Studies in Avian Biology No. 6:24–29, 1981.

THE CHRISTMAS BIRD COUNT: AN OVERLOOKED AND UNDERUSED SAMPLE

SUSAN RONEY DRENNAN[1]

ABSTRACT.—The Audubon Christmas Bird Count represents the most extensive, longest-term, continuous, and most geographically comprehensive data set in American ornithology. It provides the empirical basis for an increasing number of research studies, particularly longitudinal analyses of relative abundance of species and their dynamics, and definition and spatial abundance of the early winter ranges of the avian species of North America. This paper assesses the numerous assumptions involved in making statistical inferences from Christmas Bird Count data. Both biological and statistical constraints inherent in those data are addressed. The results suggest that the Christmas Bird Count has properties comparable to a valid probability sample, when latitudinally stratified. Further, that robust estimation methods should accommodate deviations intrinsic in the data.

The Constitution of the United States established the decennial census of human population numbers in 1790. In 1900 the Audubon Societies, parent organization of the National Audubon Society, inaugurated the Christmas Bird Count. Neither was originally conceived as a data base for statistical inference to increase knowledge of human or avian populations, but the world has changed; and they have been so used. Owing to their higher degree of reliability, the United States Census Bureau demographers prefer utilizing their sample estimates, based upon age-cohort survival tables and fertility rates, rather than the census "official" population count. Similarly, researchers should be wary of inferring too much from Christmas Bird Count (hereafter, CBC) raw numerical data, as it is not a census by strict definition of that term.

The Audubon Christmas Bird Count is the single, most popular, voluntary, early winter bird continental inventory in the world. The 1979–80 CBC had 33,022 participants who censused 1320 count units in North, Middle and South America, and Hawaii. A count unit is defined as that area contained within a circle of 24.1 km diameter. With few exceptions, each count unit is discrete, having no common parts. Counts must be made during a single calendar day within the official CBC period, two weeks centered around Christmas Day. Searching the count unit is accomplished by parties of observers of varying numbers. Every individual bird encountered is included in the inventory. Results of each count unit are reported on standardized forms, which also solicit details on weather, habitat coverage, methods of canvassing the area, natural food resources, and hours afield and miles covered. Count statistics have been published annually for the past 80 years. The CBCs represent the most extensive, longest-term, continuous, and most geographically comprehensive data set in American ornithology.

With increasing frequency, ornithologists interested in winter population trends, winter range extensions, and winter bird distribution have been turning to those data accumulated in the CBCs. Preston (1980) has used CBC data to analyze distributions of commonness and rarity. Major patterns of avian species diversity and abundance have been analyzed by Bock and Lepthien (1974, 1975b, 1976d), Bock, Mitton and Lepthien (1978), Tramer (1974), Smith (1979), and Falk (1979), to name but a few. Several recent longitudinal studies of single species density have been done by Brown (1973, 1975, 1976b), Bock and Lepthien (1975a, 1976b), Stahldecker (1975), Anderson and Anderson (1976), Johnsgard and DiSilvestro (1976), Kennard (1977), Morrison and Slack (1977), Neidermyer and Hickey (1977), Aldrich and Weske (1978), Bonney (1979), Raynor (1980), Rosahn (1980), and Stewart (1980).

Relative abundance maps seemingly provide viable results when the density of a fairly ubiquitous species is plotted by using the number of individuals recorded per 1, 10, or 100 party-hours as the basic standard (Bystrak and Drennan 1975, Plaza 1978). In 1974, the United States Air Force funded a project to map the winter distribution and relative abundance of 143 selected species considered the greatest potential threat to low-level-aircraft operation in the United States (Bystrak et al. 1974). These maps were plotted by using CBC data.

The CBC provides the empirical basis for a growing number of studies, particularly longitudinal analyses of definition and spatial abundance of the early winter ranges of the avian species of North America. Critics (e.g., Stewart 1954, Kenaga 1965, Robbins and Bystrak 1974) question the extent of scientific usefulness of CBC data, citing the lack of rigid controls as an inherent limitation. Others (Hickey 1955) argue that the CBC should be disregarded as a scien-

[1] National Audubon Society, 950 Third Ave., New York, New York 10022.

tific tool and be considered ". . . essentially a recreational activity in which distinct elements of competition, surprise, rarities and the big list are bright and personally thrilling."

