CALCULATION OF REPEATABILITY FOR SNOWSHOE HARE HIND FOOT LENGTH

A group of four field workers measured the right hind foot on the same snowshoe hares during one mark-recapture session with the following results:

	Observer					
Hare tag	A	B	C	D	n_i	$\sum X$
171	140	140			2	280
184	125	125			2	250
186	130	129	135		3	394
191	130	132	132		3	394
192	131	134			2	265
193	139	140	142		3	421
196	127	127			2	254
202	130	130	130	133	4	523
203	129	132	132		3	393
207	138	137	138	138	4	551
211	140	141	143	141	4	565
217	147	149	147		3	443
				Total	35	4733

Grand mean foot length = 135.2286 mm

1. Calculate the sums for each individual measured and the grand total, as shown above in column 7, and the sum of the individual items squared: (*i* = hare, *j* = observer)

$$\sum X_{ij}^2 = 140^2 + 140^2 + 125^2 + \cdots = 641{,}433$$

2. Calculate the *sums of squares*. The sums of squares *among* hares is:

$$\begin{aligned} \mathrm{SS}_A &= \sum\left[\frac{\left(\sum X_i\right)^2}{n_i}\right] - \frac{(\text{Grand total})^2}{\text{Total sample size}} \\ &= \frac{280^2}{2} + \frac{250^2}{2} + \frac{394^2}{3} + \cdots + \frac{443^2}{3} - \frac{(4733)^2}{35} \\ &= 641{,}379.5834 - 640{,}036.8286 = 1342.7548 \end{aligned}$$

The sum of squares *within* individual hares is:

$$\begin{aligned} \mathrm{SS}_E &= \sum\left(X_{ij}^2\right) - \sum\left[\frac{\left(\sum X_i\right)^2}{n_i}\right] \\ &= 641{,}433 - 641{,}379.5834 = 53.4166 \end{aligned}$$

3. Fill in the ANOVA table and divide the sums of squares by the degrees of freedom to get the mean squares:

Source	d.f.	Sum of squares	Mean square
Among hares	11	1342.7548	122.0686
Within individuals ("error")	23	53.4166	2.32246

4. Calculate the *variance components* from equations (16.27) to (16.29):

$$s_E^2 = \mathrm{MS}_E = 2.32246$$

$$\begin{aligned} n_o &= \frac{1}{a-1}\left(\sum n_i - \frac{\sum n_i^2}{\sum n_i}\right) \\ &= \frac{1}{12-1}\left(35 - \frac{2^2+2^2+3^2+3^2+2^2+\cdots}{35}\right) \\ &= 2.8987 \text{ (effective average number of replicates per individual hare)} \end{aligned}$$

$$\begin{aligned} s_A^2 &= \frac{\mathrm{MS}_A - \mathrm{MS}_E}{n_0} \\ &= \frac{122.0686 - 2.32246}{2.8987} = 41.31 \end{aligned}$$

5. Calculate *repeatability* from equation (16.26):

$$\begin{aligned} R &= \frac{s_A^2}{s_A^2 + s_E^2} \\ &= \frac{41.31}{41.31 + 2.32} = 0.947 \end{aligned}$$

6. Calculate the *lower confidence limit* for *R* from equation (16.32):

$$R_L = 1.0 - \frac{n_0 \text{MS}_E F}{\text{MS}_A + \text{MS}_E (n_0 - 1)(F)}$$

With n_1 = 11 d.f. and n_2 = 23 d.f from the *F*-table for α = 0.025 we get F = 2.62 and thus:

$$R_L = 1.0 - \frac{(2.8987)(2.32246)(2.62)}{122.0686 + 2.32246(2.8987 - 1)(2.62)} = 0.868$$

7. Calculate the *upper confidence limit* for *R* from equation (16.33). To calculate *F* for α = 0.975 for n_1 and n_2 degrees of freedom, note that:

$$F_{.975}(n_1, n_2) = \frac{1.0}{F_{.025}(n_1, n_2)}$$

Since $F_{.025}$ for 23 and 11 d.f. is 3.17, $F_{.975}$ for 11 and 23 d.f. is 0.3155.

$$R_U = 1.0 - \frac{n_0 \text{MS}_E F}{\text{MS}_A + \text{MS}_E (n_0 - 1)(F)}$$

$$= 1.0 - \frac{(2.8987)(2.32246)(0.3155)}{122.0686 + 2.32246(2.8987 - 1)(0.3155)} = 0.983$$

These calculations can be carried out in Program EXTRAS, Appendix 2, page 000.

16.3 CENTRAL TREND LINES IN REGRESSION

Linear regression theory, as developed in standard statistics texts, applies to the situation in which one independent variable (*X*) is used to predict the value of a dependent variable (*Y*). In many ecological situations, there may be no clear dependent or independent variable. For example, fecundity is usually correlated with body size, or fish length is related to fish weight. In these cases there is no clear causal relationship between the *X* and *Y* variables, and the usual regression techniques recommended in statistics texts are not appropriate. In these situations Ricker (1973, 1984) recommended the use of a central trend line to describe the data more accurately.

Central trend lines may also be useful in ecological situations in which both the *X*- and *Y*- variables have measurement errors. Standard regression theory assumes that the *X*- variable is measured without error, but many types of ecological data violate this simple assumption (Ricker 1973). If the *X*-variable is measured with

error, estimates of the slope of the usual regression line will be biased toward a lower absolute value.

Figure 16.5 illustrates a regression between hind foot length and log body weight in snowshoe hares. There are two standard regressions that can be calculated for these data:

1. *Regression of hind foot length (Y) on log body weight (X)*; this regression minimizes the vertical deviations (squared) from the regression line.

2. *Regression of log body weight (Y) on hind foot length (X)*; this regression could alternatively be described as a regression of *X* on *Y* in terms of (1) above.
These two regressions are shown in Figure 16.5.

