

EXPERIMENTAL DESIGN WHEN COUNTING BIRDS

DAVID G. DAWSON¹

ABSTRACT.—The number of species discovered increases by a constant increment with each doubling of observer effort; this relationship can be used to standardize incomplete lists for studies of biogeography and species diversity. The use of frequency of occurrence, rather than the total number counted, prevents discrimination of differences in density at high population levels. The counts of many species approximate a Poisson distribution, for which chi-squared tests on the totals counted may be an adequate approximation; the difference between two samples that may be detected is inversely proportional to the square root of the total number of a species counted. Monitoring long-term changes in bird numbers is better done by point or transect counts than by territory mapping; a large number of sites should be visited under standardized conditions each year. Rare species or habitats are best sampled after a reconnaissance has established their distribution, not by random or systematic sampling of all habitats. Differences in bird populations between habitats can be studied with index techniques, provided that the habitats are not too diverse and that observers, times of day, season, and other sources of bias are taken into account. Territory mapping permits fine-scale definition of habitats. Estimates of absolute density, accurate enough for the study of energetic or trophic relationships, may be obtained from point, transect or mapping methods.

No single answer can be given to the question of how to count birds. Principles common to the design of all sampling schemes, such as the choice between random and systematic sampling, or the sampling units used, are adequately covered in standard texts (e.g., Cochran 1963), so this paper concentrates on the limitations of different counting methods in answering questions about bird numbers and diversity. My other two papers in this symposium give the background to the more commonly used point, transect and territory mapping techniques.

SPECIES RICHNESS AND SPECIES DIVERSITY

The number of species in an area is of interest in quantitative biogeography and conservation (MacArthur and Wilson 1967, Diamond 1975b). The same measure is also used in studies of species "diversity," where it is termed species "richness," one of the two components of diversity, the other being "equitability" (Tramer 1969). A problem for all uses is that the list of species increases with the time spent looking.

Several researchers have fitted empirical curves to this increase (Preston 1960, Caughley 1965, Järvinen and Väisänen 1977a, Ratowsky and Ratowsky 1979) and found a linear increase in the number of species with the logarithm of effort. This semilogarithmic relationship is supported by the results of my studies (Fig. 1). Robbins (1972) suggested that the relationship is closer for the square root or even the fourth root of the effort, but presented no data to support this.

The semilogarithmic equation may be written $S_n = S_1 + a \log n$, where S_n is the number of species discovered in n units of effort (e.g.,

points, km of transect, ha of mapping plot, or time), S_1 is the average number discovered in one unit, and a is a coefficient describing the rate of increase of S_n with increasing n . Coefficient a is estimated from $a = (S_t - S_1)/\log t$, where t is the number of units of effort put into a study area.

The average number of species discovered in a unit of effort, S_1 , is also an estimate of the sum of the individual species' probabilities of discovery in a unit of effort. This will depend on the densities, d , of the individual species and on their susceptibility to discovery (conspicuousness or detectability), b . S_1 has been termed "mean richness" by Blondel (1977), but "richness" is conventionally used to denote the total number of species in an area by Tramer (1969), whom I follow. The semilogarithmic relationship between S_n and effort tells us that S_n increases quickly with increasing effort at first and then levels off to increase much more slowly at high n . Thus at large n the species list is nearly complete, much less dependent on variation in d and b , and a much better estimate of species "richness." The choice of a particular large value of n is arbitrary, but it is probably best to choose one towards the upper end of the values of n from the areas to be compared, to avoid excessive extrapolation. Frochot (1976) suggested a similar approach, but did not fit a curve to his results.

The application of this estimate of species richness can be illustrated with data from Blondel (1977). In *Quercus ilex* stands in Provence, $t = 16$, $S_{16} = 23$, $S_1 = 10.7$, so a is 10.2 and S_{30} is estimated to be 25.8. Figure 2 compares Blondel's values of species diversity (which were derived from density estimates) with values of S_{30} calculated as above. Clearly S_{30} accounts for most of the variation in species diversity, a point made also by Tramer (1969),

¹ Ecology Division, Department of Scientific and Industrial Research, Private Bag, Lower Hutt, New Zealand.

