Res Popul Ecol (1999) 41:229–234

ORIGINAL ARTICLE

Kohji Yamamura

Transformation using (x + 0.5) to stabilize the variance of populations

Received: October 16, 1998 / Accepted: June 10, 1999

Abstract Transformation is required to achieve homoscedasticity when we perform ANOVA to test the effect of factors on population abundance. The effectiveness of transformations decreases when the data contain zeros. Especially, the logarithmic transformation or the Box–Cox transformation is not applicable in such a case. For the logarithmic transformation, 1 is traditionally added to avoid such problems. However, there is no concrete foundation as to why 1 is added rather than other constants, such as 0.5 or 2, although the result of ANOVA is much influenced by the added constant. In this paper, I suggest that 0.5 is preferable to 1 as an added constant, because a discrete distribution defined in $\{0, 1, 2, \ldots\}$ is approximately described by a corresponding continuous distribution defined in $(0, \infty)$ if we add 0.5. Numerical investigation confirms this prediction.

Key words ANOVA · Box–Cox transformation · Heteroscedasticity · Iwao's $m^* - m$ regression · Taylor's power law

Introduction

Many works have been developed about which measure should be used to describe the variability of populations (Williamson 1984; McArdle et al. 1990; Gaston and McArdle 1993; Leps 1993; McArdle and Gaston 1993, 1995). If we want to analyze the cause of population dynamics, a logarithmic scale is preferable in many cases, because mortality, as well as reproduction, is a multiplicative process. Life table analyses such as key-factor/key-stage analyses adopted a logarithmic scale for this reason (Yamamura 1999). One of the difficulties of logarithmic scale is that we cannot calculate the logarithm if the data contain zeros. In such a case, $\log_e(x + 1)$ or $\log_{10}(x + 1)$ is traditionally used, where x is the number of individuals. However, the variability of $\log_e(x + 1)$ is much different from that of $\log_e(x)$ when x is small. Hence, McArdle and Gaston (1992) recommended the coefficient of variation (CV) as a measurement of population variability, because CV performs the same way as the standard deviation of $\log_e(x)$ and is unaffected by zeros.

The logarithmic transformation is also frequently used to achieve homoscedasticity or stability of variance when we perform ANOVA to test the effect of factors on the population abundance. If the data contain zero, 1 is traditionally added to each data or only to the zeros. However, there is no concrete foundation as to why 1 is added rather than another constant, such as 0.5 or 2, although the result of ANOVA is much influenced by the added constant. In this article, I first summarize the procedure to determine the transformation formulae to stabilize the variance of population counts. Then, I suggest that 0.5 is a reasonable choice as the added constant. Numerical investigation is also conducted to determine an appropriate constant.

Heteroscedasticity in populations

The variance of population increases with increasing mean. Bliss (1941) suggested two equations to describe the heteroscedasticity:

$$s^2 = gm + hm^2 \tag{1}$$
$$s^2 = am^b \tag{2}$$

where m and s^2 are the mean and variance of the number of individuals in a sample, and a, b, g, and h are constants. The general applicability of Eqs. 1 and 2 were first shown by Iwao (1968), based on his $m^* - m$ regression, and by Taylor (1961), based on his power law, respectively. A considerable amount of controversy has been held about which of the two is superior as an ecological model (Iwao and Kuno 1971; Taylor et al. 1978; Taylor 1984; Itô and Kitching 1986; Kuno 1991; Routledge and Swartz 1991; Perry and Woiwod 1992). For the practical purpose of description, however,

K. Yamamura
Laboratory of Population Ecology, National Institute of Agro-Environmental Sciences, 3-1-1 Kannondai, Tsukuba 305-8604, Japan
Tel. +81-298-38-8313; Fax +81-298-38-8199
e-mail: yamamura@niaes.affrc.go.jp

both equations fit the data equally well in most cases, and hence I later use both equations to investigate the effect of added constant on the stabilization of variance.