Ricker (1973) argues that a better description of this relationship is given by the central trend line called the *functional regression* or the *geometric mean regression* (GMR). The functional regression is simple to obtain, once the computations for the standard regression (1) have been made:

$$\hat{v} = \frac{\hat{b}}{r} = \sqrt{\frac{\hat{b}}{\hat{d}}} \qquad (16.34)$$

where $\hat{v}$ = Estimated slope of the geometric mean regression
$\hat{b}$ = Estimated slope of the regression of *Y* on *X*
r = Correlation coefficient between *X* and *Y*
$\hat{d}$ = Estimated slope of the regression of *X* on *Y*

and $\hat{v}$ has the same sign as *r* and $\hat{b}$. The term *geometric mean regression* follows from the second way of estimating *v* given in equation (16.34). Box 16.3 illustrates the calculation of a central trend line.

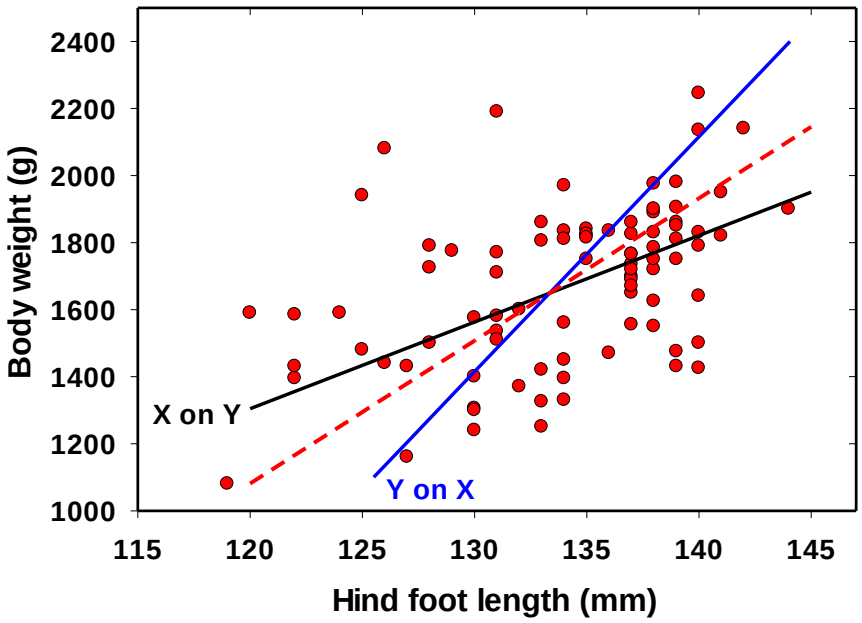

Figure 16.5 Regression of log body weight on hind foot length for snowshoe hares. Three regressions can be calculated for these data. The usual regression of *Y* on *X* is shown as a solid black line along with the less usual regression of *X* on *Y* (blue line). The functional or geometric mean regression (GMR) of Ricker (1973) is shown as the dashed red line.

Box 16.3 CALCULATION OF A GEOMETRIC MEAN REGRESSION

Garrod (1967) gives the following data for fishing effort and total instantaneous mortality rate for ages 6 through 10 for the Arcto-Norwegian cod fishery:

Year	Fishing effort	Total mortality rate
1950-51	2.959	0.734
1951-52	3.551	0.773
1952-53	3.226	0.735
1953-54	3.327	0.759
1954-55	4.127	0.583
1955-56	5.306	1.125
1956-57	5.347	0.745
1957-58	4.577	0.859
1958-59	4.461	0.942
1959-60	4.939	1.028
1960-61	6.348	0.635
1961-62	5.843	1.114
1962-63	6.489	1.492
Totals	60.500	11.524

Means	4.6538	0.8865

The total mortality rate should depend on the fishing effort, and there is a large measurement error in estimating both of these parameters. To calculate the GMR proceed as follows:

1. Calculate the usual statistical sums:

$$\sum X = 2.959 + 3.551 + 3.226 + \cdots = 60.500$$
$$\sum Y = 0.734 + 0.773 + 0.735 + \cdots = 11.524$$
$$\sum X^2 = 2.959^2 + 3.551^2 + 3.226^2 + \cdots = 298.40551$$
$$\sum Y^2 = 0.734^2 + 0.773^2 + 0.735^2 + \cdots = 10.965444$$
$$\sum XY = (2.959)(0.734) + (3.551)(0.773) + \cdots = 55.604805$$

2. Calculate the *sums of squares* and *sums of cross products*:

$$SS_x = \sum X^2 - \frac{\left(\sum X\right)^2}{n}$$
$$= 298.40551 - \frac{60.500^2}{13} = 16.847818$$

$$SS_Y = \sum Y^2 - \frac{\left(\sum Y\right)^2}{n}$$
$$= 10.965444 - \frac{11.524^2}{13} = 0.7498612$$

$$\text{Sum of cross-products} = \sum XY - \frac{\left(\sum X\right)\left(\sum Y\right)}{n}$$
$$= 55.604805 - \frac{(60.5)(11.524)}{13} = 1.9738819$$

3. Calculate the parameters of the *standard regression of Y on X*:

$$Y = a + bX$$

$$\text{Slope} = \hat{b} = \frac{\text{Sum of cross-products}}{\text{Sum of squares in } X}$$
$$= \frac{1.9738819}{16.847818} = 0.117159$$

$$Y \text{ intercept } = \hat{a} = \bar{y} - \hat{b}\,\bar{x}$$
$$= 0.8865 - (0.117159)(4.65385) = 0.34122$$

$$\text{Correlation coefficient } = r = \frac{\text{Sum of cross-products}}{\sqrt{(SS_x)(SS_y)}}$$
$$= \frac{1.9738819}{\sqrt{(16.847818)(0.7498612)}} = 0.55534$$

$$\text{Variance about regression} = s_{yx}^2$$

$$= \frac{\text{SS}_y - \left((\text{sum of cross-products})^2 \Big/ \text{SS}_x \right)}{n-2}$$

$$= \frac{0.7498612 - \left((1.9738819)^2 \Big/ 16.847818 \right)}{11}$$

$$= 0.0471457$$

$$\text{Standard error of slope} = s_b = \sqrt{\frac{s_{xy}^2}{\text{SS}_x}}$$

$$= \sqrt{\frac{0.0471457}{16.847818}} = 0.052899$$

4. Calculate the *geometric mean regression* from equations (16.34) and (16.35):

$$\text{Slope of GMR} = \hat{v} = \frac{\hat{b}}{r}$$

$$= \frac{0.117159}{0.55534} = 0.21097$$

$$Y \text{ intercept of GMR} = \hat{a}' = \bar{y} - \hat{v}\,\bar{x}$$

$$= 0.8865 - (0.21097)(4.6538) = -0.0953$$

The standard error of the slope of the GMR is the same as the standard error of $\hat{b}$.