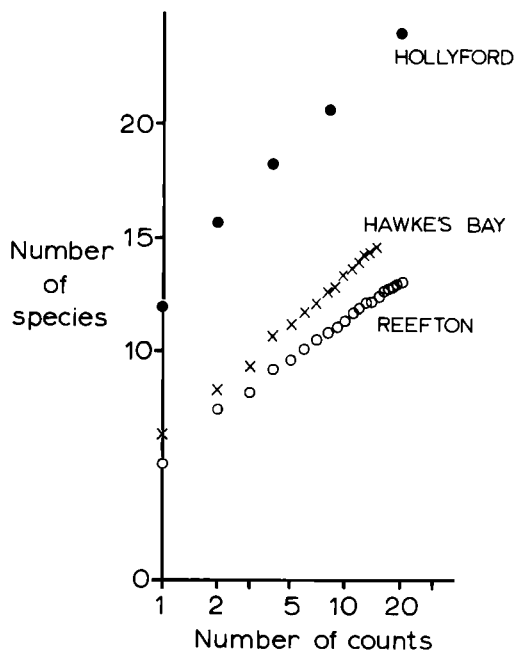


FIGURE 1. The increasing number of species detected with an increasing number of counts in three New Zealand studies using the five-minute point count (Dawson and Bull 1975): the total from 20 points in the Hollyford Valley and the average from 24 points in Hawke's Bay and 10 near Reefton (Dawson et al. 1978). All three relationships seem linear.

Blondel (1975) and Cousins (1977). Thus point, transect or mapping samples readily yield an estimate of species richness, which is also well correlated with species diversity. This approach may be preferable to estimating b for each species and calculating diversity (as was done by Blondel 1975, Järvinen and Väisänen 1976a) and is certainly preferable to the assumption that b is constant (Lancaster and Rees 1979, Wilkinson and Guest 1977, and many mapping studies).

The coefficient of the semilogarithmic relationship, a , describes the rate at which species are added to the list as effort increases. This will, like S_1 , depend on individual species' values of density and detectability, and each of these may vary with place or time. Thus it is difficult to read any meaning into the value of a , but like S_1 it tends to be large when the total number of species is large (Fig. 2).

If the equitability component of diversity is to be studied, estimates of density are needed (Tramer 1969, Taylor 1978), but the accuracy of density estimates from mapping, point or transect methods may be inadequate for this (Dawson 1981c).

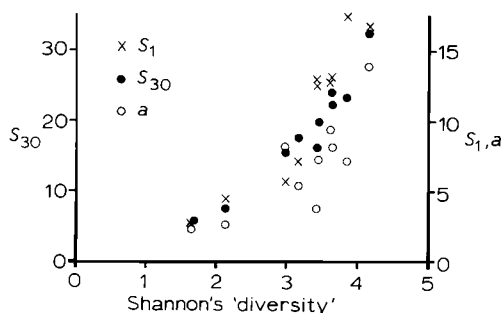


FIGURE 2. Relationship between Shannon's "diversity" (as calculated by Blondel [1977] from point counts and a conversion factor to density for each species) and the average number of species per point count (S_1), the slope of the semilogarithmic relationship (a), and the estimated species list for 30 E.F.P. counts (S_{30}). (The value of S_{30} for a given diversity varies little, so S_{30} is a good measure of diversity; a and S_1 are less good.)

Island biogeographic studies may compare the number of species in habitat patches of different size (MacArthur and Wilson 1967). Here two additional problems occur; as no area is sufficiently homogeneous for the species list at one point to be representative of the whole area, the list increases with effort in both space (Cousins 1977) and time. I believe that patches should be sampled in proportion to their area. This is impossible when the patches commonly range in size over several orders of magnitude, so the increase in the number of species with increasing area is likely to be underestimated. The problem should be studied to establish its magnitude and to show how it could be avoided.