Derivation of the transformation formula

Taylor series expansions

Let us assume that the variance of a variable x is given by a function of the mean $d(m)$. Let $f(x)$ be a function of x. Using Taylor series expansions around the mean, m, we obtain:

$$f(x) = f(m) + f'(m)(x - m) + \ldots \tag{3}$$

where $f'(m)$ is the first derivative of $f(x)$ evaluated at $x = m$. By squaring the above equation, we obtain an approximation of the variance of $f(x)$:

$$\begin{aligned} \mathrm{V}[f(x)] &= \mathrm{E}\left\{[f(x) - \mathrm{E}[f(x)]]^2\right\} \approx \mathrm{E}\left\{[f(x) - f(m)]^2\right\} \\ &\approx [f'(m)]^2 \mathrm{E}\left\{(x - m)^2\right\} = [f'(m)]^2 d(m) \end{aligned} \tag{4}$$

where E and V indicate the expectation and variance, respectively. This method to obtain the variance is usually called the delta method because of the reliance upon first derivatives (Stuart and Ord 1994, p 350). Our present concern is to find the function $f(x)$ that yields a constant variance irrespective of m. Then, we obtain from Eq. 4:

$$f(x) = \int \frac{w}{\sqrt{d(x)}}\, dx \tag{5}$$

where w is an arbitrary constant (Beall 1942; Bartlett 1947). If we know the form of $d(x)$ beforehand, therefore, we can derive a transformation formula using Eq. 5. When the distribution of x is a Poisson distribution, for example, we have $d(m) = m$, and hence we obtain the transformation formula, $f(x) = \sqrt{x}$. If the coefficient of variation (CV) of the distribution is constant, we obtain $f(x) = \log_e(x)$, because $d(m)$ is proportional to m^2. Iwao and Kuno (1968) derived the transformation formulae based on Eq. 1:

$$\left.\begin{aligned} f(x) &= \sqrt{x} && (g > 0,\ h = 0) \\ f(x) &= \log_e(x) && (g = 0,\ h > 0) \\ f(x) &= \sin^{-1}\left(\sqrt{-\frac{hx}{g}}\right) && (g > 0,\ h < 0) \\ f(x) &= \log_e\left(\sqrt{\frac{hx}{g}} + \sqrt{\frac{hx}{g} + 1}\right) && (g > 0,\ h > 0) \end{aligned}\right\} \tag{6}$$

Bliss (1941) derived the transformation formulae based on Eq. 2:

$$\left.\begin{aligned} f(x) &= x^{1-b/2} && (b \neq 2) \\ f(x) &= \log_e(x) && (b = 2) \end{aligned}\right\} \tag{7}$$

Equation 7 includes the square root transformation as its special case of $b = 1$.

Box–Cox transformation

Box and Cox (1964) proposed a procedure for determining a transformation formula, which is applicable when we do not know the form of $d(x)$ beforehand. They used a modified form of Eq. 7:

$$\left.\begin{aligned} f(x) &= \frac{(x^\lambda - 1)}{\lambda} && (\lambda \neq 0) \\ f(x) &= \log_e(x) && (\lambda = 0) \end{aligned}\right\} \tag{8}$$

This transformation is continuous around $\lambda = 0$, although Eq. 7 is discontinuous around $b = 2$. Hence, we can obtain a series of transformations, including $\sqrt{x}$ and $\log_e(x)$, by changing λ continuously. They estimated the parameter λ by the maximum-likelihood method based on the assumption that the distribution after the transformation follows a normal distribution. To obtain the estimate, the working variable, y, is first calculated:

$$y = \frac{x^\lambda - 1}{\lambda G^{\lambda - 1}} \tag{9}$$

where G is the geometric mean of x. ANOVA is performed for this working variable. Box and Cox (1964) showed that the maximum-likelihood estimate of λ coincides to the λ that minimizes the residual sum of squares in this ANOVA. Hence, we can easily find the maximum-likelihood estimate by comparing the residual sum of squares for various λ.