This paper evaluates the present CBC and finds that it may provide useful and valid statistical estimates. This is suggested by its size, design (unplanned), and consistency of results. This paper does not address the issue of unstandardized count methods, at which most of the criticism has been directed. However, the results here suggest that that lack of standardization and imperfections in some areas of data collection may not be critical for use in certain types of analysis.

THE CBC AS A SAMPLE

Inasmuch as a census is a total count of all the members in an unambiguously defined universe, or a measure of a characteristic of all the members in a population, the CBC is not a census of North American birds. There have been no bird censuses of any continent for any widespread or numerous species, a fact which needs emphasis at this symposium. Practicality limits the use of a census to situations where the population being counted is very narrowly defined. Otherwise, drawing inferences about the population under study is accomplished through use of a sample, or subset of the individual items in an unambiguously defined population. By this means, unknown population parameters (e.g., its size: N, the mean: μ, and standard deviation: σ, of some population characteristic), can be estimated and the likely range of error of such estimates computed, provided that certain conditions are satisfied.

For example, let K equal the total number of subareas or count units, each of fixed area A, into which North America can be partitioned, such that each point on its surface lies within one and only one uniquely identified count unit. Let k equal a subset or sample from the K count units, such that $0 < k < K$. Let n_{ij} equal the number of birds of species j observed in the ith count unit ($i = 1, \ldots k$), on a given day in December. The mean absolute density of species j per area A (recalling that all count units are of equal size, A), for the sample of k count units is computed as

$$\bar{n}_j = \frac{\sum_{i=1}^{k} n_{ij}}{k} \qquad (1)$$

Then the estimate of N_j, the total number of birds of species j in North America, is: $\hat{N}_j = K(\bar{n}_j)$. Because $\bar{n}_j$ is computed from a sample, it is subject to error, and consequently, the estimate of N_j is subject to error. The standard error (SE), for n_j is given by:

$$SE = \sqrt{\frac{\sum_{i=1}^{k} (n_{ij} - \bar{n}_j)^2}{k(k - 1)}} \qquad (2)$$

So the 95% confidence interval for N_j is: $\hat{N}_j \pm 1.96K$ (SE). That is, there is a 95% probability that the true total number of wintering birds of species j in North America lies within the stated numerical limits.

At the outset of this example it was noted that such sample estimates of unknown population values are valid provided that certain conditions are met. The main conditions are: (1) the k count units selected for the sample represent a random sample from the total K count units (i.e., every count unit in K has an equal likelihood of being selected for the sample), and (2) the coverage and method of data collection in each of the k sample count units is the same, and every sample count unit is, in fact, covered.

The issue at hand is whether or not the CBC meets these conditions.

Regarding condition (1), the randomness of the sample; we have already established that the CBC is not a continental census. Additionally, it is not a simple random sample, because not every possible count circle in North America has an equal chance of being selected for the sample. The use of circles instead of hexagons or rectangles as the actual CBC count unit does not bias the results, as all of the circles are of equal size and are virtually mutually exclusive. Condition (2), method of coverage, is discussed elsewhere in this volume by Arbib (1981). Although the set of CBC circles covered each year is an accident of history, highly correlated with human population centers, it has properties comparable to a valid probability sample. A probability sample (including simple random samples) is a sample design in which the sampling units are selected according to laws of chance, such that the probability of any unit being included is known and not zero for every unit in the population.