MEASURES OF ABUNDANCE AND WHEN AND HOW TO USE THEM

FREQUENCY OF OCCURRENCE OF EACH SPECIES

The results for any given species may be reported either as the frequency with which it is recorded in a series of counts, or as the average number detected per count. Blondel (1975) noted the curvilinear relationship between these two measures (Fig. 3) but did not recognize that a curve could be fitted to it by assuming a Poisson or binomial distribution. In practice this relationship means that either frequency of occurrence or average number is an adequate measure for species which occur usually as one or none in each counting unit. On the other hand, frequency becomes an increasingly insensitive measure for species found in larger numbers. Robbins and Van Velzen (1974) claimed that frequency was more suitable for statistical testing with species found in colonies or flocks, but this

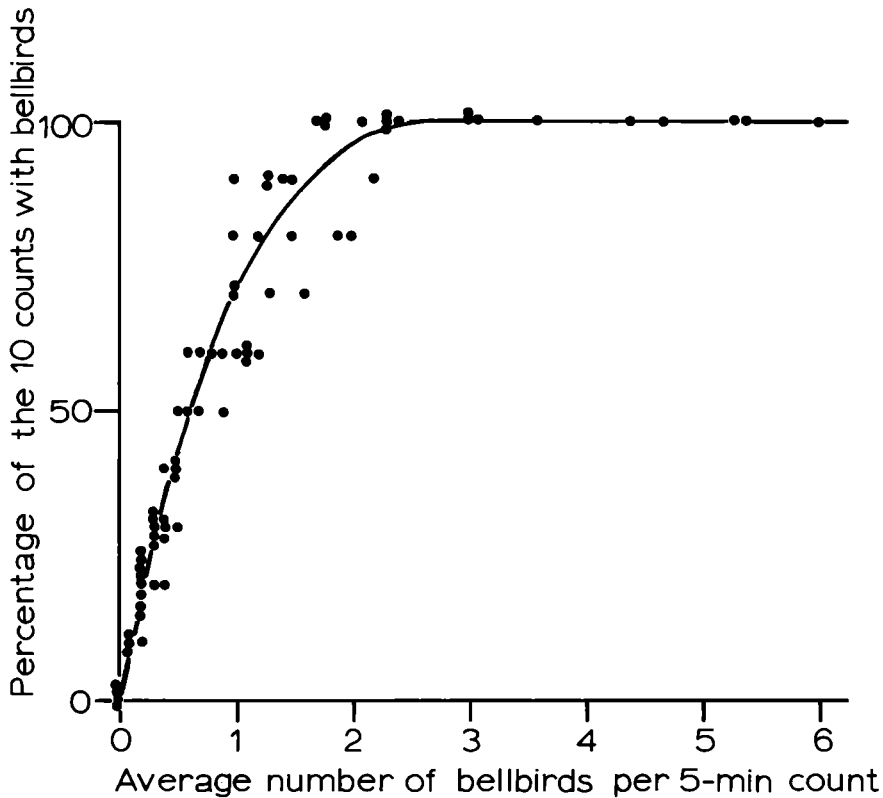


FIGURE 3. Relationship between the mean number of Bellbirds (*Anthornis melanura*) recorded in a series of counts and the percentage of the counts with bellbirds. The line is the expected percentage if Bellbird counts fit a positive binomial distribution with $K = 3$ (Elliott 1971). Other species are better described by Poisson or negative binomial models.

should not be so, as frequency uses less of the available information than do other measures (e.g., rank order).

Blondel et al. (1981) use frequency mainly to estimate the average number of species at a station, which they say is related to density (but see my discussion of this measure [my S_1] above). They claim that frequency is less affected by diurnal changes in detectability than is the average number, but this simply reflects the insensitivity of frequency and so is not an advantage.

Frequency better reflects density when the sampling time it is calculated from is shorter, but I can find no study of units less than five minutes and its deficiencies are still evident at this level (Fig. 3).

