J. Field Ornithol., 62(3):357-366

ESTIMATING SEX RATIOS WITH DISCRIMINANT FUNCTION ANALYSIS: THE INFLUENCE OF PROBABILITY CUTPOINTS AND SAMPLE SIZE

LEONARD A. BRENNAN
Department of Wildlife and Fisheries
P.O. Drawer LW
Mississippi State University
Mississippi State, Mississippi 39762 USA

JOSEPH B. BUCHANAN
Cascadia Research Collective
218½ West Fourth Avenue
Olympia, Washington 98501 USA

CHARLES T. SCHICK
Department of Biological Sciences
University of California
Santa Barbara, California 93106 USA

STEVEN G. HERMAN
The Evergreen State College
Olympia, Washington 98505 USA

Abstract.—The influence of different probability cutpoints and associated variations in sample size on population sex ratios were inferred with birds sexed by discriminant function analysis (DFA). Using two data sets from Dunlins (*Calidris alpina*) collected in western Washington state as examples, it was observed that selection of high probability (≥0.95) cutpoints resulted in biased population sex ratio estimates. Overall, approximately 80% of the samples from the two data sets (in the probability cutpoint range of ≥0.7) used in the analyses were required for accurate estimation of sex ratios. Calculation of sample sizes required for a 5% error bound on the estimate of the proportion of females in the study population indicated that 200-300 samples were required. This figure will probably vary for other species of birds with different sex ratios and population sizes. A technique from the statistical literature for assessing the statistical limitations of sex ratios based on data sets with fixed sample sizes is presented. DFA has great potential as a passive, physically nondestructive technique for sexing species of monochromatic birds, but sample size must be considered when population parameters such as sex ratio are inferred using birds sexed by DFA.

ESTIMANDO LA PROPORCIÓN DE SEXOS MEDIANTE EL USO DE UN ANÁLISIS DE FUNCIÓN DISCRIMINANTE: LA INFLUENCIA DE PUNTOS DIVISORIOS DE PROBABILIDADES Y TAMAÑO DE LA MUESTRA

Sinopsis.—La influencia de diferentes puntos divisorios de probabilidades y las variaciones asociadas al tamaño de una muestra en la proporción de sexos en una población, fueron inferidos con aves cuyo sexo fue determinado mediante el uso de un análisis de función discriminante. Utilizando dos conjuntos de datos de individuos de *Calidris alpina,* tomados en la parte oeste de Washington, se observó que la selección de puntos divisorios de alta probabilidad ≥0.95, resultaron en estimados con sesgo de la proporción de sexos. En general, se requieren aproximadamente el 80% de las muestras en los dos conjuntos de datos (en donde el campo de valores de los puntos divisorios de probabilidades sean ≥0.7) para estimar con precisión la proporción de sexos. El cálculo de los tamaños requeridos de las muestras para tener un margen de error de 5%, en el estimado de la proporción de hembras, indicó que se necesitaban al menos de 200–300 individuos. Estos números probablemente varíen para otras especies de aves con diferentes proporciones de sexos y tamaños poblacionales. En este trabajo se presenta una técnica de la literatura sobre estadísticas, para determinar las limitaciones de ésta en la proporción de sexos, basados en conjuntos de datos con una muestra fija. El análisis de función discriminante tiene gran potencial como técnica pasiva (en donde no se destruya la muestra), para determinar el sexo de aves monocromáticas. Sin embargo, el tamaño de las muestras debe ser considerado cuando parámetros poblacionales, como la proporción de sexos, son inferidos utilizando aves cuyo sexo es determinado por análisis de función discriminante.

There has been a recent surge of interest in the use of discriminant function analysis (DFA) as a tool for sexing species of birds that are sexually monochromatic (Brennan et al. 1984, Edwards and Kochert 1986, Green 1982, Hanners and Patton 1985, Maron and Myers 1984, Reese and Kadlec 1982, Scolaro et al. 1982, Skeel 1982). Typically, one develops and tests a discriminant function (predictive equation) using external morphometric measurements (e.g., bill or tarsus lengths) taken from birds of known sex, such as museum specimens, or collections made for other purposes. After being tested for accuracy with known samples, the discriminant function is used to classify live birds to a particular sex. Thus, such a technique can be applied readily to measurements from birds that are trapped or netted as part of a banding program. The primary appeal of DFA as a sexing tool is that monochromatic species can be sexed at a preselected level of statistical confidence without relying on techniques that are physically invasive or destructive to birds.