As stated, condition (1) is not met by the CBC because the count areas included annually do not constitute a random sample from the set of all possible North American count areas. However, because geographical distribution of bird species is so uneven (nonrandom), a simple random sample would be inappropriate. The more appropriate sample design called for is a stratified sample, in which the probability that any count unit is selected is higher for high species density areas and lower for low species density areas. The most efficient distribution of sam-

TABLE 1
1979–80 CBC DISTRIBUTION: NUMBER OF SPECIES BY LATITUDINAL BELTS

Number of species	Number of counts by latitudinal belts							Total counts
	25–29°	30–34°	35–39°	40–44°	45–49°	50–59°	60+°	
0–9	0	0	0	0	2	2	1	5
10–19	0	0	2	5	13	4	6	30
20–29	0	2	7	30	46	11	2	98
30–79	6	45	207	376	87	20	2	743
80–89	2	25	31	29	5	0	0	92
90–99	2	36	16	12	4	1	0	71
100–149	35	59	49	38	17	0	0	198
150+	22	14	17	2	0	0	0	55
Total counts	67	181	329	492	174	38	11	1292

pling effort is contingent on concentration of effort increasing as avian density increases. As Caughley (1977) and others have shown, it borders on the absurd to intensively sample an area containing few animals.

It can be shown that, although the present CBCs are not the result of a rigid sample design, sampling effort is clearly more intensive in high species density areas than in low density areas. Table 1 presents the distribution of CBCs by reported species density (i.e., number of species) and latitude belt for 1979–80. Of 1292 CBCs (excluding the handful of counts south of latitude 25°N), 324 or 25.1% had 90 or more species. By contrast, 133 CBCs or 10.3% had 29 or fewer species. Thus, the sampling effort is more concentrated in high density areas.

If the goal of the CBC was to estimate density and absolute number of a single avian species, the ideal stratification of count areas would be based upon density of that species. However, if the goal was to estimate density and absolute number for as many different species as possible, then the ideal stratification of count areas would be based upon density of all species. The implication here is that sampling effort would not be randomly distributed over North America but would tend to be concentrated on coastal areas and in lower rather than higher latitudes. As we cannot know at the outset that which we would like to measure, density of species for every area of North America, the stratification should be based upon one or more characteristics which are measurable and are known to be correlated with number of species, e.g., latitude. Table 1 certainly supports the hypothesis of an inverse relationship between species density and latitude (see also Bock and Lepthien, 1974). Table 2 compares the distribution of the 1979–80 CBC with stratification based upon latitude belts. With the population domain defined as North America (excluding Greenland), the total area of 20.5 million km^2 is partitioned into latitude belts, and areas within each are approximated in column 2. Every CBC circle encompasses an area of 457.9 km^2. If A_1 equals the area of any latitude belt (in thousands km^2), then column 3 shows the population of count units (K) within each latitude belt, i.e., $K_1 = A_1/0.4579$. Column 4 shows the actual number of CBCs conducted in each latitude belt (k) in 1979–80. The total of 1292 sample count units conducted represents 2.9% of the land area of the continent. In other words, one out of every 34 possible count units was included in the sample.

With regard to flocking birds, an observation by Caughley (1977) is entirely applicable: that the higher the sum of sampled units, the more accurate the estimate; and the more clustered the animals, the greater the number of sampling units required to render a density estimate that is reasonably accurate. Note that percent of land area included in the sample tends to increase with decreasing latitude. The historical reasons for this are, of course, that: (1) the geographic distribution of birders is similar to the geographic distribution of people, and (2) birders tend to pick CBC areas relatively near their homes. Those two unchallenged principles are fundamental to past criticisms of the CBC as not representing a random selection of count areas. But the historical reasons are not germane. The de facto outcome is a sample in which the sampling effort (i.e., the percent of area covered) tends to increase with decreasing latitude. That is a highly desirable feature as avian winter populations are not randomly distributed, and their distribution is inversely correlated with latitude. Therefore, although ca. 54% of the area of North America lies above 50°N latitude, it would be inefficient and wasteful to devote ca. 54% of the