THE NUMBER COUNTED OF EACH SPECIES

The numbers of individuals of a species counted in several sampling units (mapping plots, points, transects, or parts of a transect) may approximate a normal distribution, but typically

the distribution is positively skewed, especially if the average number counted per unit is low (Elliott 1971), or if the species is confined to some parts of the sampling area (Dawson and Bull 1975, Ferry 1974, Gur'ev and Zubtvochkii 1974, Robbins and Van Velzen 1970). Given the large number of factors that may influence the numbers counted (Dawson 1981a), it is difficult to use the distribution of counts to obtain accurate details of the actual distribution of birds. For example, the use by Källander et al. (1977) of the average of positive records in five-minute stops as an index of flock size for a species is probably misleading.

If the counts are normally distributed and their variance is not correlated with their mean, parametric statistical techniques such as t -tests, regression and analysis of variance may be used (Conner and Dickson 1980). However, the variance of counts usually increases with increasing mean count (Fig. 4); the variance commonly approximates the mean, as is characteristic of a Poisson distribution (e.g., the Tomtit in Fig. 4).

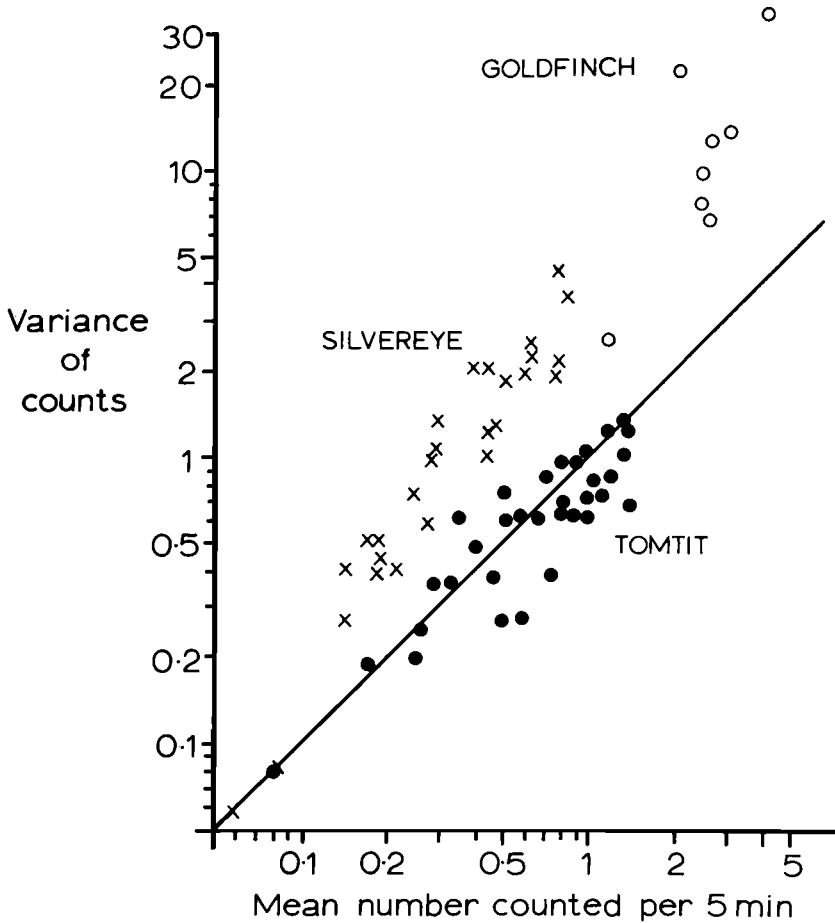


FIGURE 4. Relationship between the mean number counted per 5 minutes, and the variance of the counts, for three New Zealand species. Silvereyes (*Zosterops lateralis*) and Tomtits (*Petroica macrocephala*) were counted in forest on the Victoria Range, north Westland, with 12 to 42 counts contributing to each point (H. A. Best pers. comm.). Goldfinches (*Carduelis carduelis*) were counted in orchards and pastoral land in Hawke's Bay, with 96 counts contributing to each point. For the Tomtit, means approximately equal variances as in a Poisson distribution. The other two species' counts are more aggregated, so statistical tests assuming Poisson are not appropriate.