Both regression lines can be plotted from the two points ($(\bar{x}, \bar{y})$ and the *Y*-intercept $(\hat{a}')$, i.e. the value of *Y* when *X* = 0.

Program EXTRAS (Appendix 2) can do these calculations for the geometric mean regression.

For the functional regression, the *y*-intercept is calculated in the usual way:

$$Y \text{ intercept} = \bar{y} - \hat{v}\,\bar{x} \qquad (16.35)$$

where $\bar{y}$ = Observed mean value of *Y*

$\bar{x}$ = Observed mean value of *X*

$\hat{v}$ = Estimated slope of the GMR from equation (16.34)

Ricker (1973) showed that the standard error of the slope of the GMR is the same as that of the slope $\hat{b}$ in a standard regression and that confidence limits for $\hat{v}$ can be obtained in the usual way from this standard error.

The central trend line described by the functional regression has three characteristics that are essential for any descriptive line (Ricker 1984):

1. The line must be *symmetrical*; thus, if *X* and *Y* are interchanged, there will be no change in the position of the line relative to the data.

2. The line must be *invariant to linear changes of scale* so that its position among the data points does not depend on whether inches or centimeters are used in measuring.

3. The line should be *robust* so that deviations from the statistical assumption of a bivariate normal distribution are not fatal.

Ordinary regressions discussed in all the statistics books are scale-invariant but they are not symmetrical nor are they robust (Schnute 1984). By contrast, the geometric mean regression has all of these traits.

Ricker (1984) and Jensen (1986) discuss several situations in which the functional regression is superior to ordinary regression procedures. Figure 16.6 illustrates a common ecological situation in which incomplete samples are taken from a population. The functional regression is obviously superior to the regression of *Y* on *X* in describing this relationship.

When should you use the geometric mean regression? Ricker (1973, 1984) has put together some recommendations for guidance that are summarized in Table 16.2. The decision on what type of regression to choose depends on the answer to three questions:

1. Are there serious measurement errors in the *X*- or *Y*- variables? Or are measurement errors quite small?

2. Is the statistical population well-defined so that it can be sampled randomly?

3. Does the statistical population show a frequency distribution that is approximately bivariate normal?

TABLE 16.2 RICKER'S (1984) RECOMMENDATIONS REGARDING THE TYPE OF LINEAR REGRESSION TO USE FOR DIFFERENT KINDS OF DATA[a]

A. No serious measurement errors in the X or Y variables

B. *Random sample from a bivariate normal population:*
- use GMR to describe general trend
- use ordinary regression for prediction.

B'. *Random sample from a population that is not bivariate normal:*
- use GMR or Schnute (1984) method for description and for prediction.

B". *Nonrandom sample:*
- use GMR method for description and for prediction.

A'. *Measurement errors in Y but not in X*
- use ordinary regression for description and for prediction.

A". *Measurement errors in both Y and X variables*

C. Error variances available for both *X* and *Y*:
- use Jolicoeur and Heusner (1971) method

C'. Error variances of *X* or *Y* unknown:
- use GMR method for description and prediction

[a] GMR = geometric mean regression of RIcker (1973); ordinary regression = least-squares regression of *Y* on *X*.

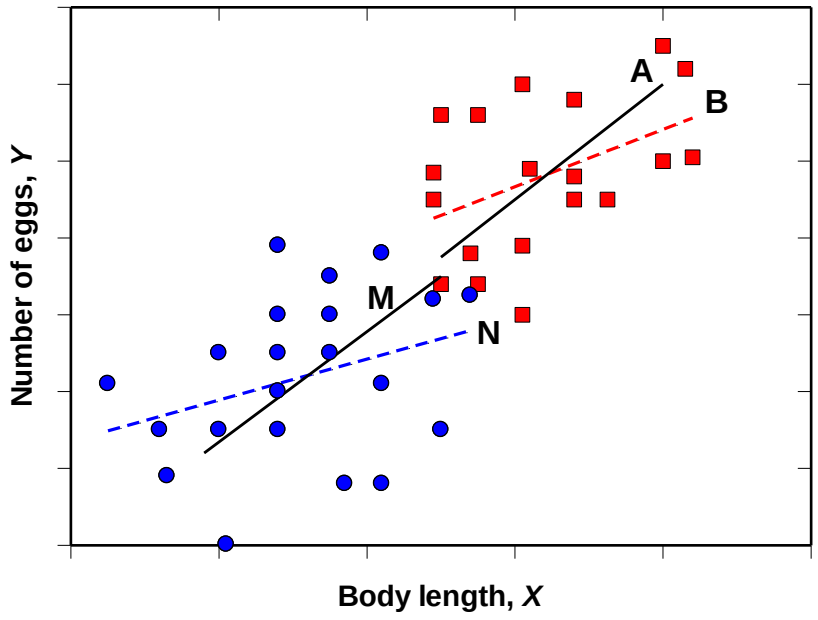

Figure 16.6 Hypothetical example of two incomplete samples from a population in which there is a linear regression between *X* and *Y*. Lines B (red) and N (blue) (dashed lines) show the usual regression of *Y* on *X* for the two groups of data indicated by the two symbols. If you used these two regressions you would probably conclude that these data come from two distinct statistical populations. Solid lines A and M show the geometric mean

regressions for the two groups and suggest strongly that there is only one statistical population with a common relationship between body length and egg production. The GMR is a better way of describing this relationship for this population. (Modified from Ricker, 1984)

The main consideration is whether you are trying to predict the *Y*-variable with minimal error, or whether you wish to fit a functional relationship to show the trend. The decision as to which method to use in linear regression is not serious if there is a tight correlation between the *X* and *Y* variables (Jensen 1986). But in much ecological data, correlations of 0.5 to 0.7 occur and it becomes important to decide which of the available regression lines is most appropriate to the data at hand.

