

Random distribution in birds in flocks: significance testing

J.J.D. GREENWOOD

Department of Biological Sciences, University of Dundee, Dundee DD1 4HN, UK

Citation: Greenwood, J.J.D. 1981. Random distribution of birds in flocks: significance testing. *Wader Study Group Bull.* 31: 35.

Furness & Galbraith (1980) dyed part of a catch of Redshanks *Tringa totanus* and subsequently counted the numbers of dyed individuals in groups of 10 birds in the roosting flock. The distribution of dyed birds appeared to be clumped; more groups than chance would dictate contained no or several dyed birds and fewer contained average numbers. This is an interesting finding and the technique should certainly be used more widely to see if the phenomenon is general. The purpose of this note is point out that the statistical test used by Furness and Galbraith is unduly conservative when the birds are clumped – i.e. it gives rise to a significant result less often than it should. I shall indicate a less conservative test and point out that even this is somewhat conservative, so that other, less straightforward, methods may be needed in some circumstances.

The data obtained by Furness & Galbraith are shown in Table 1 with their “expected numbers”, based on the Poisson distribution. This distribution is based on the assumption that it is possible for an infinite number of dyed birds to occur in any one group. Since the sampling technique sets an upper limit of 10, this assumption is violated. The appropriate distribution to use is the binomial distribution, since this is concerned with the number of dyed birds in a sample of finite size. Binomial expectations are also shown in the table. It is clear that, as theory predicts, data which show clumping are fitted even worse by the binomial distribution than by the Poisson: the result is thus actually more significant than Furness and Galbraith supposed.

The calculation of the binomial expectation is easy. Suppose that the frequency of dyed birds, estimated by dividing the total number of dyed birds in all the groups by the total number of birds in the groups, is p (0.0772 in the present case). Then, if n is the group size (10 in the present case) and N is the number of groups counted (177), the expected number of groups with no dyed birds is:

$$E(O) = N(1-p)^n$$

The expected numbers with 1, 2, 3, ... n dyed birds are:

$$E(1) = [n/1] [p/(1-p)] [E(O)]$$

$$E(2) = [(n-1)/2] [p/(1-p)] [E(1)]$$

$$E(3) = [(n-2)/3] [p/(1-p)] [E(2)]$$

...

$$E(n) = [1/n] [p/(1-p)] [E(n-1)]$$

In practice, if p is small the expected numbers fall off rapidly as one goes through this series. Given the restrictions of the χ^2 test (see below), such small expectations are individually of no interest. In general, once one has reached a value of x such that $E(x+1)$ is likely to be much less than 5, one calculates the expected number of groups containing more than x dyed birds by adding up the expectations $E(O)$ to $E(x)$ inclusive and subtracting from the total number of groups (N). In the present case:

$$E(\text{more than } 3) = 177 - (79.3 + 66.3 + 25.0 + 5.6)$$

To test the significance of the departure of the expected numbers from the observed data, one uses the chi-squared test. As always, this should not be used when more than 1 in 5 of the expected values is less than 5 or when any expected is less than 1. To overcome this problem, one may combine adjacent rows: in Table 1, I have combined the rows for “3 dyed birds” and “more than 3 dyed birds”, since the latter has an expectation of only 0.8. For each pair of observed (O) and expected (E) values one calculates $(O-E)^2/E$ and sums them to obtain χ^2 in the usual way. This is compared with the tabulated chi-squared with χ^2 degrees of freedom, χ being the number of values of $(O-E)^2/E$ used. For the present data, $\chi^2 = 19.6$ with 2 d.f., which is less than the tabulated value at $P = 0.001$: the result is highly significant.

Strictly speaking, even the binomial distribution is not appropriate for these data, since it assumes that the total number of birds in the population from which the N groups are drawn is infinite. The correct distribution to use, though it is not at all easy, is the hypergeometric. Fortunately, if the number of birds remaining unsampled is several times greater than the size of individual samples, then the binomial is a good approximation. Furthermore, application of the binomial will result in a conservative test: if clumping is demonstrably significant using the binomial approximation,

Table 1. Observed and expected numbers of dyed birds per group of ten.