TABLE 2
STRATIFIED SAMPLE DESIGN: 1979–80 CBC

North latitude	Area, million km² (±)	Potential CBC units	Actual CBC units	% Potential CBC units sampled	Number of species		Coefficient of variation ($s/\bar{X}$)
					Mean ($\bar{X}$)	Std. Dev. (s)	
60°+	5.4	11,790	11	0.1%	19.6	13.2	0.673
50–59°	5.7	12,450	38	0.3	32.9	16.1	0.489
45–49°	2.6	5680	174	3.1	46.2	30.1	0.652
40–44°	2.3	5020	492	9.8	58.8	24.3	0.413
35–39°	2.1	4590	329	7.2	77.8	34.5	0.443
30–34°	1.6	3490	181	5.2	99.8	32.5	0.326
25–29°	0.8	1750	67	3.8	129.9	33.8	0.260
Total average	20.5	44,770	1292	2.9%	70.3[a]	37.1	0.529

[a] Weighted average.

sampling effort, namely ca. 697 CBC circles, to that area.

Thus the problem of latitudinal nonrandomness of sampling is not as critical as has been previously supposed. Nonetheless, it is important to note other elements of bias, e.g., the concentration of effort expended sampling urban or semi-urban habitats and coastal habitats. Through statistical weighting of disproportionate habitats (biased samples), perhaps errors in estimation can be minimized.

Means ($\bar{X}$) and standard deviations (s) of species density for each belt indicate that: (1) species density increases with decreasing latitude, and (2) variability is decidedly higher below latitude 50°N than above. Because variability is not homogeneous, a stratified sample is more appropriate, assuming that our aim is to measure species density or absolute density, for each of many species. In the last column of Table 2, the gain from stratification is clearly seen in the coefficients of variation ($s/\bar{X}$). The coefficient of variation for the total area is .529, which is larger than the coefficient in five of the seven latitude belts, indicating that variability is reduced by stratification. When the absolute density of every species in an area to be sampled is heterogeneous, some stratification is preferable to none. It is certainly appropriate to a sample design that estimates species density by direct counting. An advantage that stratified sampling has over simple random sampling is that it divides broad heterogeneously dense areas into discrete sampling units, within which species density is roughly homogeneous. Species densities of each unit are inventoried. The mean density of inventoried units is used as an estimate of mean density for inventoried and uninventoried units combined. This consequently increases the precision of the estimate, as it is now a function of density within stratified zones, as opposed to density over a broad area. Precision of density estimates for each species is thus inversely related to the density variability within the entire area under survey.

The CBC sample size and sample design (planned or accidental) may then be useful. What about the sampling results? Tables 3 and 4 compare broad measures of results from the previous two CBCs (1978–79 and 1979–80). They show that the allocation of the sample is closely consistent for these two different periods. In Table 3, the mean number of species observed by latitude belts are almost identical in all but the lowest latitude belt. In Table 4, the percent distributions of counts by reported species-density classes are practically identical.

Thus the evidence supports the hypothesis that CBC data may be useful when treated as a stratified sample. There are three reasons supporting that conclusion. First, the de facto sample design closely conforms to a stratified sample, in which sampling effort is more intensive in high-density areas than in low-density areas. Second, the sample is impressively large (one out of 34 potential count units are included). Third, the allocation of the sample for two different periods is highly consistent.

One might argue that the validity of the sample design is limited because the sample count units, within each latitude belt, are not selected by random process. The response to that objection is pragmatic. A random selection of count units would of course, require assigning birders to those sample areas selected. Theoretically, under such conditions, it is entirely possible that the number of willing participants would plummet dramatically. Perhaps 200± count units would be covered. This would result in a significant rise in the standard error (SE) of estimates. If ca. 200 CBCs were covered, then SE would tend to increase by a factor of ±2.6, as shown.