When DFA is used as a sexing procedure, morphometric data for an individual bird are used to calculate a specific discriminant score for each bird. Associated with each DFA score is a probability coefficient (ranging between 0.5 and 1.0) that corresponds to the likelihood that the bird has been correctly sexed. If the objective of an investigation is to obtain behavioral observations of known sex birds, then one should choose a high probability (generally ≥ 0.90) of correct classification, or use a technique such as laparotomy (see Maron and Myers 1984) to determine sex. If, however, the objective of the research is to infer a population parameter such as a sex ratio, then one must consider the size of the sample on which the inferred sex ratio is based. This is because accurate estimation of sex ratios often requires large sample sizes and because there is a direct, inverse relationship between sample size (based on the proportion of the total sample) and the probability of correctly classifying birds by sex (Fig. 1). Thus, a high probability of correct classification results in a concomitant decrease in the sample size on which the inferred sex ratios would be based.

To our knowledge, no one has examined the potentially confounding influence that variation in selection of probability cutpoints (and subsequently sample size) might have on population sex ratios when they are inferred with birds sexed by DFA. Therefore, the purpose of this paper is to illustrate how sample size can affect estimates of population sex

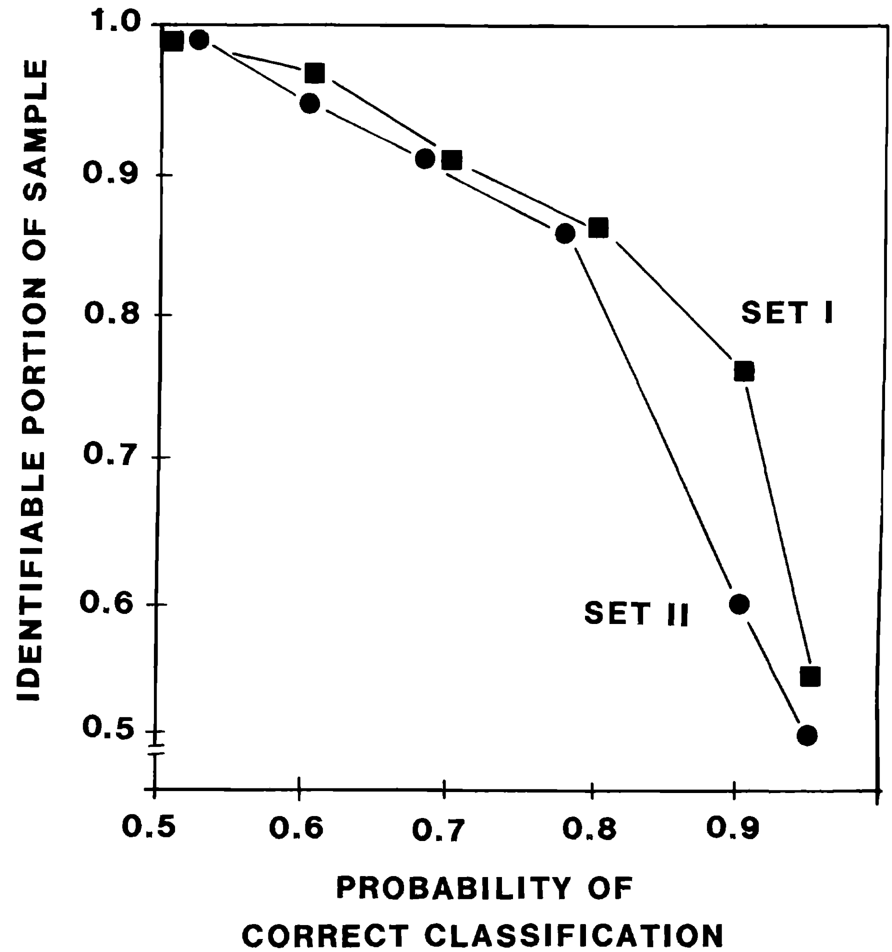

FIGURE 1. Relationship between proportions of samples and increases in the probability of correctly sexing Dunlins from western Washington using discriminant function analysis. Sample size decreases as the probability of correctly sexing the birds with DFA increases. Squares represent data from Set I (n = 200), circles represent data from Set II (n = 78). Actual number of samples in each probability cutpoint interval is given in Table 1.

ratios when they are inferred from birds sexed by DFA. We use data from our research on Dunlins (*Calidris alpina*) as an example. Our objectives are to: (1) assess the tradeoffs between choosing different probability cutpoints and obtaining a realistic population sex ratio from two data sets of birds sexed by DFA; (2) illustrate a method for calculating the sample size necessary to be statistically confident that a sex ratio inferred from birds sexed by DFA reflects the actual sex ratio of a population; and (3) illustrate a method for determining the statistical limitations of sex ratios inferred from data sets with fixed sample sizes.

METHODS