Problem caused by zeros

The above transformations meet serious difficulties when the data contain zeros because Eqs. 6 and 7 contain $\log_e(x)$ transformation. We cannot use the Box–Cox transformation in this case, either, because Eq. 9 containing the geometric mean of x in the denominator becomes infinity if data contain zeros. Hence, we should use $(x + c)$ instead of x to avoid these problems. Box and Cox (1964) suggested that the maximum-likelihood method is available to select an appropriate value of c in the modified Box–Cox transformation:

$$\left.\begin{aligned} f(x) &= \frac{(x + c)^\lambda - 1}{\lambda} && (\lambda \neq 0) \\ f(x) &= \log_e(x + c) && (\lambda = 0) \end{aligned}\right\} \tag{10}$$

However, Hill (1963) showed that the maximum-likelihood estimation of c is not acceptable even when $\log_e(x + c)$ exactly follows a normal distribution. Let us denote the smallest data by $x_{\min}$. Then, the likelihood becomes infinity when $c = -x_{\min}$ in this case. Thus, $-x_{\min}$ always becomes a global maximum-likelihood estimate of c, but such an estimate is not acceptable. Several alternative principles to determine the parameter c have been proposed (Hill 1963; Griffiths 1980; Berry 1987). Although these procedures are applicable under a certain range of assumptions, they require complicated calculations. Thus, we need another practical principle to determine the value of c.

A general approximation

The above difficulties seem to occur because zeros do not meet the assumption that is involved in the transformation. The slope of these transformation functions (Eqs. 6–8) becomes larger as x approaches zero. Let us consider a small difference in x, which is denoted by Δx. The difference in the transformed value, $f(x + \Delta x) - f(x)$, decreases with increasing x. In that sense, the quantity of Δx is more compressed if the position of Δx is far from zero, but it is more expanded if the position is near zero. The expansion effect of transformation becomes larger as the position of Δx approaches zero. Hence, these transformations are most suitable for a continuous distribution defined in $(0, \infty)$ such as shown by the curve in Fig. 1. However, the distribution of individuals is a discrete distribution defined for $\{0, 1, 2, \ldots\}$. As shown in Fig. 1a, such a discrete distribution cannot be well approximated by a continuous distribution. If we want to describe the discrete distribution by a continuous distribution, the discrete distribution should be shifted by 0.5, as shown by Fig. 1b. Such a shifted discrete distribution is approximately described by a continuous distribution with the same variance if the mean is large. Therefore, we can expect that $c = 0.5$ is a reasonable choice to enhance the effect of transformation.

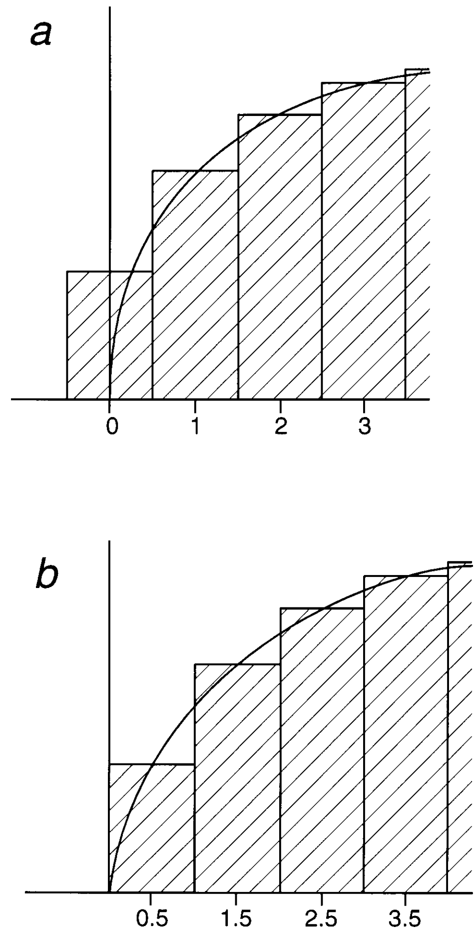

Fig. 1. Approximation of a discrete distribution defined in $\{0, 1, 2, \ldots\}$ by a continuous distribution defined in $(0, \infty)$. **a** Insufficient approximation without adding constant. **b** Improved approximation by adding 0.5