Ricker (1973, 1984) gives a detailed discussion of the linear regression problem in ecological research and provides many examples from fisheries work in which the GMR is the appropriate regression to use. Sokal and Rohlf (1995 p. 541) also discuss this regression problem.

Program EXTRAS (Appendix 2, page 000) can do these calculations for a geometric mean regression.

16.4 MEASURING TEMPORAL VARIABILITY OF POPULATIONS

Some populations fluctuate greatly in abundance from year to year and others are more stable in numbers. How can you measure this variability? This question arises in a number of ecological contexts, from theories that predict that species of smaller body size have more variable populations than larger species (Gaston and Lawton 1988), to suggestions that endangered species that fluctuate more are prone to extinction (Karr 1982). Given a set of population estimates for a species, it would seem to be a simple matter to estimate the variability of these estimates in the standard statistical manner. Unfortunately, this is not the case (McArdle *et al.* 1990), and it is instructive to see why.

Since populations typically change in size by proportions, rather than by constant amounts, the first suggestion we might make is to log-transform the density estimates. We can then use the standard deviation to measure variability with the usual formula first suggested by Lewontin (1966):

$$s = \sqrt{\frac{\sum_{t=1}^{k}\left(\log N_t - \log \bar{N}\right)^2}{k-1}} \qquad (16.36)$$

where s = Standard deviation of log abundances = index of variability
N_t = Population size at time t
$\log \bar{N}$ = Mean of the logarithms of population size = $\sum(\log N_t)/n$
k = Number of observations in the time series (sample size)

This is the most commonly used measure of variability of populations and it has serious problems which suggest it should never be used (McArdle *et al.* 1990, Stewart-Oaten *et al.* 1995). The most obvious problem is what to do when the population estimate is zero (McArdle and Gaston 1993), since the log of zero is not defined. Most authors sweep over this difficulty by adding a constant to the population estimate (e.g. $N + 1$) for each sample. But the value of s in fact changes with the value of the constant added and this is not satisfactory. A second problem is that the variance of populations is typically related to the mean (c.f. Taylor's Power Law, page 000) and unless the slope of Taylor's Power Law is 2, the log transformation is not entirely effective in removing this relationship. We need to search elsewhere for a better index of variability.

One measure that is not affected by zeros is the coefficient of variation. It is widely used in biometrics as a scale independent measure of variability, and is defined in the usual way:

$$\mathrm{CV}(N) = \frac{\text{standard deviation of } N}{\bar{N}} \qquad (16.37)$$

where $\bar{N}$ = Mean population size for a series of time periods

The coefficient of variation is similar to the s statistic in being independent of population density if the slope of Taylor's Power Law is 2. Note that the usual measure of the coefficient of variation defined in equation (16.37) is slightly biased (Haldane 1955) and should be corrected by:

$$\mathrm{CV}(N) = \left(1 + \frac{1}{4n}\right)\left(\frac{\text{standard deviation of } N}{\bar{N}}\right) \qquad (16.38)$$

where *n* is sample size.

The recommended procedure to measure population variability is as follows:

1. Plot the regression of the log of the coefficient of variation (*Y*) vs. the log of population density (*X*) for the species of interest. This plot uses the same information as Taylor's Power Law, and thus requires several samples of populations that have different average densities (a great deal of data!).
2. If the variability does not depend on the mean (see Figure 16.7) you can use either the coefficient of variation or the standard deviation of log abundances (if no zeros) to measure variability in this population. If variability does not depend on mean density, Taylor's Power Law will have a slope of 2.0 (McArdle *et al.* 1990).
3. If variability increases, or decreases, with mean density (Figure 16.7) you cannot compare population variability for this species with that of other species. A population at high density will of necessity be more or less variable than one at low density. Instead of comparing population variability you should turn your attention to interpreting the pattern of relationships illustrated in Figure 16.7. Two situations are common.

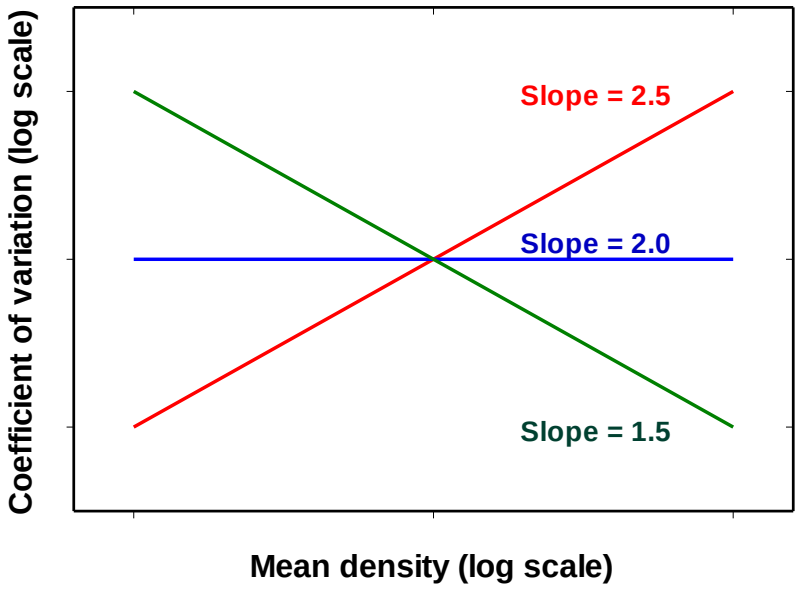

Figure 16.7 The problem of estimating variability in a population. The relationship between the coefficient of variation (log scale) and the mean density of the population (log scale) that

is expected under different slopes from Taylor's Power Law. Only if the Taylor slope = 2 is the measure of variability independent of population density so that you could use the coefficient of variation as a measure of population variability.