DFA is a multivariate statistical method that uses a linear combination of variables to maximize the variation, and hence separation, between two or more groups. The discriminant function is then used in a separate analysis to classify samples into their groups (see Green 1978 for conceptual background and computational formulae of DFA). Before being used to classify unknown samples, the accuracy of a particular discriminant function is tested with samples of known group membership (Green 1978).

Estimation of sex ratios: proportions of males and females.—We used two data sets (referred to below as Set I and Set II) from our research on Dunlins (Brennan et al. 1984, Buchanan et al. 1986) to evaluate the influence of probability cutpoints on the accuracy of inferred population sex ratios. Set I was initially used to develop a discriminant function (see Brennan et al. 1984 for details) that predicted the sex of 183 of 200 (91.5%) Dunlins based on measurements of exposed culmen, wing-length and weight. Set II was obtained from data presented in Buchanan et al. (1986), and consists of 78 additional Dunlins of known sex. The discriminant function based on the data from Set I correctly classified 69 of the 78 (89%) Dunlins in Set II. Thus, approximately 10% of both data sets contained birds that were improperly sexed by DFA. These incorrectly sexed birds were included in our subsample sex ratio estimates for two reasons: (1) they were distributed evenly throughout all probability cutpoint intervals (e.g., each cutpoint interval contained only one or two improperly sexed birds), and (2) the inclusion of improperly sexed birds had no effect on the inferred sex ratios.

We treated the data in Sets I and II as separate, hypothetical "populations." Both of these "populations" had male : female sex ratios of 1.6:1 (62% males, 38% females), which is the same as the overall sex ratio of Dunlins wintering in western Washington based on a total of 588 birds (Buchanan et al. 1986). In this example, we considered sex ratios from the standpoint of the proportion of male (0.62) and female (0.38) Dunlins in our study populations.

In the analyses presented here, we categorized subsamples from Sets I and II according to a set of probability cutpoint intervals (Table 1). We then calculated estimates of sex ratios by successively eliminating the subsamples within each probability cutpoint interval, starting with all data having a probability ≥ 0.51 of being correctly sexed, and continuing until only data from the 0.96–1.0 interval were included. Each sex ratio estimate was then graphically contrasted with the actual sex ratio of both hypothetical populations.

Estimation of required sample sizes.—The following equation from Scheaffer et al. (1986:59; see also Cochran 1977 for sampling concepts involving ratios and proportions) was used to calculate the sample size

TABLE 1. Breakdown of sample sizes and percentages of total samples in relation to probability cutpoints for correctly classifying male and female Dunlins with discriminant function analysis.

