Studies in Avian Biology No. 13:193–198, 1990.

PRECISION, CONFIDENCE, AND SAMPLE SIZE IN THE QUANTIFICATION OF AVIAN FORAGING BEHAVIOR

LISA J. PETIT, DANIEL R. PETIT, AND KIMBERLY G. SMITH

Abstract. We used equations presented by Tortora (1978) to estimate minimum sample sizes for avian foraging data. Calculations using absolute precision provided considerably lower estimates of sample size than those using relative precision. When sample sizes were estimated using absolute precision more observations were required to accurately represent foraging behavior of a generalist than of a specialist, but, for a precision of $\pm 5\%$ with $k = 3$ categories, no more than 572 observations were ever required. The opposite trend was observed with relative precision, such that, for extreme specialists, with $k = 3$ categories, $>100{,}000$ observations were needed to achieve relative precision of 5% around extremely rare behaviors. Because foraging studies typically focus on common behaviors, absolute precision is usually adequate for estimating sample size. Estimates of sample size acquired using Tortora's (1978) equations are dependent upon desired levels of confidence and precision. The estimation method can also be used *a posteriori* to determine precision associated with a sample.

Key Words: Sample size; avian foraging behavior; generalist; specialist; precision.

The increased use of statistics over the last two decades to analyze avian foraging behavior has heightened awareness of the problem of obtaining enough observations for proper analysis. Sample size clearly has a considerable effect on one's ability to make statistical inferences; yet, few attempts have been made to determine the number of observations needed to quantify avian foraging behavior. It would appear that most researchers simply gather the greatest number of observations possible, without much regard for which sample sizes may be appropriate for their analyses. Thus, a great variation in sample sizes of foraging behavior has been reported in the literature, ranging from 20–30 (e.g., Eckhardt 1979, Tramer and Kemp 1980, Maurer and Whitmore 1981) to >1000 (e.g., Holmes et al. 1979b, Sabo 1980, Landres and MacMahon 1983) single point and sequential foraging observations on individual species. Data collected in two or more field seasons are often combined to increase sample sizes, but that practice may not be appropriate because of between-year differences (e.g., Landres and MacMahon 1983).

Only Morrison (1984a) has directly assessed influence of sample size. Based on stabilization of means and narrowing of confidence intervals with increasing sample size, he suggested that a minimum of 30 independent observations (i.e., individual birds) were necessary to quantify foraging behavior of two species of warblers. The point at which confidence intervals are sufficiently narrowed, however, may be difficult to ascertain through simple inspection. In addition, because avian foraging behavior data often are made up of multiple variables dissected into many categories (e.g., "glean," "hover," and "hawk" within the variable, "foraging mode"), Morrison's (1984a) method involved calculating confidence intervals for each category of observations separately, such that minimum sample sizes in his study varied among different categories within the same variable. Further, it is not clear whether Morrison's estimate of sample size is readily generalizable to other passerine species.

Another factor that may influence sample size is variation of behavioral repertoires among species. For example, for a species with a fairly limited repertoire, with most observations falling into one or very few categories (i.e., a specialist; Morse 1971a), adequate sample sizes might be smaller relative to those required to quantify the more diverse repertoire of a foraging generalist. On the other hand, Tacha et al. (1985) indicated that large sample sizes were needed to capture rare behavioral events. If so, more observations will be needed to characterize a specialist's behavior compared to that of a generalist because of difficulty associated with quantification of rare events.

To maximize efficiency in collecting foraging data, some criteria are needed to determine a minimum sample size necessary to quantify such behaviors. Goodman (1965) introduced a procedure based on calculation of simultaneous confidence intervals for a multinomial population. Tortora (1978) modified that procedure for application to the situation in which a random sample of observations (i.e., independent and unbiased observations) are classified into k mutually-exclusive categories, and the proportions in those categories sum to one. (While we acknowledge that there are difficulties associated with obtaining a truly random sample of behaviors in avian foraging studies [e.g., Altmann 1974, Wagner 1981a, Morrison 1984a, Tacha et al. 1985], this is an assumption of all sample size estimation techniques [e.g., Cochran 1977, Steel

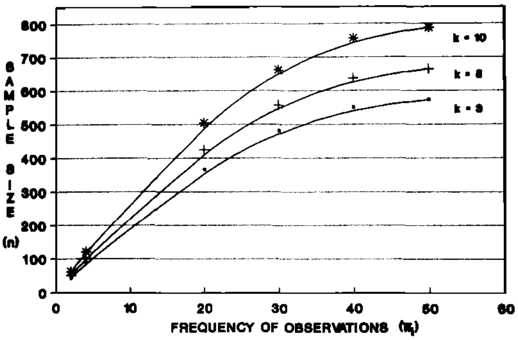

FIGURE 1. Estimation of sample sizes with absolute precision (b_i) of 5% as a function of the frequency of observations in one of 3, 5, or 10 mutually-exclusive categories (k). Confidence level (α) for these estimations is 0.05. See text for further explanation.