McCaughran and Jeffrey (1980) outline a binomial test for use with two small samples, and a large sample, normal approximation is available. Provided the total number of birds counted in each sample (nm) is greater than 30, a normal deviate is given by: $z = (m_1 - m_2)(m_1/n_1 + m_2/n_2)^{-1/2}$, where m is the mean number per unit and n is the number of sampling units counted (Elliott 1971). The special case where $n_1 = n_2$ is optimum for detecting differences, and the equation simplifies to $z = n^{1/2}(m_1 - m_2)(m_1 + m_2)^{-1/2}$. This is identical to a chi-squared test of equality of the total number counted in the two samples (nm):

$$\chi^2 = n(m_1 - m_2)^2 / (m_1 + m_2) \dots (1)$$

observing that, for one degree of freedom, χ is a normal deviate.

Rearrangement of these equations gives a form which assists the design of counting studies. We define the overall mean number per count as:

$$m = (m_1 + m_2) / 2 \dots (2)$$

and the percentage difference between the mean numbers in the two samples as:

$$d = 100 |m_1 - m_2| / [(m_1 + m_2) / 2]$$

using the overall average as the base for the percentage. This simplifies to

$$d = 200 |m_1 - m_2| / (m_1 + m_2) \dots (3)$$

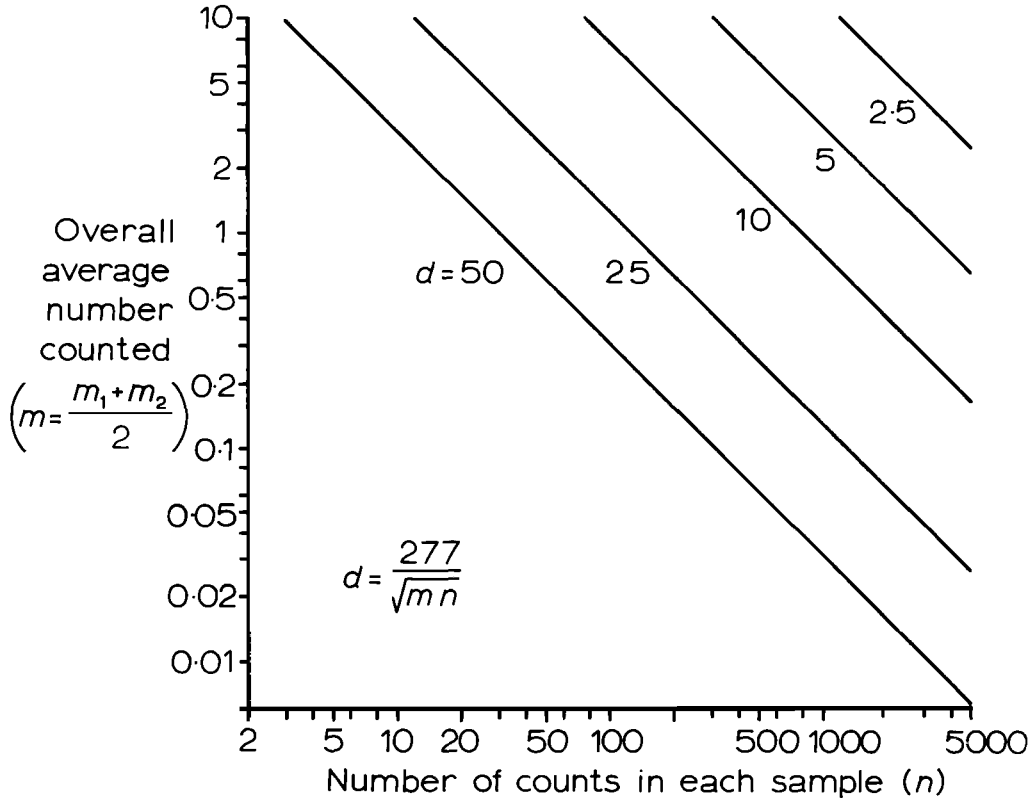


FIGURE 5. The percentage difference (*d*) between the counts obtained in two areas that can be detected with a chi-squared test for $P = 0.05$, given the number of counts in each area and the overall mean count. For example, if a species averages one per count, 770 counts are needed in each of two samples to permit a 10% difference to be detected in a chi-squared test; i.e., the average counts in the two areas could be as close as 0.95 and 1.05 before the test failed to reach the significance level.

Substitution in (1) for $m_1 - m_2$ from (3) and then for $m_1 + m_2$ from (2) gives:

$$\chi^2 = nd^2m/20,000.$$

For a result significant at the 5% level, χ^2 must be greater than 3.84: $3.84 < nd^2m/20,000$, or $n > 76,800/d^2m$; a form that permits estimation of the number of units needed to detect a given percentage difference between the samples when the average count of the species is known (Fig. 5). The requirement that $nm > 30$ corresponds to the area above and to the right of the line $d = 50$ in the graph. Dawson and Bull (1975) first proposed this method in their Table 3. The same relationship may be expressed as $d > 277(mn)^{-1/2}$, from which it can be seen that the percentage difference that may be detected decreases with the increasing total count of a species (nm), so that more units must be counted to detect the same difference with a rare species than are needed with a common one.