Probability cutpoint	0.51–0.60	0.61–0.70	0.71–0.80	0.81–0.90	0.91–0.94	0.95–1.00
Set I						
Percent of sample	100	97	90	85	77	55
Sample size	200	194	180	170	154	110
Set II						
Percent of sample	100	95	91	85	61	50
Sample size	78	74	71	66	48	39

(n) needed to estimate the proportion of females in the total population (N) within a 5% error bound:

$$n = \frac{N(pq)}{[(N - 1)D] + (pq)}, \tag{1}$$

where p = the proportion of females in the population (in our case p = 0.38), $q = 1 - p$, and $D = B^2/4$ (here we set $B = 0.05$, for a 5% error bound on the estimation of p). For values of N, we used the range of population sizes documented for Dunlins in western Washington (250–15,000; Brennan et al. (1985)) and elsewhere in the world (up to 200,000; Cramp and Simmons (1983)). We used the range of regional values to illustrate the relationship between the sizes of populations and required samples for sex ratio estimation. Sampling large populations would also require careful consideration of spatial variation in sex ratios. In our example, consider a population of 2000 birds where the proportion of females is 0.38. Calculations using equation (1) indicate that 317 birds would be required to be 95% confident that the observed proportion would be within 0.05 of this value. That is, 95% of the time a sample of 317 birds would give an estimate of 0.33–0.43 for the proportion of females in the population.

Estimation of error bounds.—The number of birds collected or trapped from a particular location is often fixed (i.e., it is not possible to collect additional samples). It is possible, however, to estimate the error bound for an estimated sex ratio obtained from a fixed number of samples. We used the following equation (also from Scheaffer et al. 1986:57) to estimate the error bound on the estimate of the proportion of females:

$$E = \pm 2\sqrt{\left(\frac{pq}{n - 1}\left[\frac{N - n}{N}\right]\right)}, \tag{2}$$

where E = the error bound and remaining symbols follow equation (1) above. In equation (2), we considered N (total population size) to represent a population of 2000 birds, and n (the sample taken from the population) to represent different fixed samples ranging from 10–500 birds sampled from the overall population of 2000. Thus, if only 100 birds were sampled from the overall population of 2000 (rather than the 317 required for a 5% error bound as indicated by equation (1) where the proportion of females is 0.38, equation (2) indicates that the error bound on the proportion of females obtained from a sample of 100 birds would be 9%.

RESULTS

Estimating proportions of males and females with DFA.—Higher probability cutpoints had a drastic effect on the estimated sex ratio and hence the proportions of males and females within each remaining subsample (Fig. 2). Proportions of males and females inferred from Set I approximated the actual proportions of both sexes at and below the probability cutpoint of 0.9 (Fig. 2). Proportions of males and females inferred from Set II approximated the actual sex ratio at probability cutpoint levels of 0.7 and 0.5, but deviated greatly at other cutpoints. Small sample sizes resulting from high (>0.9, see Table 1) cutpoints were apparently inadequate for obtaining a realistic sex ratio estimate. With the exception of a wide deviation from the actual sex ratio at the 0.6 probability cutpoint with Set II, the estimates that were closest to the actual population sex ratios were obtained when ≥80% of each data set was used. The directions of the deviations from actual sex ratios were quite different between Sets I and II (Fig. 2). With Set I, the sex ratio at the >0.95 probability cutpoint interval was much greater than the actual sex ratio. With Set II, estimated sex ratios were both greater and lesser than the actual sex ratio.

Estimation of sample sizes and error bounds.—Using a set of population size values ranging from 200 to 200,000 individuals with equation (1), we observed that the required n for estimation of sex ratios with a 5% error bound ranged from 131 for a population of 200 birds to 377 for a population of 200,000 birds (Fig. 3). When we considered the effect of fixed sample sizes on error bounds (using sample sizes ranging from 10 to 500 for a population of 2000 birds) using equation (2) we observed that error bounds ranged from 37 to 4% for samples of 10 to 500 birds. Samples ≥100 had error bounds of <10%.