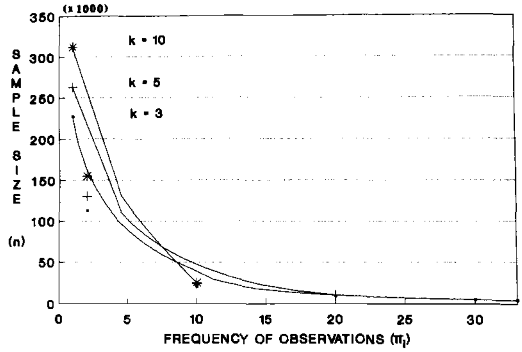

FIGURE 2. Estimation of sample sizes with relative precision (b_i') of 5% as a function of the minimum frequency of observations in one of 3, 5, or 10 mutually-exclusive categories (k). Confidence level (α) for these estimations is 0.05. See text for further explanation.

and Torrie 1980], and it is our intention only to present one of these techniques rather than to discuss the related but separate question of how foraging data are obtained.) In contrast to the methods used by Morrison (1984a), Tortora's (1978) procedure considers all categories simultaneously and allows for estimation of the sample size needed to achieve a specified level of confidence (α-level) such that percentages in all k categories are within some specified range (precision) of the true population values.

Our objectives were to: (1) determine a minimum sample of independent observations necessary to quantify foraging behavior, and (2) determine whether minimum sample sizes are different for specialist and generalist species.

METHODS

Tortora (1978) presented equations for calculating sample sizes based on either absolute or relative precision. (Precision is a measure of variance around the true population mean. Therefore, for these equations we assume that the true population mean is known [i.e., representation of the true mean is accurate]. We discuss below what can be done when the true mean is not known.) Absolute precision refers to the situation in which the acceptable variation around a small proportion is relatively greater than that around a larger proportion. This means that we are more interested in the ability to quantify the most common behavior at the expense of the precision associated with the rarest behaviors. For example, assume that gleans, hovers, and hawks occur with frequencies of 96%, 2%, and 2%, respectively, for a hypothetical foliage-gleaning bird. If we specify an absolute precision of 5%, we would accept foraging behavior estimates of 91–100% (96% ± 5%) for glean and 0–7% (2% ± 5%) for both hover and hawk. The equation given by Tortora (1978) for calculating sample size (n_a; the subscript refers to the type of precision used) with absolute precision is:

$$n_a = B\Pi_i(1 - \Pi_i)/b_i,$$

where B is the critical value of a χ^2 with 1 degree of freedom at a probability level of α/k (k = number of categories), b_i is a specified absolute precision (i.e., acceptable deviation from the true value) for each category i, and Π_i is the proportion of observations in the ith category. Sample sizes (n_a) increase to a maximum as Π_i approaches 0.50 (see Results). Thus, if $b_i = b$ for all categories, one calculates n_a using the Π_i closest to 0.50. If that frequency is >50%, its complementary frequency (i.e., 1 − percent frequency) is used. If $b_i = b$ for all categories, the largest n_a is chosen as the minimum sample size, and if the true population mean is unknown, one can calculate a "worst case" sample size by using $\Pi_i = 0.50$ (see also Discussion).

Relative precision refers to when the acceptable relative variation around the smallest proportion is the same as around the largest proportion. For the example mentioned above, we would accept estimates between 91.2–100% for glean for a relative precision of 5% (i.e., ±5% of 96%), but we would now only accept estimates between 1.9–2.1% for hover and hawk (i.e., ±5% of 2.0%). Here, sample sizes will be greatly influenced by attempting to quantify precisely the rarest foraging event. Tortora's (1978) equation for calculating sample sizes (n_r) with relative precision is:

$$n_r = B(1 - \Pi_i)/\Pi_i b_i'^2,$$

where $b_i'^2 = b_i/\Pi_i$, and, if $b_i' = b'$ for all categories, Π_i is the minimum proportion of the k observed proportions (e.g., 2% in the example above). As with absolute precision, if $b_i' = b'$ for all k, choose the largest n_r calculated for the sample size.