Numerical evaluation of approximation

To evaluate the effectiveness for using $c = 0.5$, I conducted numerical calculations for several combinations of parameters. It is known that the distribution of individuals can be approximately described by a negative binomial distribution in most cases. McArdle et al. (1990) used a negative binomial distribution, whose parameters are subjected to the constraint of Eq. 2, to evaluate the stabilization effect of the logarithmic transformation. In a similar way, I use a negative binomial distribution defined by

$$P(x) = \binom{k + x - 1}{x}\left(1 + \frac{m}{k}\right)^{-k}\left(\frac{m}{m + k}\right)^{x} = \frac{1}{\Gamma(k)} \cdot \frac{\Gamma(k + x)}{\Gamma(x + 1)}\left(1 + \frac{m}{k}\right)^{-k}\left(1 + \frac{k}{m}\right)^{-x} \qquad (11)$$

whose parameter k is subjected to the constraint of Eq. 1 or 2. We should also calculate the effect of transformation for a corresponding continuous distribution to evaluate how a discrete distribution approaches a continuous distribution by adding 0.5. Eq. 11 can be approximately described by a gamma distribution with a shape parameter k and a scale parameter k/m, if m is large:

$$P(x) = \frac{1}{\Gamma(k)} x^{k-1}\left(\frac{k}{m}\right)^{k} \exp\left(-\frac{k}{m}x\right) \qquad (12)$$

Hence, we use Eq. 12 whose parameters have the same constraint in its mean and variance. Bartlett (1936) used a similar approach to evaluate the effect of square root transformation. I calculated only the transformation for a realistic range, $s^2 \geq m$, for each combination of parameters.

When we have a relation $s^2 = m^2$, which corresponds to $g = 0$ and $h = 1$ in Eq. 1 or $a = 1$ and $b = 2$ in Eq. 2, Eqs. 6 and 7 recommend a logarithmic transformation, $\log_e(x)$. In this case, the transformed variable of a gamma distribution showed a conspicuous homoscedasticity as indicated by the horizontality of the dotted line in Fig. 2. The variance of the transformed variable of a negative binomial distribution converges to that of the gamma distribution as the mean increases. The convergence is much influenced by the value of c. Among calculated values of c, $c = 0.2$ seems to be most preferable in this situation, because it shows superior horizontality. $c = 0.5$ is not the best choice in this case, but it is preferable to $c = 1$.

