Res. Popul. Ecol. (1968) X, 210–214

USE OF THE REGRESSION OF MEAN CROWDING ON MEAN DENSITY FOR ESTIMATING SAMPLE SIZE AND THE TRANSFORMATION OF DATA FOR THE ANALYSIS OF VARIANCE[1,2]

Syun'iti IWAO and Eizi KUNO

Entomological Laboratory, College of Agriculture,
Kyoto University, Kyoto, Japan

In the planning of a sample survey of an animal population, it is important to decide how many sample units should be taken to estimate the population mean with a desired precision. For this decision, we need some information concerning the mean and variance of the population to be sampled, which is usually obtained by preliminary sampling or by guess work. Also, when applying the analysis of variance to the data obtained from a sample survey, the individual counts should be transformed in a proper way to stabilize the variance unless the variance is independent of the mean. Since the variance in most biological populations changes in some relation with the mean density, it would be useful if we can find out a definite relation between mean and variance for the spatial distribution pattern of the species concerned.

As shown in a previous paper (IWAO, 1968), the relation of mean crowding to mean density can be fitted to a linear regression in a variety of both theoretical and biological populations. From this relation, we can derive an approximate method for estimating sample size in simple random sampling and a systematic way of transformation of raw data, both of which can be applied widely without restrictions as to the distribution pattern.

ESTIMATION OF SAMPLE SIZE IN SIMPLE RANDOM SAMPLING

The mean crowding, $\overset{*}{m}$, is related with the variance, V, in the following way (LLOYD, 1967):

$$\overset{*}{m}=m+(V/m-1), \tag{1}$$

where m is the mean density per habitat unit (quadrat).
From this, we get

$$V=m(\overset{*}{m}-m+1). \tag{2}$$

If $\overset{*}{m}$ is linearly related with m over the range of m's, $\overset{*}{m}$ can be expressed by

[1] Contribution from the Entomological Laboratory, Kyoto University No. 418.

[2] Contributions from JIBP-PT No. 52. Aided in part by a grant from the Ministry of Education for the special project research, 'Studies on the dynamic status of biosphere'.

$$\overset{*}{m}=\alpha+\beta m, \tag{3}$$

as shown in the previous paper (Iwao, *l. c.*).
Substituting (3) into (2),

$$V=(\alpha+1)m+(\beta-1)m^2. \tag{4}$$

When m, V, α and β are replaced by their sample estimates $\bar{x}$, s^2, $\hat{\alpha}$ and $\hat{\beta}$ respectively, the standard deviation of the mean density, $s_{\bar{x}}$, is given by

$$s_{\bar{x}}=\sqrt{\frac{(\hat{\alpha}+1)\bar{x}+(\hat{\beta}-1)\bar{x}^2}{q}}, \tag{5}$$

where q is the number of sample units taken.

Since the half-width of confidence interval, d, is given by $d=ts_{\bar{x}}$, where t is the value of the normal deviate corresponding to the desired confidence probability (Student's t),[1] we get

$$d=t\sqrt{\frac{(\hat{\alpha}+1)\bar{x}+(\hat{\beta}-1)\bar{x}^2}{q}}, \tag{6}$$

and hence the number of quadrats to be sampled is

$$q=\frac{t^2}{d^2}\left\{(\hat{\alpha}+1)\bar{x}+(\hat{\beta}-1)\bar{x}^2\right\}. \tag{7}$$

If we are interested in the magnitude of error relative to mean density, $D=d/\bar{x}$, which is more widely used in population studies, we have

$$D=t\sqrt{\frac{1}{q}\left(\frac{\hat{\alpha}+1}{\bar{x}}+\hat{\beta}-1\right)}, \tag{8}$$

and

$$q=\frac{t^2}{D^2}\left(\frac{\hat{\alpha}+1}{\bar{x}}+\hat{\beta}-1\right). \tag{9}$$

Equation (7) or (9) can be used for the determination of sample size by knowing the approximate level of mean density of the population that is being sampled.

If the proportion of q to the total number of units in the population Q is appreciable, the corrected number of sample units is given by

$$q'=\frac{q}{1+\frac{q}{Q}}. \tag{10}$$

Strictly speaking, the equations (6) through (9) are valid only when the sample mean $\bar{x}$ is normally distributed about the true mean of the population. They are, however, applicable more generally for our practical purpose, because the normal approximation can be allowed even for highly skewed distributions as far as the number of sample units is sufficiently large and because a considerable amount of skewness in the distribution of $\bar{x}$ can be tolerated if we are more interested in the total amount of the error of estimate rather than the upper and lower limits of it

[1] Put $t \fallingdotseq 2$ for 95% confidence probability, if $q>30$.