RESULTS

Application of Equations

We calculated sample sizes necessary to represent with absolute precision means for six different frequency combinations, for k = 3, 5, and 10 categories (Fig. 1). The relationship between Π_i and sample sizes with absolute precision (n_a)

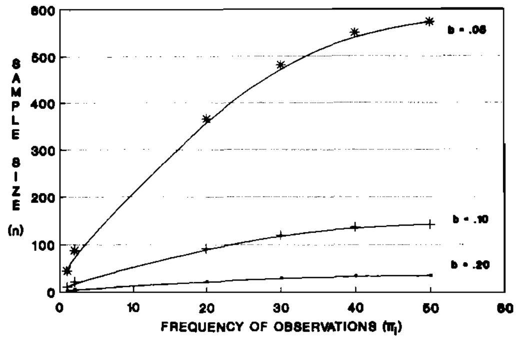

FIGURE 3. Effect of variation in absolute precision (b_i) on estimation of sample size for different frequencies of observations at $\alpha = 0.05$ and for $k = 3$ categories.

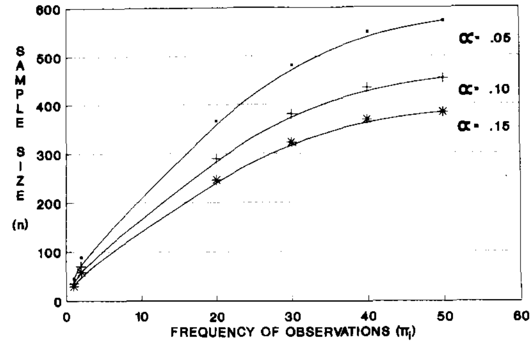

FIGURE 4. Effect of variation in confidence level (α) on estimation of sample size for different frequencies of observations, with precision (b_i) of 5% and $k = 3$ categories.

is such that, as any one categorical frequency approaches 50%, sample size increases for a given α and b_i. Thus, the curve in Figure 1 is symmetrical around $\Pi_i = 0.50$. Consider the situation in which $k = 3$ categories, $\alpha = 0.05$, $B = 5.724$ (χ^2 critical value for P = 0.05/3 = 0.0167), and $b = 0.05$ (absolute precision of 5%). If 98% of the observations are in one category and 1% are in each of the two remaining categories, about 45 independent observations would be necessary to have 95% confidence that the observed (sample) mean is within 5% of the true population mean (Fig. 1). Based on this approach, no more than 572 independent observations would ever be needed to quantify a variable with $k = 3$ categories (e.g., glean, hover, and hawk) at our specified levels of b (=0.05) and α (=0.05). Note, however, that n_a increases as number of categories (k) increases, particularly as Π_i approaches 0.50 (Fig. 1). Assuming those frequency combinations are representative of specialist or generalist species, the results suggest that: (1) minimum sample size is smaller for a species that is specialized in its foraging behavior (i.e., frequency in any category diverges substantially from 50%); and (2) influence of k on minimum sample size is greater for a generalist than for a specialist (Fig. 1).

A potential problem with an absolute precision of 0.05 is that, for example, in the extreme specialist case (98%, 1%, 1%), an acceptable mean would range from 93–100% for the first category and 0–6% for the others, which produces an acceptable range of 600% around the means for the two "rare event" categories. This problem can be remedied by calculating n_r with a relative precision (b_i') for each category. Unfortunately, this results in a large increase in minimum sample sizes (Fig. 2). Those data show that, contrary to estimations using absolute precision, sample sizes estimated with relative precision increase substantially as a species becomes more specialized (i.e., min $[\Pi_i, \ldots, \Pi_k]$ approaches 0). Thus, in the case of an extreme specialist with a repertoire of three foraging modes with percent frequencies of 98%, 1%, and 1%, the minimum required sample size (with $b_i' = 0.05$) is 226,670 independent observations. Again, as with absolute precision, sample sizes calculated with relative precision increase as number of categories (k) increases (Fig. 2). Increases in both specified α and b_i levels cause decreases in sample size estimates with the greatest influence being exerted by changes in b_i (Figs. 3 and 4).