TABLE 3
CBC RESULTS: MEAN NUMBER OF SPECIES BY LATITUDE BELTS

North latitude	Mean number of species	
	1978–79	1979–80
60°+	20.3	19.6
50–59°	33.7	32.9
45–49°	45.4	46.2
40–44°	57.7	58.8
35–39°	77.6	77.8
30–34°	97.5	99.8
25–29°	120.0	129.9

TABLE 4
CBC RESULTS: COUNT DISTRIBUTION BY NUMBER OF SPECIES

Number of species	Percent distribution CBCs	
	1978–79	1979–80
0–9	0.6	0.4
10–19	2.4	2.3
20–29	7.4	7.6
30–79	57.6	57.5
80–89	7.9	7.1
90–99	5.4	5.5
100–149	14.1	15.3
150+	4.6	4.3
	100.0	100.0

The basic formula for SE is

$$SE = \frac{s}{\sqrt{k}}\sqrt{1 - \frac{k}{\Sigma K}} \quad (3)$$

With the present (1979-80) CBC, the number of sample count units, k, is 1292; ΣK, the total number of potential count units in North America, is 44,770. The standard deviation, s, for some variable measured with the sample, such as a particular species density, need not be numerically specified to illustrate the point. So with the present count,

$$SE = \frac{s}{\sqrt{1292}}\sqrt{1 - \frac{1292}{44{,}770}} \quad (4)$$

$$SE = 0.0274s$$

That is, SE is ±2.7% of s. With a random selection of count units within each latitude belt and the assignment of participants to sample areas resulting in a drop of k to ca. ±200 (which may be optimistically high), the new standard error (SE′) would be

$$SE' = \frac{s}{\sqrt{200}}\sqrt{1 - \frac{200}{44{,}770}} \quad (5)$$

$$SE' = 0.0706s$$

That is, SE′ is ±7% of s. Thus, by substitution, SE′ = 2.575 SE. The SE′ is >2.6 times larger than SE, which is a marked reduction in precision of sample estimates. It is difficult to imagine that any gains from instituting a random selection process would be worth that loss of precision.

Another objection is that the tendency for birders to select count areas of high habitat variation and, consequently, avian density, may introduce an upward bias in sample estimates within any latitude belt. However the size of that bias could be estimated with some controlled field work and, in the interest of standardization, could be used to derive more satisfactory sample estimates.

An additional point is worth making here: although the CBC data appear to be useful as an estimation scheme when treated as a stratified sample, stratification by latitude belts is neither the only nor necessarily the optimal stratification. As already noted, bias which may be introduced by disproportionate habitat coverage (e.g., urban vs rural, coastal vs interior) is not eliminated by latitudinal stratification. To determine the best stratification requires further work. Researchers might try stratification schemes based on distinct physiographic regions, riverine systems or coastal zones. Stratification based on fixed or variable distances from urban centers may help in removing the well known urban sampling bias. Stratification based on botanical or ecological homogeneity may provide stimulating insights into avian densities. These few designs suggest the potential for CBC analysis according to stratification.

CONCLUSIONS

On balance then, the CBC is an enormously rich data source which may be useful for estimating population parameters, provided that researchers are aware of its limitations. Biostatisticians sometimes caution that only random sampling ensures true representation of the population under study. That is, only randomness ensures consistency, efficiency, sufficiency, and unbiasedness. The data of the CBC do not perfectly conform with the underlying assumptions (e.g., random selection of count areas, standard procedures of coverage, equal effort in all count areas, etc.) assumed in applying statistical procedures such as calculating mean number of species or absolute density for a species. But those statistical procedures are known to be so robust that departures from the assumptions within the data can be encompassed, especially when the sample is large. The methods of statistics are exact; the real world data to which they are applied, here CBCs, are inexact. So although we know that within, say the 40°–45°N latitude belt, some counts had two or three parties, others six or more parties, and not all had the same number of parties as last year, much less the same weather conditions, the same routes, the same participants, party composition, or collective skill, nevertheless, the mean number of species for the belt was 58.8 in 1979–80, almost identical to the mean of 57.7 in the prior year (1978–79). In other words, the lack of standardization may be so much statistical "noise" that tends to disappear when results are aggregated for broad areas. Curiously, the limitations of the CBC, this "noise," have received so much casual attention that the baby has too often been discarded with the bath water. There is no other branch of field zoology which has any sample comparable in size, scope, and regularity to the CBC, and yet ornithologists have hardly begun to exploit it.