One problem in the comparison of temporal population variabililties is that extreme events can exert a large effect on the estimate of variability (Heath 2006). For example, a large population crash in one year will affect estimates of population variability even if all the other years in the time series have exactly identical population sizes. Heath (2006) proposed a simple measure of population variability that can be applied to time series in which the population abundance is estimated at a fixed time interval (e.g. yearly). A time series of length *n* will have the number of possible pairwise combinations abundances given by:

$$C = \frac{n(n-1)!}{2} \tag{16.39}$$

Given this many possible pairwise combinations of abundances, $z = 1 \ldots C$, we calculate for each of the *z* pairs:

$$D(z) = 1 - \left(\frac{\min\left[z_i, z_j \right]}{\max\left[z_i, z_j \right]} \right) \tag{16.40}$$

And after calculating all these possible combinations, we average them to obtain a measure of population variability:

$$PV = \frac{\sum_{i=1}^{C} D(z)}{C} \tag{16.41}$$

where

$PV =$ estimate of population variability
$C =$ number of possible pairwise combinations (eq. 16.40)
$D(z) =$ array of all possible differences among C pairs

PV varies from 0 to 1.0 where 0 is complete stability and values near 1.0 indicate great instability.

There are two advantages to using PV as a measure of temporal variability. If a time series has an estimate of 0 abundance in it, the log transformed standard deviation (eq. 16.36) is undefined and cannot be used. The coefficient of variation

(eq. 16.38) is sensitive to extreme, rare changes in population size, but PV is much less sensitive (Heath 2006). Finally, simulation studies suggest PV performs better than the other two estimates of variability in small samples.

Comparing Temporal Variability

In this case a number of independent sites are sampled repeatedly through time and the samples are taken far enough apart in time that they are independent statistically. For each site calculate the mean density and the coefficient of variation over time. Plot the data as in Figure 16.7 and calculate Taylor's Power Law (page 000).

(a) if $b = 2$, the variability over time of the population is constant from site to site.

(b) if $b > 2$, temporal variability is greater at good sites (= high density sites).

(c) if $b < 2$, temporal variability is greater at poor sites, and low density populations are subject to more fluctuations than those in good sites.

Comparing Spatial Variability

In some cases interest centers on a series of populations in different sites (e.g. Perry 1988). In this case, as in the previous case, a number of independent sites are sampled repeatedly through time and the samples are taken far enough apart in time and space that they are independent statistically. Calculate the mean and coefficient of variation over all the sites in a given year. Plot the data as in Figure 16.7 and calculate Taylor's Power Law for the spatial data..

(a) if $b = 2$, the variability over space of the population is constant from year to year.

(b) if $b > 2$, spatial variability is greater in good years (= high density years).

(c) if $b < 2$, spatial variability is greater in bad years, and in low density years populations are more variable spatially. In good years densities tend to even out in space.

Clearly if we sample in the same sites at the same times in the same set of years, we can compare both temporal and spatial variability of our population.

Ecologists also wish to compare variability between different species, for example to determine if bird populations fluctuate more than mammal populations. The same principles given above for single species comparisons apply here. For simplicity consider the case of two species. If for both of the species the slope of Taylor's Power Law is 2 for both temporal and spatial variability, then it is possible to use the coefficient of variation to compare population variability in the two species. Much background information is clearly needed before you can make this assumption. If you have only two sets of samples, one for each of the species, you cannot make any legitimate statement about relative variability. If the slope of Taylor's Power Law is not 2, comparisons are more difficult and must be limited to the observed range of densities (Figure 16.8). If many species are to be compared, one can only hope that the Taylor's slope is 2, and that you do not have to deal with a situation like that shown in Figure 16.8. Ecologists need to pay attention to these details to make proper statistical statements about variability of populations. Recent reviews have indeed suggested that most of the existing literature comparing population variability is invalid and the conclusions artifacts of sampling (McArdle and Gaston 1992, Stewart-Oaten *et al.* 1995). It is important that we use better methods for future research on relative variability of populations.

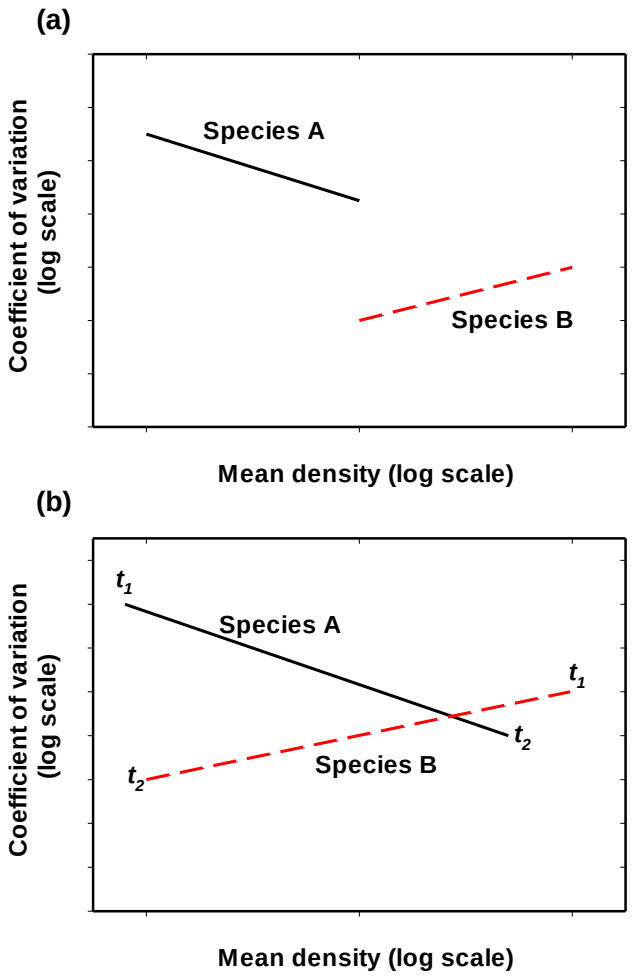

Figure 16.8 Problems of comparing temporal population variability between species. In all cases the slope of Taylor's Power Law differs from 2.0. (a) Different slopes are present but species A is always more variable than species B, so they may be compared statistically. (b) If the species abundances are correlated, the species cannot be compared directly. In this instance the two species are negatively correlated so that when the density of species A goes up from time t_1 to time r_2 , the density of species B goes down. No general conclusions can be drawn from (b). (After McArdle *et al.* 1990).

16.5 JACKKNIFE AND BOOTSTRAP TECHNIQUES

The advent of computers has opened up a series of new statistical techniques which are of great importance ot ecologists because they release us from two restrictive assumptions of parametric statistics: (1) that data conform to a normal frequency distribution, and (2) that statistical measures must have good theoretical properties so that confidence limits can be derived mathematically. The price of giving up these

traditional assumptions is a massive increase in computations so that a computer is essential for all these new methods (Diaconis and Efron 1983).