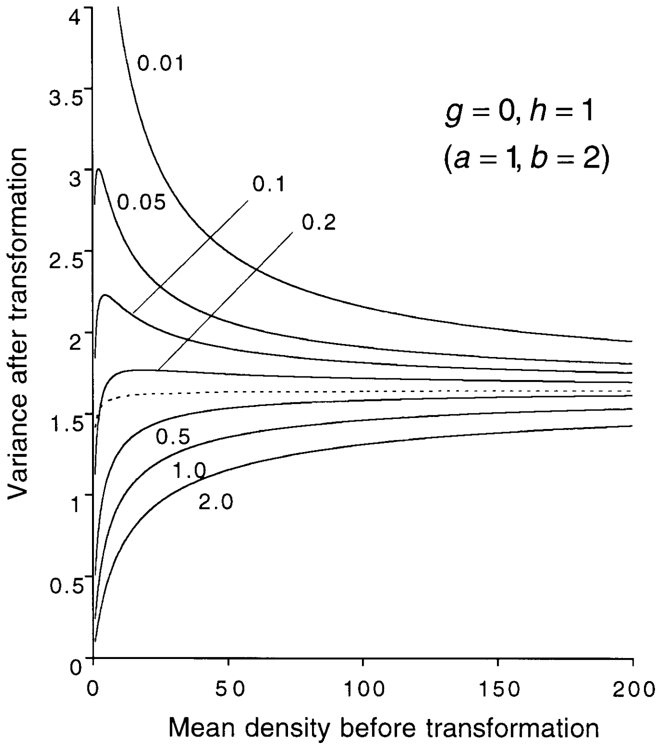

Fig. 2. Effect of adding constant (c) on the stabilization of variance of a negative binomial distribution with a constraint $s^2 = m^2$ that corresponds to $g = 0$ and $h = 1$ in Eq. 1 or $a = 1$ and $b = 2$ in Eq. 2. A logarithmic transformation, $\log_e(x + c)$, is used. Each number beside a *solid curve* indicates the c used in the calculation. The *dotted curve* is that of a gamma distribution with the same constraint for variance

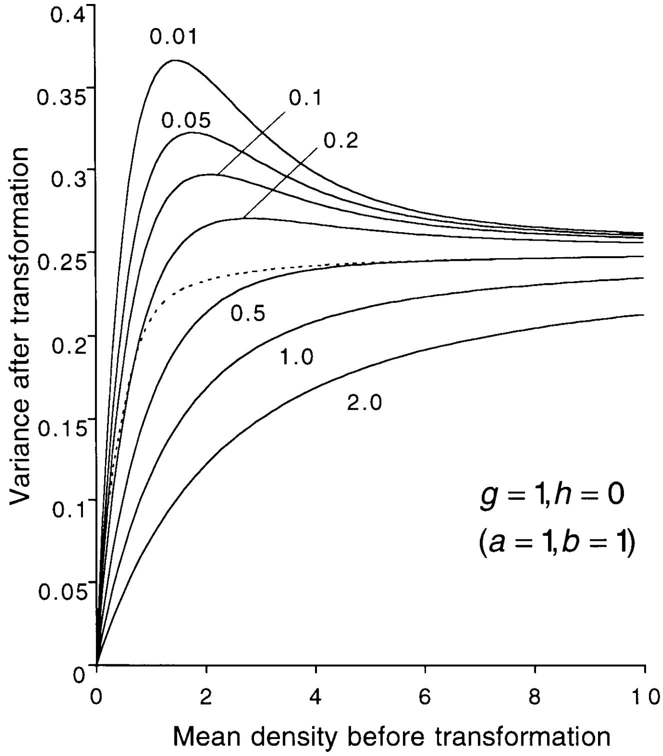

Fig. 3. Effect of adding constant (c) on the stabilization of variance of a negative binomial distribution with a constraint $s^2 = m$ that corresponds to $g = 1$ and $h = 0$ in Eq. 1 or $a = 1$ and $b = 1$ in Eq. 2. A square root transformation, $\sqrt{x+c}$, is used. Meaning of each curve is the same as in Fig. 2

The square root transformation is recommended by Eqs. 6 and 7 for the relation $s^2 = m$ that corresponds to $g = 1$ and $h = 0$ in Eq. 1 or $a = 1$ and $b = 1$ in Eq. 2. This case has been discussed by Bartlett (1936). The variance after transformation for $c = 0.5$ quickly converges to that of a gamma distribution with increasing mean (Fig. 3). The stabilization effect for $c = 1$ is considerably worse than that of $c = 0.5$. When the mean density is too small, the variance after the transformation is small irrespective of the choice of c, indicating that any transformation of this type is unsuccessful for such a case.