DISCUSSION

Our analyses indicated that the effect of sample size must be considered when inferring population sex ratios using birds sexed by DFA. Sex ratios from Sets I and II deviated from actual sex ratios in different, and apparently unexplainable directions (Fig. 2), especially when a small number of samples were used. Choosing a high probability of correctly sexing individuals with DFA caused a large proportion of birds to be excluded from the analysis, and thus reduced the sample size on which the sex ratio was ultimately based. Maron and Myers (1984; see their Fig. 3) also observed a similar pattern of decreasing sample sizes associated

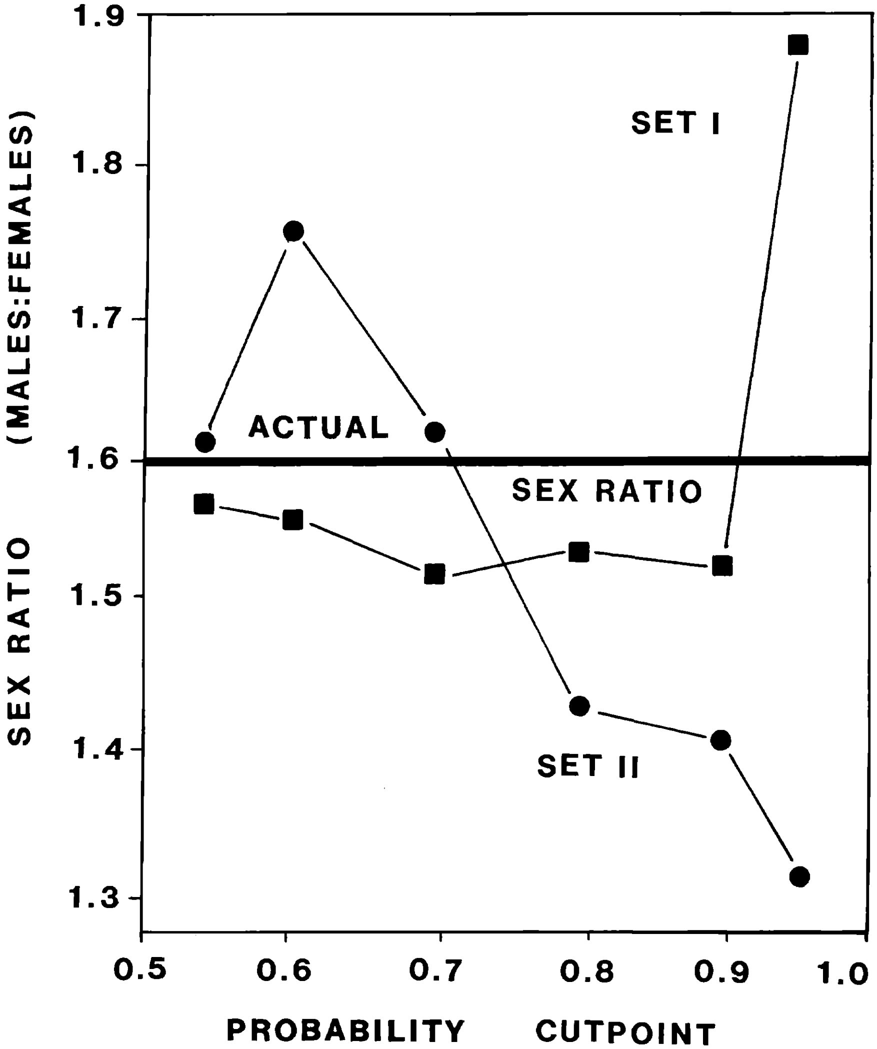

FIGURE 2. Deviations of inferred from actual sex ratios in two samples of Dunlins from western Washington. Percentages and counts of the total number of samples associated with each probability cutpoint interval are given in Table 1. Squares represent data from Set I ($n = 200$); circles represent data from Set II ($n = 78$). Actual sex ratio of both data sets is indicated by the line at 1.6:1 (38% females, 62% males).

with an increase in the probability of correctly classifying Sanderlings (*C. alba*) by sex with DFA. We thus urge investigators to test the generality of this pattern with other species of monochromatic birds.