If a different number of counts is made in each sample, chi-squared tests approximate the normal deviate test (χ/z is less than 1.1) so long as the smaller sample is at least half the size of the larger one. Outside this range the normal deviate should be used.

Chi-squared tests have commonly been used outside of the restricted range discussed above—when the counts may not fit a Poisson distribution, for more than two samples, or for sample sizes differing more than twofold (Taylor 1965, Dawson et al. 1978)—with little or no justification that the assumptions of the test are met. Elliott (1971) advises the use of appropriate transformations in these situations to bring the distribution nearer normal and to stabilize the variance, or the use of nonparametric methods. The choice will depend on the complexity of analysis needed, as the number of factors readily taken into account by nonparametric methods (Siegel 1956) is limited. Robbins and Van Velzen (1969) used square-root transformations.

Empirical study shows that chi-squared tests remain a reasonable approximation provided the Lefkovich index of dispersion is in the range of -0.2 to $+0.2$ (Fig. 6); the test is conservative below -0.2 and optimistic above $+0.2$. Counts of many species lie within these limits, but for those that do not, routine use of transformations or nonparametric tests is not laborious in these days of electronic computers.

SAMPLING DESIGN

COMPARISON OF YEARS

Studies that seek to monitor long-term changes in bird numbers have used a variety of techniques: mapping (Svensson 1974a, Williamson and Homes 1964), point counts (Robbins and Van Velzen 1967, Källander et al. 1977, Svensson 1977b) and line transects (Sammalisto 1974, Järvinen and Väisänen 1976c). All such studies seek an index of density rather than an estimate of the actual population, as the interest is simply in whether numbers increase or decrease each year. For this, point or transect counts are more cost-effective than mapping (Dawson 1981c).

An index of numerical change between years is better the more replicate sites are used to assess it, and the better control there is over other possible influences on the counts (e.g., time of year, observer, time of day, and weather; Cochran 1963, Taylor 1965, Robbins and Van Velzen 1969, Källander et al. 1977).

THE EFFECT OF A TREATMENT

Shields (1979) discusses the use of "control" areas and treated areas to distinguish the effect of a treatment from normal seasonal changes in density and conspicuousness. Here too replication is important, and Connor and Dickson (1980) discuss details of this for transect counts.