When we have a relation $s^2 = m^{1.5}$, which corresponds to intermediate values of parameters, $a = 1$ and $b = 1.5$, in Eq. 2, Eq. 7 recommends a power transformation $x^{0.25}$. In this case, the stabilization effect of transformation is excellent for a gamma distribution, as shown by the horizontality of the dotted curve in Fig. 4. The variance after transformation for $c = 0.5$ quickly converges to that of a gamma distribution with increasing mean. The convergence is very slow for $c = 1$.

When we use intermediate values of parameters, $g = 0.5$ and $h = 0.5$, in Eq. 1, Eq. 6 recommends a transformation $\log_e(\sqrt{x} + \sqrt{x+1})$. The variance after transformation for $c = 0.5$ is similar to that of a gamma distribution (Fig. 5). The superiority of $c = 0.5$ over $c = 1$ is also clear in this case.

Discussion

I recommended the transformation using $(x + 0.5)$, such as $\sqrt{x + 0.5}$ and $\log_e(x + 0.5)$, to stabilize the variance of populations for ANOVA. Figures 2–5 indicate that $c = 0.5$ is preferable to $c = 1$, although it is not always the best choice. The square root transformation with $c = 0.5$ was first recommended by Bartlett (1936) as the analogy with Yates' (1934) continuity correction that is used to approximate a tail probability of a discrete distribution by a tail probability of the corresponding continuous distribution. I recommended the use of $c = 0.5$ by a different reason – a discrete distribution defined in $\{0, 1, 2, \ldots\}$ is approximately described by a continuous distribution defined in $(0, \infty)$ if we use $c = 0.5$. Anscombe (1948) studied optimal values of c for several specified distributions. If the distribution is a Poisson distribution, for example, $c = 3/8$ is optimal in a sense that the variance converges most quickly as the mean increases. If the distribution follows a negative binomial distribution with constant k, another transformation, $\log_e(x + k/2)$, may be recommended for a large m. If the form of distribution is not known, however, $c = 0.5$ seems to be a reasonable choice.

The $\log_e(x + 1)$ transformation will be thus less preferable for ANOVA, because the inconstancy of variance does

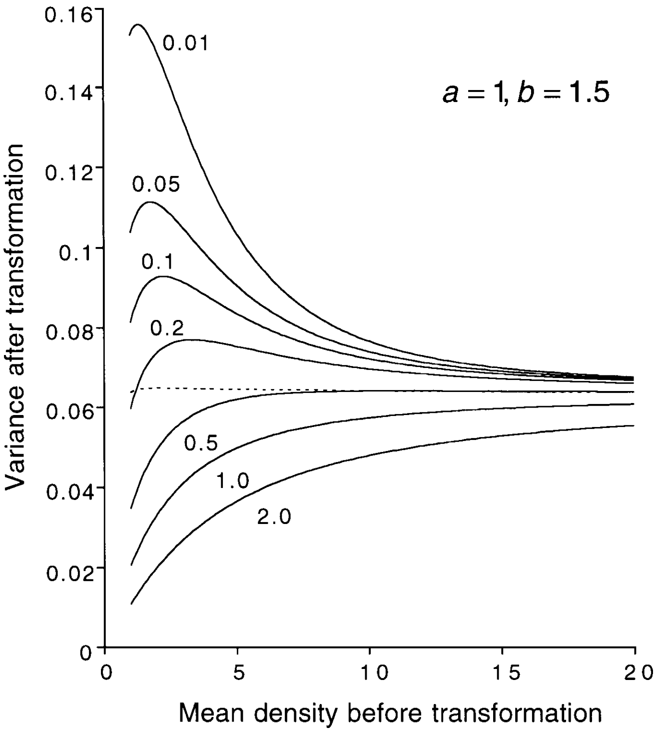

Fig. 4. Effect of adding constant (c) on the stabilization of variance of a negative binomial distribution with a constraint $s^2 = m^{1.5}$ that corresponds to $a = 1$ and $b = 1.5$ in Eq. 2. A power transformation, $(x + c)^{0.25}$, is used. Meaning of each curve is the same as in Fig. 2