(COCHRAN, 1953).[1]

It is apparent from the equation (9) that the sample size needed for a specified degree of precision D depends on the values of α and β as well as on the level of mean density; the influence of α on sample size gradually decreases as the mean density increases, while the effect of β is independent of the density. Since $\alpha = 0$ and $\beta = 1$ in the POISSON distributions, the equation (9) is reduced to the form $q = \frac{t^2}{D^2} \frac{1}{\bar{x}}$. If the distributions follow the negative binomial series with a common k (i.e. $\alpha = 0$ and $\beta = 1 + \frac{1}{k}$), the equation becomes $q = \frac{t^2}{D^2}\left(\frac{1}{\bar{x}} + \frac{1}{k}\right)$. When putting $t = 1$,[2] this is identical with the equation given by KUNO *et al.* (1963) and ROJAS (1964) for this particular distribution model. For the positive binomial distributions with a common k', we get $q = \frac{t^2}{D^2}\left(\frac{1}{\bar{x}} - \frac{1}{k'}\right)$ because $\alpha = 0$ and $\beta = 1 - \frac{1}{k'}$.

MORISITA (1964) derived the basic relation of I_δ to sample size, and ONO (1967) described the equation for simple random sampling as $q = \frac{t^2}{D^2}\left(I_\delta - 1 + \frac{1}{\bar{x}}\right)$. When I_δ in this equation is replaced by $\left(\frac{\alpha}{\bar{x}} + \beta\right)$, which is valid as far as q is sufficiently large (see IWAO, *l. c.*), we have the same equation as (9). Since I_δ shows a constant value over a range of different mean densities only in the limited cases (IWAO, *l. c.*), the equation (9) would be applicable more generally if a linear regression has been proved to be fitted to the relation of mean crowding to mean density for the species concerned.

METHOD OF TRANSFORMATION FOR STABILIZING VARIANCE

When the variance is closely related to the mean intrinsically as in the cases discussed here, the analysis of variance cannot, strictly, be applied to the raw data unless individual counts would be transformed in such a way that the variance can be stabilized, because the analysis is based on the assumption that the variance is independent of the mean. It is known that when the variance is expressed by a function $g(m)$ of the mean m theoretical transformation $f(x)$ satisfying this purpose can be derived from

$$f(x) = c \int \frac{1}{\sqrt{g(x)}}\, dx,$$

where c is an arbitrary constant. Using this $f(x)$ instead of x we can expect the new variance, $Var\,[f(x)]$, having a constant value c^2, irrespective of the mean (see, for example, BEALL, 1942). Fortunately, this type of transformation also improves considerably the normality of distribution and the additivity of effect as well, which

[1] If the more accurate estimation is desired, we can find out an appropriate way of transformation of original data, as will be shown in the next chapter.

[2] This means that the ratio of standard error to mean is set at the desired level.

are the other two assumptions underlying the analysis of variance.

Now since $g(m)$ in the present case is $(\alpha+1)m+(\beta-1)m^2$, the transformation required is given by

$$f(x)=c\int\frac{1}{\sqrt{(\alpha+1)x+(\beta-1)x^2}}\,dx. \qquad (11)$$

This generates several different forms of transformation listed below in accordance with the values of α and β.

i) $\alpha>-1,\ \beta>1$:

$$f(x)=\sinh^{-1}\sqrt{\frac{\beta-1}{\alpha+1}x}=\log_e\left(\sqrt{\frac{\beta-1}{\alpha+1}x}+\sqrt{\frac{\beta-1}{\alpha+1}x+1}\right), \qquad (12)$$

$$Var[f(x)]=\frac{\beta-1}{4};$$

ii) $\alpha=-1,\ \beta>1$:

$$f(x)=\log_e x, \qquad (13)$$

$$Var[f(x)]=\beta-1;$$

iii) $\alpha>-1,\ \beta=1$:

$$f(x)=\sqrt{x}, \qquad (14)$$

$$Var[f(x)]=\frac{\alpha+1}{4};$$

iv) $\alpha>-1,\ \beta<1$:

$$f(x)=\sin^{-1}\sqrt{\frac{1-\beta}{\alpha+1}x}, \qquad (15)$$

$$Var[f(x)]=\frac{1-\beta}{4}.$$

It is obvious that a set of these transformations covers a wide variety of biological data, as expected from the wide applicability of linear regression between the mean crowding and the mean to both theoretical and natural populations (Iwao, *l. c.*). For example, the negative binomial series with a common k offers a special case of (i) ($\alpha=0,\ \beta=1+1/k$) resulting the Beall's (1942) transformation $f(x)=\sinh^{-1}\sqrt{x/k}$ where $Var[f(x)]=1/4k$. Similarly, the familiar square-root and arcsine transformation for the Poisson and the positive binomial distribution——$f(x)=\sqrt{x}$ with $Var[f(x)]$ equal to 1/4, and $f(x)=\sin^{-1}\sqrt{x/k'}$ with $Var[f(x)]$ equal to $1/4k'$, respectively[1]——can be derived as special cases of (iii) and (iv) by putting $\alpha=0$, and $\alpha=0,\ \beta=1-1/k'$, respectively.

When we are concerned with the analysis of field populations whose distributions are almost invariably contagious, the equation (12) above will provide a standard method of transformation for stabilizing variance. Apparently it has by far wider applicability than the Beall's transformation, and also the theoretical

[1] The parameter k' indicates the maximum number of individuals a sample unit could contain, but in the sampling for proportions it means the sample size or the number of trials.

basis for the form of $g(m)$ is rather valid as compared with the TAYLOR's (1961) transformation, $f(x)=x^{1-b/2}$, based on the purely empirical relationship $g(m)=am^b$ where a and b are constants. The calculation of $f(x)$ by the equation (12) may be somewhat laborious, but this does not become a serious difficulty since at present we can leave the entire process of calculation to the computer.

SUMMARY

An approximate method for estimating the sample size in simple random sampling and a systematic way of transformation of sample data are derived by using the parameters α and β of the regression of mean crowding on mean density in the spatial distribution per quadrat of animal populations (IWAO, 1968). If the values of α and β have been known for the species concerned, the sample size needed to attain a desired precision can be estimated by simply knowing the approximate level of mean density of the population to be sampled. Also, an appropriate variance stabilizing transformation of sample data can be obtained by the method given here without restrictions on the distribution pattern of the frequency counts.

REFERENCES

BEALL, G. (1942) The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. *Biometrika*, **32**: 243-262.

COCHRAN, W. G. (1953) *Sampling techniques*. New York.

IWAO, S. (1968) A new regression method for analyzing the aggregation pattern of animal populations. *Res. Popul. Ecol.*, **10**: 1-20.

KUNO, E., S. YAMAMOTO, H. SATOMI, Y. ŌUTI and T. OKADA (1963) On the assessment of insect populations in a large area of paddy field based on the negative binomial distribution. *Proc. Assoc. Pl. Prot. Kyushu*, **9**: 33-36. (In Japanese with English summary).

LLOYD, Monte (1967) 'Mean crowding'. *J. Anim. Ecol.*, **36**: 1-30.

MORISITA, M. (1964) Application of I_δ-index to sampling techniques. *Res. Popul. Ecol.*, **6**: 43-53.

ONO, Y. (1967) Estimation of animal numbers. In '*Seitaigaku Jisshu-sho*' (*Guide for experiments in ecology*), pp. 87-107, Tokyo (In Japanese).

ROJAS, B. A. (1964) La binomial negativa y la estimación de intensidad de plagas en el suelo. *Fitotecnia Latinamer.*, **1**(1) : 27-36. (Cited from SOUTHWOOD, T. R. E., 1966, *Ecological methods*. London).

TAYLOR, L. R. (1961) Aggregation, variance and the mean. *Nature*, **189**: 732-735.

平均密度に対する平均こみあい度の回帰に基づく抽出標本数の決定とデータの変換法

巌　俊一・久野英二

平均密度に対する平均こみあい度の回帰は，種々の確率分布モデルや動物の枠当り分布において直線となる（巌，1968)。この関係からえられる基本集合度示数 α と密度一集合度係数 β を用いて，単純任意抽出において一定精度をうるに必要な標本数の決定と，分散分析などを行なうに必要な分散安定化のためのデータ変換の一般式を導いた。これらの式は枠当り個体数の分布様式の如何にかかわらず使用できる。

Erratum: IWAO (Vol. X, No. 1, 1968), page 17, line 7 from bottom, *for* $\alpha=0$, $\beta=0$ *read* $\alpha=0$, $\beta=1$.