Two computer-intensive methods have been particularly important in ecological statistics: the *jackknife* and the *bootstrap* and I will describe each of these briefly. We have already used these methods in Chapter 13 to estimate species richness (page 000) and pointed out their utility for estimating niche overlap in Chapter 14 (page 000). Ecologists should realize these methods exist; they are sufficiently complicated however that you should consult a statistician before applying them uncritically to your data.

The *jackknife* technique was first suggested by Tukey (1958). We would like to know how much better our estimate would be if we had one more sample. But we do not have any more samples, so we ask the converse question — how much worse would we be if we had one less sample. Beginning with a set of *n* measurements, the jackknife is done as follows:

1. *Recombine the original data*: this is done by omitting one of the *n* replicates from the jackknife sample.

2. *Calculate pseudo-values of the parameter of interest* for each recombining of the data:

$$\Phi_i = nS - (n-1)S_T \qquad (16.42)$$

where Φ_i = Pseudo-value for jackknife estimate *i*
n = Original sample size
S = Original statistical estimate
S_T = Statistical estimate when original value *i* has been discarded from sample *i* = sample number (1, 2, 3 ... *n*)

3. *Estimate the mean and standard error of the parameter of interest* from the resulting pseudo-values.

The jackknife technique has been applied to several ecological parameters. One good example is the estimation of population growth rates (Meyer *et al.* 1986). Population growth rates can be estimated from the characteristic equation of Lotka:

$$1 = \sum_{x=0}^{\infty} e^{-rx} l_x m_x \qquad (16.43)$$

where r = Per capita instantaneous rate of population growth
l_x = Probability of surviving to age x
m_x = Fecundity at age x

Meyer *et al.* (1986) had data on 10 individual *Daphnia pulex* females giving the reproductive output and the age at death. To calculate the jackknife estimate of r they proceeded as indicated above, discarding in turn one female from each jackknife sample, to generate 10 pseudo-values of r. By averaging these 10 values and obtaining a standard error from them, Meyer *et al.* (1986) could estimate the r value for each population of *Daphnia* studied.

The *bootstrap* technique was developed by B. Efron in 1977 (Efron 1982). The bootstrap method asks what another sample of the same size would look like, if indeed we had one. But we do not, and so we pretend that the sample we have is a universe and we sample from it with replacement to estimate the variability of the sampling process. The bootstrap method follows the same general procedure as for the jackknife:

1. *Recombine the original data*: the original data of n measurements are placed in a pool and then n values are sampled *with replacement*. Thus any measurement in the original data could be used once, twice, several times, or not at all in the bootstrap sample. Typically one repeats this bootstrap sampling at least 500 times and often several thousand times.

2. *Calculate the parameter of interest* from each bootstrap sample.

3. *Estimate the mean and standard error of the parameter of interest* from the replicate bootstrap estimates.

Bootstrap estimates of parameters are known to be biased, If the true population value for the mean (for example) is μ, and if the observed mean of the whole original sample is $\bar{x}_S$, the bootstrap estimate of the mean $\bar{x}_B$ will be biased because it estimates $\bar{x}_S$ rather than μ. The bias of $\bar{x}_B$ is defined as:

$$\text{Bias}(\bar{x}_S) = \bar{x}_S - \mu \qquad (16.44)$$

which can be estimated by:

$$\text{Bias}(\bar{x}_S) = \bar{x}_B - \bar{x}_S \qquad (16.45)$$

Because of this bias, bootstrap estimates are usually bias-adjusted by combining equations (16.41) and (16.42):

$$\text{Bias-adjusted bootstrap mean} = 2\bar{x}_S - \bar{x}_B \qquad (16.46)$$

where $\bar{x}_S$ = Observed mean of original sample
$\bar{x}_B$ = Bootstrap estimate of the mean

The precision of a bootstrap estimate will depend on how many times the original data are randomly recombined, and the bootstrap estimate will converge on a stable estimate as the number of recombinations becomes large. Note that repeated bootstrap calculations performed with the same number of recombinations and the same original data will vary somewhat because the items randomly chosen differ in every sample of recombinations.

Confidence limits for bootstrap estimates can be obtained in the usual way with standard errors calculated from the replicate bootstrap samples. Alternatively, Efron (1982) suggested measuring confidence limits directly from the frequency distribution of bootstrap estimates - if a large number (n > 500) of bootstrap estimates are done, the 2.5^{th} and 97.5^{th} percentile values of this frequency distribution delimit a confidence belt of 95%. This empirical approach – the *percentile method* of Efron (1982) – may produce somewhat biased confidence limits when sampling distributions of bootstrap estimates are skewed. Efron (1982) discusses ways of correcting for this bias.

There is as yet no general agreement on when jackknife estimates are better and when bootstrap methods are better, and more empirical work is needed for ecological measures. Meyer *et al.* (1986) found that the jackknife and the bootstrap were equally effective for estimating population growth rates (*r*) for *Daphnia pulex*. Since jackknife estimates require much less computing than bootstrap estimates

(which typically need 500 or 1000 replicates), it may be more useful to use jackknife procedures when computing time is limited.

We have already seen cases in which jackknife estimators have been useful in ecological statistics (e.g. Chapter 14, pg. 000, for species richness measures). There is no doubt that these computer-intensive techniques will be used more and more to improve the estimation of difficult ecological parameters.

Box 16.4 illustrates the use of the bootstrap method to estimate niche breadth in mammals. Virtually all the good statistical packages now contain routines to calculate jackknife and bootstrap estimates to estimate confidence intervals for variables that do not fit the standard normal distribution.