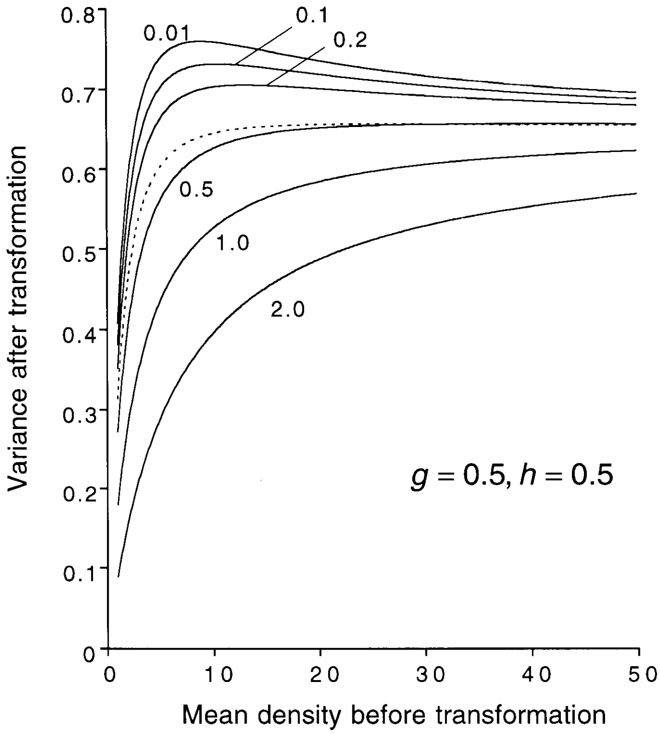

Fig. 5. Effect of adding constant (c) on the stabilization of variance of a negative binomial distribution with a constraint $s^2 = 0.5(m + m^2)$ that corresponds to $g = 0.5$ and $h = 0.5$ in Eq. 1. An arc-hyperbolic transformation, $\log_e(\sqrt{x+c} + \sqrt{x+c+1})$, is used. Meaning of each curve is the same as in Fig. 2

not guarantee the assumption of the F-test. This transformation may not also be suitable for describing the population dynamics. Notice that the population dynamics should be defined for the total population in an area but not defined for the number of observed individuals, because the latter is influenced by the sampling variability that is not of interest to us. In most cases, the zero observation indicates that the population density is very low but does not indicate that the population is truly zero. If the sample size is extremely large, zero will not occur in many cases. In that sense, zero data are artifacts that derive from the deficiency in sampling effort. Then, it is preferable to use an approximation for the dynamics of the logarithm of true total population. Let us imagine that the sample size becomes r times larger to obtain the total population N. Then, the frequency distribution of N/r is a continuous alternative to the distribution of x, if r is extremely large. By the same argument shown in Fig. 1, therefore, $\log_e(x + 0.5)$ transformation yields an approximation for the distribution of $\log_e(N/r)$, i.e., the distribution of $\log_e(N) - \log_e(r)$. Hence, $\log_e(x + 0.5)$ transformation seems to be preferable to $\log_e(x + 1)$ for evaluating the population dynamics as well as for performing ANOVA.

One of the possible misuses of the logarithmic transformation is to add a constant to the mean population instead of the total population before transformation. As an illustration, let us consider that 100 plants are examined and x individuals are observed on them. In this case, if we use logarithmic transformation after adding 0.5 to the mean population (x/100), the transformation corresponds to the logarithmic transformation using $c = 50$, because we have $\log_e(x/100 + 0.5) = \log_e(x + 50) - \log_e(100)$. Taylor series expansions about x around 0 yield:

$$\log_e(x + c) \approx \log_e(c) + \frac{x}{c} - \frac{x^2}{2c^2} + \ldots \qquad (13)$$

If c/x is large, therefore, the transformation formula approaches $f(x) = \log_e(c) + x/c$, i.e., no transformation. Thus, the logarithmic transformation may become meaningless if 0.5 is added to the mean population in this case. I recommended $c = 0.5$ because it is half of the discrete unit (see Fig. 1). When we analyze mean population, x/100, however, the discrete unit is 1/100, and hence we should add 0.5/100 before the logarithmic transformation in such a case. Thus, we should more carefully select an appropriate constant when we analyze the mean density instead of total population.