We applied the equations above to foraging data (Table 1) to determine how precisely sample sizes have allowed estimations of "true" population values. Note that all but Morrison's (1984a) are based upon sequential observations. Thus, the assumption of independence of observations for Tortora's equations may be violated, such that precisions we report probably are lower (i.e., better) than the actual precisions associated with those data sets (Tacha et al. 1985).

Table 1 shows that, for example, Morrison (1984a) reported that Hermit Warblers gleaned 78.8% of the time, hover-gleaned 11.5%, fly-caught 3.8%, and performed some other maneuver 5.8% of the time. Assuming those are the true proportions for the population then, based on a sample of 60 independent observations, with $k = 4$ and $B = 6.239$ (for $\alpha/k = 0.0125$), we calculated an absolute precision of 0.1319, or 13.2% (Table 1), meaning that one can expect to estimate within 13.2% of the true values for that distribution of proportions using 60 observations. To achieve 5% absolute precision, Morrison would have needed approximately 417 independent observations (n_a). To achieve relative precision of 5%, he would have required 63,178 independent observations (n_r)!

TABLE 1. REPORTED SAMPLE SIZES (N) WITH ESTIMATED SAMPLE SIZES (n_a AND n_r; BASED ON EQUATIONS PRESENTED IN TORTORA [1978]) FOR THE VARIABLE, FORAGING MODE, FOR SELECTED "SPECIALIST" AND "GENERALIST" SPECIES IN STUDIES OF AVIAN FORAGING BEHAVIOR. PRECISIONS ASSOCIATED WITH REPORTED SAMPLE SIZES ARE REPRESENTED BY b_i (ABSOLUTE) AND b_i' (RELATIVE). CALCULATIONS OF n_a AND n_r ARE BASED ON $\alpha = 0.05$ AND b_i AND $b_i' = 0.05$; k IS THE NUMBER OF BEHAVIORAL CATEGORIES WITHIN THE VARIABLE FORAGING MODE

Species	k	Greatest percent	Smallest percent	n	n_a	b_i	n_r	b_i'	Study
Hermit Warbler									
Dendroica occidentalis	4	78.8	3.8	60	417	0.13	63,178	1.62	Morrison (1984a)
Bushtit									
Psaltriparus minimus	3	96.0	2.0	270	88	0.03	112,190	1.02	Landres and MacMahon (1983)
White-breasted Nuthatch									
Sitta carolinensis	3	51.0	3.0	1430	572	0.03	74,030	0.36	Landres and MacMahon (1983)
Solitary Vireo									
Vireo solitarius	3	66.6	2.6	114	506	0.11	85,772	1.37	Holmes et al. (1979b)
Least Flycatcher									
Empidonax minimus	3	75.4	3.0	609	425	0.04	74,030	0.55	Holmes et al. (1979b)
Western Wood-Pewee									
Contopus sordidulus	3	94.0[a]	1.0[a]	419	129	0.03	226,670	1.16	Eckhardt (1979)
Wilson's Warbler									
Wilsonia pusilla	3	75.0[a]	8.0[a]	427	429	0.05	26,330	0.39	Eckhardt (1979)
Red-eyed Vireo									
Vireo olivaceous	3	61.0	11.5	150	545	0.09	17,620	0.54	James (1976)
White-eyed Vireo									
Vireo griseus	3	59.0	10.0	132	553	0.10	20,606	0.62	James (1976)
Black-throated Blue Warbler									
Dendroica virens	3	43.3	26.7	30	562	0.22	6,286	0.76	Maurer and Whitmore (1981)
Acadian Flycatcher									
Empidonax virescens	3	59.7	1.4	72	551	0.14	161,253	2.37	Maurer and Whitmore (1981)
Prothonotary Warbler									
Protonotaria citrea (males)									Petit et al. (1990)
Pre-nestling period	3	91.0	0.4	1393	188	0.02	570,110	1.01	
Nestling period	3	77.8	3.5	630	396	0.04	63,128	0.50	

[a] Proportions estimated from histograms.

Absolute precisions (b_i) associated with reported sample sizes (n) in Table 1 ranged from 0.02 (Petit et al., unpubl. data) to 0.22 (Maurer and Whitmore 1981) and, in general, most observed sample sizes corresponded to absolute precisions within 10% of the true proportions (with 95% confidence) in each category of foraging mode (Table 1). It is perhaps not surprising that none of the observed sample sizes (n) provided acceptable relative precisions.