COMPARISON OF HABITATS

Point counts have been used to discover sites of conservation value and to document the distribution of rare species (Crook et al. 1977, Ramsey et al. in press, Svensson 1977a). A very large number of sites must be counted to achieve an accurate index of the abundance of a rare species; actual densities are even more difficult to assess. Figure 5 shows that the detection of a 50% difference in numbers between two areas would require 3000 units to be counted in each for a bird that is found in only one unit out of a hundred, but those same 3000 counts would permit a difference of 5% to be detected for a species that averages one per unit. The solution to this problem in sampling rare species is a two-stage process: (i) a reconnaissance survey to establish the distribution of the rare species, and

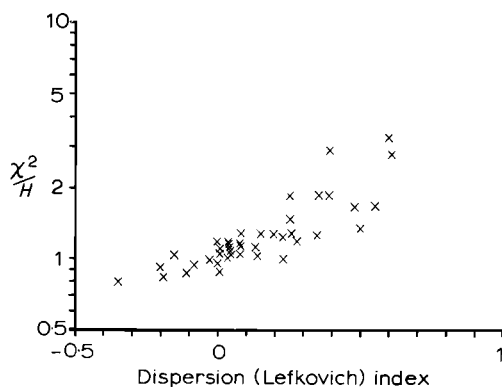


FIGURE 6. The effect of dispersion on the performance of chi-squared tests—a comparison of chi-squared tests with a more rigorous Kruskal-Wallis one-way analysis of variance. Each point is based on counts at nine counting stations for one species in the Victoria Range study of Figure 5. The Kruskal-Wallis "H" is distributed as chi-squared with 8 degrees of freedom (Siegel 1956), so the tests are equivalent when the fraction χ^2/H is unity. This is approximately so for Lefkovich (1966) indices of -0.2 to $+0.2$. The Lefkovich index is $1/45 \tan^{-1}(\text{variance/mean}) - 1$.

(ii) counts of those areas alone. Similar reasoning can be applied to habitats as well as species: random or systematic sampling of large areas results in too many counts in the large habitats and too few in small ones. Again the solution is to establish the distribution of habitats first and then to sample each with a different intensity depending on its extent (Cochran 1963).

An index of differences in bird numbers between habitats can be less variable the more replicates it is based upon (Cochran 1963 gives methods for optimal allocation of effort within and between replicates). And other factors that influence the counts, such as the time of year, should be standardized or studied. For example, Dawson et al. (1978) used four observers to count in four study areas. On each visit each observer counted in all four areas (one per day) and all four areas were counted simultaneously on each of the four consecutive days. This design held observer, time of year, and weather more or less constant. Seasonal changes were studied by visiting the areas at two-month intervals through the year.

For some studies both time and place are factors in the sampling design. For example Williamson (1969) studied the change of habitat preferences with time and was able to demonstrate that Wren (*Troglodytes troglodytes*) numbers varied between years much more in sub-optimal than in optimal habitat. Dawson et al.

(1978) suggested that some New Zealand forest birds show altitudinal migration.

Techniques giving an index of abundance are usually adequate for comparing habitats, but mapping techniques (Cousins 1977) permit greater precision of habitat definition than do point or transect counts. Another problem with using indices to study habitat selection is that detectability can vary between habitats (Dawson 1981a). An estimate of density may be needed to provide a correction factor.

COMPARISON OF SPECIES

To compare species with each other also requires estimates of density, as indices measure each species on a different scale. Densities may also be needed to examine species diversity (but

see my discussion of this above), energetic or trophic relationships.

Estimates of density from mapping, or by means of the more cost-effective point or transect methods (Dawson 1981c) are imprecise, but may often be sufficient, given the magnitude of the errors in measuring other factors, such as food intake. More precise methods, such as capture-recapture (Frochot et al. 1977), or intensive observation of marked individuals, are also more time-consuming but may be needed if the imprecision of the estimate of density limits the accuracy of the overall calculation.

ACKNOWLEDGMENTS

A. E. Aldridge, P. C. Bull, M. Harrison, E. B. Spurr and R. H. Taylor made constructive comments on the manuscript.