Box 16.4 BOOTSTRAP CALCULATION OF FOOD NICHE BREADTH FOR COYOTES

Mark O'Donoghue measured the frequency of 8 food sources for coyotes in the southwestern Yukon, Canada, in the winter of 1992-93 as follows:

	Number of scats with species	Proportions p_i
Snowshoe hares	91	0.421
Red squirrels	17	0.079
Arctic ground squirrels	21	0.097
Field voles	54	0.250
Red-backed voles	14	0.065
Least chipmunk	3	0.014
Moose	7	0.032
Willow ptarmigan	9	0.042
Totals	216	1.000

1. We calculate first the observed food niche breadth.

Levin's Measure of Niche Breadth

For the coyote data, from equation (14.1):

$$B = \frac{1}{\sum p_j^2}$$

$$= \frac{1}{0.421^2 + 0.079^2 + 0.097^2 + 0.250^2 + 0.065^2 + 0.014^2 + \cdots}$$
$$= \frac{1}{0.26267} = 3.807$$

2. We next resample the observed distribution of 216 observations at random with replacement to obtain another sample of observations with the same sample size. We obtained the following data in this first random resampling:

	Number of scats	Proportions
Snowshoe hares	85	0.394
Red squirrels	19	0.088
Arctic ground squirrels	26	0.120
Field voles	61	0.282
Red-backed voles	11	0.051
Least chipmunk	4	0.019
Moose	6	0.028
Willow ptarmigan	4	0.018
Totals	216	1.000

Use these resampled data to estimate niche breadth in the same manner as previously:

$$B = \frac{1}{\sum p_j^2}$$

$$= \frac{1}{0.394^2 + 0.088^2 + 0.120^2 + 0.282^2 + 0.051^2 + 0.019^2 + \cdots}$$
$$= \frac{1}{0.26076} = 3.835$$

3. We repeat this procedure in the computer 1000 times to obtain 1000 resampled estimates of niche breadth. From this sample of 1000 we calculate the mean niche breadth in the usual way to obtain:

$$\bar{x} = \frac{\sum B}{1000} = 3.816$$

$$\text{Bias-adjusted bootstrap mean} = 2\bar{x}_S - \bar{x}_B$$
$$= 2(3.807) - 3.816 = 3.798$$

Since we have generated 1000 bootstrap estimates (in the computer) we can use the

percentile method of Efron (1982) to provide estimates of confidence limits. For 95% confidence limits we rank all the 1000 estimates and locate the 2.5% and the 97.5% points in the frequency distribution. These will be (in ranked form) the 25^{th} and the 975^{th} value, and for these particular data these were found to be:

2.5^{th} percentile $= 3.772$
97.5^{th} percentile $= 3.829$

and these values can be used as empirical estimates of the 95% confidence limits for this estimate of niche breadth.

Bootstrap estimates can be obtained from Microsoft EXCEL as well as a variety of statistical packages you can search for on the web..

16.6 SUMMARY

Transformations are essential for much ecological data that are not normally distributed and in which variances are not equal in all groups. Traditional transformations, like the logarithmic and square root, can be applied to known classes of data or used *ad hoc* to improve the fit of data to the assumptions of parametric statistics. Alternatively, one can use a general purpose transformation like the Box-Cox transformation to produce a specially-tailored exponential transformation for a particular data set.

Repeatability is a statistical measure of how similar duplicate measurements on the same individual are, in relation to differences among individuals. It can be a useful measure in quantitative genetics, where it is an upper bound on heritability, and in data analysis in which one is looking for the best way of measuring complex traits.

Linear regression data do not often satisfy the standard assumption that the *X* variable is measured without error. Central trend lines may be a better description of the regression and give more accurate predictions. The *geometric mean regression* is one simple central trend line that should be used on more ecological data.

Estimating the relative variability of populations in time and space would appear to be a simple task but is fraught with statistical problems. The coefficient of variation should be used for estimates of variability rather than the standard

deviation of log abundances. Only if the slope of Taylor's Power Law is 2 can one make simple comparisons among populations or between species.

Computer-intensive estimates of statistical parameters can be achieved in two general ways using *jackknife* or *bootstrap* methods. By discarding some of the original data or by sampling the original data with replacement, one can calculate nearly unbiased estimates of complex ecological parameters like the rate of population growth, along with confidence limits. This new area of statistics holds much promise for dealing with ecological parameters that have been intractable to normal statistical methods.

SELECTED REFERENCES

Berry, D.A. 1987. Logarithmic transformations in ANOVA. *Biometrics* **43**: 439-456.

Dochtermann, N.A., and Peacock, M.M. 2010. Differences in population size variability among populations and species of the family Salmonidae. *Journal of Animal Ecology* **79**(4): 888-896.

Efron, B. & Tibshirani, R.J. 1993. *An Introduction to the Bootstrap.* Chapman and Hall, New York.

Heath, J.P. 2006. Quantifying temporal variability in population abundances. *Oikos* **115**(3): 573-581.

Jensen, A.L. 1986. Functional regression and correlation analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **43**: 1742-1745.

Lessells, C.M. & Boag, P.T. 1987. Unrepeatable repeatabilities: a common mistake. *Auk* **104**: 116-121.

McArdle, B.H., Gaston, K.J., and Lawton, J.H. 1990. Variation in the size of animal populations: patterns, problems and artifacts. *Journal of Animal Ecology* **59**: 439-454.

Meyer, J.S., Ingersoll, C.G., McDonald, L.L. & Boyce, M.S. 1986. Estimating uncertainty in population growth rates: jackknife vs. bootstrap techniques. *Ecology* **67**: 1156-1166.

Ricker, W.E. 1984. Computation and uses of central trend lines. *Canadian Journal of Zoology* **62**: 1897-1905.

Smith, R.J. 2009. Use and misuse of the reduced major axis for line-fitting. *American Journal of Physical Anthropology* **140**(3): 476-486.

Stewart-Oaten, A., Murdoch, W.W. & Walde, S.J. 1995. Estimation of temporal variability in populations. *American Naturalist* **146**: 519-535.

QUESTIONS AND PROBLEMS

16.1. Twelve samples of forest soil were cultured to estimate the abundance of a species of protozoa in the soil, with these results:
56, 160, 320, 640, 900, 900, 5200, 5200, 7200, 20800, 20800 and 59000.
Do these data require transformation? What transformation do you recommend?

16.2. Aerial counts of caribou from photographs taken sequentially produced these counts for four adjacent flight paths:

Flight Path

A	B	C	D
240	80	10	72
600	34	54	35
50	58	250	27
135	90	4	100
82	73	70	660
98	430	180	20
32	33	150	160
220	250	92	90

(a) Do these data require transformation?