Acknowledgments I thank Dr. K. Kiritani for his comments on the manuscript.

References

Anscombe FJ (1948) The transformation of Poisson, binomial and negative-binomial data. Biometrika 35:246–254

Bartlett MS (1936) The square root transformation in analysis of variance. J R Stat Soc Suppl 3:68–78
Bartlett MS (1947) The use of transformations. Biometrics 3:39–52
Beall G (1942) The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. Biometrika 32:243–262
Berry DA (1987) Logarithmic transformations in ANOVA. Biometrics 43:439–456
Bliss CI (1941) Statistical problems in estimating populations of Japanese beetle larvae. J Econ Entomol 34:221–232
Box GEP, Cox DR (1964) An analysis of transformations. J R Stat Soc B Met 26:211–252
Gaston KJ, McArdle BH (1993) Measurement of variation in the size of populations in space and time: some points of clarification. Oikos 68:357–360
Griffiths DA (1980) Interval estimation for the three-parameter lognormal distribution via the likelihood function. Appl Stat 29:58–68
Hill BM (1963) The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic. J Am Stat Assoc 58:72–84
Itô Y, Kitching RL (1986) The importance of non-linearity: a comment on the view of Taylor. Res Popul Ecol 28:39–42
Iwao S (1968) A new regression method for analyzing the aggregation pattern of animal populations. Res Popul Ecol 10:1–20
Iwao S, Kuno E (1968) Use of the regression of mean crowding on mean density for estimating sample size and the transformation of data for the analysis of variance. Res Popul Ecol 10:210–214
Iwao S, Kuno E (1971) An approach to the analysis of aggregation pattern in biological populations. In: Patil GP, Pielou EC, Waters WE (eds) Statistical ecology, vol 1. Spatial patterns and statistical distributions. Pennsylvania State University Press, London, pp 461–513
Kuno E (1991) Sampling and analysis of insect populations. Annu Rev Entomol 36:285–304
Leps J (1993) Taylor's power law and the measurement of variation in the size of population in space and time. Oikos 68:349–356
McArdle BH, Gaston KJ (1992) Comparing population variabilities. Oikos 64:610–612
McArdle BH, Gaston KJ (1993) The temporal variability of populations. Oikos 67:187–191
McArdle BH, Gaston KJ (1995) The temporal variability of densities: back to basics. Oikos 74:165–171
McArdle BH, Gaston KJ, Lawton JH (1990) Variation in the size of animal populations: patterns, problems and artefacts. J Anim Ecol 59:439–454
Perry JN, Woiwod IP (1992) Fitting Taylor's power law. Oikos 65:538–542
Routledge RD, Swartz TB (1991) Taylor's power law re-examined. Oikos 60:107–112
Stuart A, Ord JK (1994) Kendall's advanced theory of statistics, vol 1. Distribution theory, 6th edn. Arnold, London
Taylor LR (1961) Aggregation, variance and the mean. Nature (Lond) 189:732–735
Taylor LR (1984) Assessing and interpreting the spatial distribution of insect populations. Annu Rev Entomol 29:321–357
Taylor LR, Woiwod IP, Perry JN (1978) The density-dependence of spatial behaviour and the rarity of randomness. J Anim Ecol 47:383–406
Williamson M (1984) The measurement of population variability. Ecol Entomol 9:239–241
Yamamura K (1999) Key-factor/key-stage analysis for life table data. Ecology 80:533–537
Yates F (1934) Contingency tables involving small numbers and the χ^2 test. J Roy Stat Soc Suppl 1:217–235