Number dyed in group of ten	Observed number of group	Poisson expectation	Binomial expectation	$(O-E)^2/E$
0	97	81.8	79.3	$(97-79.3)^2/79.3 = 3.95$
1	40	63.1	66.3	$(40-66.3)^2/66.3 = 10.43$
2	28	24.4	25.0	$(28-25.0)^2/25.0 = 0.36$
3	9	6.3	5.6	$(12-6.4)^2/6.4 = 4.90$
3	3 } 12	1.2	0.8 } 6.4	
				$\chi^2 = 19.64$



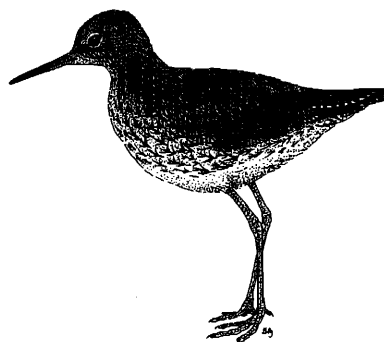
it would certainly be significant if one applied the hypergeometric distribution. However, if p is large or the number of birds remaining unsampled is small, one may miss what is actually significant clumping by using the binomial.

Acknowledgements

I am grateful to Cynthia Greenwood for typing this note and to G.E. Thomas for checking its content.

References

Furness, R.W. & Galbraith, H. 1980. Non-random distribution in roosting flocks of waders marked in a cannon net catch. *Wader Study Group Bull.* 29: 22–23.



A potential bias in log-transformed allometric equations

A.G. WOOD

British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK

Citation: Wood, A.G. 1986. A potential bias in log-transformed allometric equations. *Wader Study Group Bull.* 47: 17–19.

Introduction

Many recent studies in avian ecology have used regressions with logarithmic transformations to estimate various biological parameters, particularly metabolic rate (Lasiewski & Dawson 1967, 1969, Zar 1968, Aschoff & Pohl 1970, Kendeigh *et al.* 1977). This paper examines a potential bias resulting from the use of such equations, and presents the appropriate methods for converting estimates from logarithmic equations back to untransformed units.

The problem

In the general case, we have two variables X and Y which are related by the allometric equations:

$$Y = kX^b \quad (1)$$

where k and b are constants. Although the relationship between X and Y is non-linear, the transformed variates $\log X$ and $\log Y$ are connected by the straight line relationship:

$$\log Y - \log k = b \cdot \log X \quad (2)$$

This equation implies a linear relation between the logarithms of X and Y based on three assumptions:

1. The expected value of Y , for a given X is $(E(\log Y) = \log k + b \cdot \log X)$.
2. The variance V of $\log Y$, given $\log X$, is constant.
3. For each value of $\log X$, $\log Y$ is normally distributed.

The parameters of transformed equation (2) can now be estimated using the biological data and standard least-squares regression techniques.

When a logarithmic transformation is used it is usually necessary to be able to express estimated values of Y in

untransformed units. Such a back transformation is not direct, because if the distribution of $\log Y$ at a given $\log X$ is normal, the distribution of Y cannot be normal, but will be skewed. In fact the solution of equation (2) for a given X , and determining the antilogarithm of $\log Y$, yields the median of the skewed distribution of Y rather than the mean (Baskerville 1971). The correction factor (CF) by which this median must be multiplied to obtain the mean of Y , has been derived by a number of authors (Baskerville 1971; Mountford & Bunce 1973, Sprugel 1983), and is calculated from:

$$CF = e^{(V/2)}$$

where V is the variance of $\log Y$, e is the base of natural logarithms 2.718.

In practice V is not known, but can be estimated from the square of the standard error of the estimate of the regression, giving equation (3).

$$CF = e^{(SEE^2/2)} \quad (3)$$

where SEE is the standard error of estimate of the regression. The values for $\log k$ and b in equation (2) also have errors associated with them. However, they can be considered insignificant if a large enough sample size for the regression is obtained.

The value of SEE depends on the base to which logarithms are taken when the values of Y are transformed (Sprugel 1983). To obtain the correct value for the correction factor (CF), SEE must be based on natural logarithms. Therefore, using a base 10 standard error does not give the correct value; this base 10 SEE should be converted to base e (multiply by $\log_{.10} e = 2.303$) and this value used in equation (3).

In energetic studies, estimates are made of a species' metabolic rate from its weight using the following equation,