(b) If a logarithmic transformation is to be used, what value of the constant *c* would you recommend?

16.3. J.N.M. Smith took three measurements repeatedly on bill size in song sparrows (*Melospiza melodia*) on Mandarte Island, B.C. with these results:

Bird no.	Bill length (mm)	Bill depth (mm)	Bill width (mm)
10993	8.5	6.2	6.6
	8.5	6.1	6.5
	8.5	6.1	6.7

	8.1	6.0	6.6
	8.6	6.2	6.5
	8.8	6.2	6.7
10994	8.0	5.9	6.6
	8.4	5.6	6.3
	8.5	5.7	6.3
10996	8.5	6.0	6.8
	8.5	5.8	6.5
	9.0	6.0	6.8
10999	8.2	5.6	6.5
	8.4	5.8	6.3
11000	9.1	5.9	6.9
	8.7	5.8	6.6
	8.7	5.8	6.5
10982	7.7	5.4	5.9
	7.9	5.3	6.0
	7.9	5.5	6.0

Calculate repeatability for each of these measures. Is one measurement better than the others, as far as repeatability is concerned?

16.4. Standard length (cm) and total weight (g) were measured for 34 herring caught in a purse seine off the B.C. coast. Age was determined from growth rings in the scales.

Age 1+		Age 2+		Age 3+		Age 4+		Age 5+	
SL	WT	SL	WT	SL	WT	SL	WT	SL	WT
17.2	74	19.7	119	19.3	115	20.8	159	22.0	177
16.8	62	19.0	106	19.7	125	20.1	120	22.0	164
16.2	64	18.8	101	19.2	118	19.5	115	20.7	146
16.3	52	18.2	92	19.7	118	21.0	145		
16.1	58	18.1	92	19.8	124	21.0	141		
16.4	48	18.8	91	19.3	115	21.5	156		

16.1	57	19.6	117	20.6	142	21.7	146		
16.4	55	18.7	100	19.7	119				
17.3	73	19.3	106	20.2	136				
		18.1	82	19.8	111				
		19.1	100	19.1	110				
		17.0	75	20.9	143				
		18.0	96	19.3	110				
		19.3	110	17.7	91				
		17.9	85	20.1	134				
				19.7	120				

Data from Ricker (1975, page 213).

(a) Is length related to weight in these fish? Is a transformation necessary to do a linear regression?

(b) Calculate a ordinary regression and a functional regression for these data. Which is more appropriate if you wish to estimate weight from length?

16.5. Calculate repeatability for the data on aerial counts of moose in Table 16.1.

16.6. Compare the temporal variability of these 10 grey-sided vole (*Clethrionomys rufocanus)* populations on Hokkaido, Japan (Saitoh *et al.* 1997): Values are number caught per 100 trap-nights in each area in the autumn of each year.

Area

	1	2	3	4	5	6	7	8	9	10
1963	9.75	7.91	0.42	3.14	1.78	1.44	1.44	1.78	2.80	0.76
1964	9.32	6.21	8.25	7.23	3.48	8.93	1.10	7.23	6.89	2.29
1965	6.43	5.18	9.27	11.65	8.93	6.21	9.27	14.38	11.65	5.53
1966	6.97	7.57	3.82	12.68	10.29	0.08	1.78	0.76	2.80	3.48
1967	16.25	12.34	4.67	9.95	23.23	6.55	1.78	3.14	6.55	3.31
1968	5.87	0.42	1.10	4.16	0.42	2.46	6.89	1.10	10.29	1.44
1969	19.94	19.15	19.74	13.70	24.59	2.29	10.63	2.80	6.21	4.42
1970	17.78	18.46	16.88	3.14	14.38	12.34	36.17	14.38	9.61	16.08
1971	12.11	1.44	19.83	13.70	9.27	6.21	17.44	2.80	4.84	7.34
1972	18.69	16.06	18.12	5.18	19.49	4.84	6.21	0.76	3.48	6.55
1973	10.29	13.70	16.08	7.91	7.91	0.76	9.27	13.70	3.48	7.23

1974	18.92	16.74	6.89	4.84	5.53	4.16	9.95	8.25	11.31	6.72
1975	7.34	1.44	2.46	3.48	2.80	0.08	3.48	0.08	3.48	4.67
1976	20.28	8.93	9.61	11.65	16.74	4.16	8.25	2.80	5.18	1.27
1977	9.67	17.00	4.00	2.00	23.00	42.00	14.00	14.00	14.00	17.25
1978	23.80	30.00	11.33	39.00	10.00	45.00	11.00	30.00	12.50	26.67
1979	14.40	0.00	5.67	13.00	4.00	3.00	3.00	1.00	1.00	2.33
1980	12.00	3.00	4.33	6.00	7.00	1.00	3.00	6.00	1.50	3.00
1981	17.00	17.00	16.67	19.00	24.00	12.00	9.00	31.00	13.50	9.00
1982	5.80	5.00	16.00	10.00	14.00	8.00	16.00	16.00	3.00	9.33
1983	8.80	8.00	4.00	8.00	21.00	2.00	3.00	0.00	8.00	9.33
1984	5.20	0.00	2.33	0.00	3.00	1.00	0.00	0.00	1.50	0.33
1985	21.60	10.00	14.67	8.00	19.00	12.00	11.00	5.00	23.00	10.67
1986	1.80	3.00	2.00	2.00	6.00	3.00	16.00	6.00	19.50	16.33
1987	4.50	0.00	1.00	4.00	0.00	0.00	0.00	0.00	0.00	0.00
1988	17.75	10.00	10.50	14.00	8.00	10.00	11.00	7.00	12.00	9.67
1989	2.50	7.00	5.50	1.00	16.00	21.00	6.00	16.00	19.00	5.67
1990	6.25	6.00	0.00	6.00	1.00	0.00	3.00	1.00	0.00	2.00
1991	5.33	6.00	2.00	16.00	2.00	1.00	0.00	0.00	9.00	0.00
1992	2.00	4.00	1.00	0.00	2.00	1.00	1.00	1.00	4.00	0.00

(a) What is the slope of Taylor's Power Law for this population?

(b) Is population 1 more variable than population 4?