DISCUSSION

Tortora's (1978) equations provide a useful and straightforward method for estimating sample sizes for quantifying foraging behavior. However, such dramatic differences between sample sizes calculated using relative and absolute precision prompts the question: How much precision is necessary? A minimum necessary sample size of 600 is infinitely more attractive (and attainable) for field researchers than is one of 50,000. Although some attention has been paid to methods that quantify rare events (e.g., Wagner 1981a, Morrison 1984a, Tacha et al. 1985), most studies have focused only on common behaviors, because extremely rare behaviors (e.g., 1–5% of all maneuvers) are usually relatively unimportant in characterizing the general foraging behavior. Thus, for most studies, it may be sufficient to calculate sample size based on absolute precision, provided that the acceptable confidence interval is relatively small. The decision of what constitutes an acceptable absolute precision or confidence level may depend on the objectives of the study in question and is always at the discretion of the investigator. We chose $\alpha = 0.05$ and $b_i = 0.05$ based on standard statistical criteria (i.e., α-level of significance [α/k is similar to calculating an experimental error rate]). However, these specifications may be unnecessarily stringent. Several recent papers (e.g., Thompson 1987; Angers 1979, 1984) have criticized Tortora's method for being too conservative (i.e., estimating larger sample sizes than necessary), and proposed variations in the estimation technique, making it more liberal (i.e., lowering estimated sample sizes). The technique proposed by Angers (1979, 1984), however, involves tedious calculations. Moreover, the methods proposed by both Thompson (1987) and Angers (1979, 1984) do not improve greatly on the applicability of Tortora's original modification of the estimation technique, and thus, do not decrease its validity.

Given the conservative nature of Tortora's method, one may be justified in relaxing levels of confidence or precision or both when using the equations. It is reasonable to set $\alpha/k = 0.05$ and/or to accept a precision of 10% or even 15%, either of which will lower the minimum number of samples needed (Figs. 3 and 4).

An implicit assumption in using Tortora's equations is that the theoretical frequency to be observed in each category does not change through time. This is difficult to meet in foraging studies because a species' behavior can differ between sexes (e.g., Morse 1968), within a season (Morse 1968, Sherry 1979), and between years (Landres 1980). To meet that assumption, sample sizes would have to be estimated for each category depending on the temporal or spatial scale at which the research is conducted and the objectives of that research. Using the equations presented in this paper, researchers can estimate a required sample size at any required confidence level (α) or precision.

Although sample sizes calculated using absolute precision are considerably lower than those using relative precision, it still may be difficult for researchers to obtain even 100 independent observations (depending on how one achieves that independence; e.g., single point observations) for a population. The estimation method presented here allows researchers to assign a precision, *a posteriori,* to any sample of independent observations, thereby getting an idea of the "power" of their sample and attaining a certain level of confidence in their data.

Suggested Sampling Protocol

To estimate sample size using techniques described above, one must have some *a priori* idea of the number of categories (k) and the proportions of observations that will be found in each category. Because those proportions usually are not known, one may consider using the "worst case" (e.g., using $\Pi_i = 0.50$ in the equation for absolute precision above) sample size in order to ensure an adequate sample. While this approach is justifiable, it could lead to gross oversampling. One might also rely on published data to gain an idea of the proportions for a particular species, provided that those data are accurate representations of behaviors exhibited by the species. However, many species exhibit highly plastic foraging behaviors (Petit, Petit, and Petit, this volume), such that predicting foraging behaviors for one population based on previous studies conducted at other locations, or even at the same location using different methods or observers, may be tenuous.

A more reasonable approach would be to collect a preliminary sample of observations (say $N = 100$; these would not necessarily have to be independent observations) to estimate the proportions $\Pi_1, \ldots, \Pi_k$. For each estimate of Π_i, decide the acceptable absolute precision, b_i, and confidence (α) levels (see above) for Π_i and calculate the estimated sample size (n_a) using the formula above, realizing that it will be necessary to then collect n_i − N additional observations (if N is made up of independent observations). As for the formula above, if $b_i = b$ for all categories, the Π_i closest to 0.50 should be used. Because the required sample size will increase with an increase in number of categories within a variable, researchers perhaps should calculate a required sample size based on the minimum n_i required for the variable with the most k categories.

ACKNOWLEDGMENTS

We thank W. M. Block, J. E. Dunn, M. L. Morrison, and especially C. A. O'Cinneide and J. T. Rotenberry for comments which greatly improved the quality of this manuscript. K. G. Smith was supported by National Science Foundation Grant BSR 84-08090 during preparation of this paper. This work is dedicated to the memory of Richard E. Petit.