With our data, it appeared that at least 80% of the individuals from each set were required for a realistic assessment of the sex ratio. Lowering

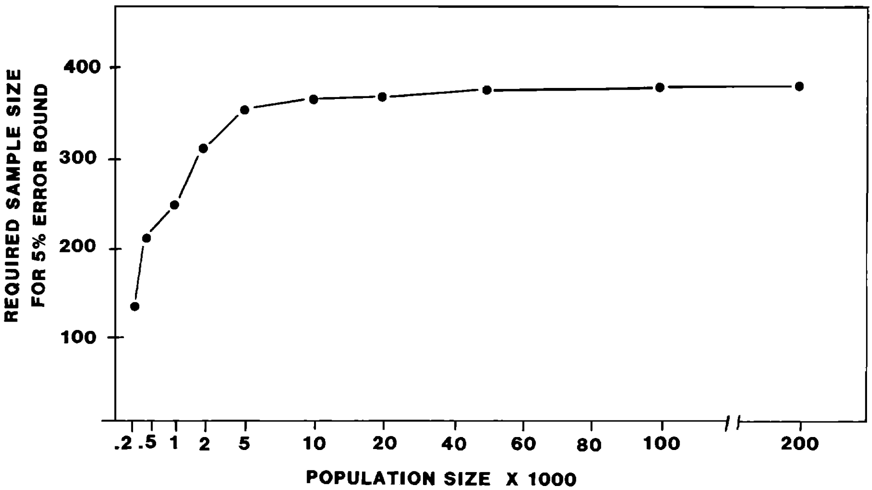

FIGURE 3. Number of samples required to estimate the sex ratio of populations of Dunlins with 5% error bound when the ratio of males to females is 1.6:1 (38% females and 62% males) and population size ranges from 250 to 200,000 birds, based on equation (1) in text.

the probability cutpoint to 0.70–0.80 provided a much more realistic sex ratio estimate than the estimates obtained from birds in the cutpoint intervals >0.9. This pattern occurred despite the fact that the lower probability cutpoints presumably allow more error in the classification of birds by sex with DFA. However, if a large number of samples are available, then it might be possible to set a high (≥0.9) cutpoint and still obtain a realistic sex ratio estimate.

In Dunlins from western Washington, approximately 300 birds were required to estimate the proportion of females with a 5% error bound for a population of 2000 birds. These results also point to the need for considering the size of the population from which samples will be taken for estimating a sex ratio (see Fig. 3). Required sample size estimates for these parameters will probably vary according to the particular species and/or population being studied because between-sex differences in morphology and population sex ratios vary widely across different species of birds (e.g., Desrochers 1990). In a strict sense, our analyses were based on reasoning that was, in part, circular because we had prior information about the sex ratios of Dunlin populations in western Washington. Nevertheless, this approach was warranted because our goal was to assess required sample sizes for populations of different sizes when proportions of males and females were known. In cases where prior information on proportions of males and females in a population is not available, Sheaffer et al. (1986) advise setting $p = q = 0.5$ to give an overestimate of sample size. The initial sample size obtained from equation 1 with $p = q = 0.5$

could then be used to obtain a statistically reliable estimate of the proportions of males and females in a population.

Sex determinations based on DFA have been developed for a wide variety of bird taxa (e.g., eagles: Edwards and Kochert 1986, shorebirds: Brennan et al. 1984, Maron and Myers 1984, Skeel 1982, gulls: Hanners and Patton 1985, penguins: Scolaro et al. 1982, and corvids: Reese and Kadlec 1982, Green 1982). Therefore, it should be possible to apply the general methodology presented here to most species of monochromatic birds that can be sexed with DFA. The equations we used to determine the sample size required to estimate the proportions of males and females in a population with a preselected level of statistical confidence, and to determine the statistical limitations of sex ratios based on a fixed sample size, can be applied to any animal species. Although it appears that DFA has great potential as a passive, non-invasive, tool for sexing monochromatic birds, it is also apparent that sample size must be considered when population parameters are inferred using birds sexed by this technique.

ACKNOWLEDGMENTS

E. H. Burtt, Jr., W. M. Block, M. A. Colwell, R. J. Cooper, T. C. Edwards Jr., F. Hagan, H. A. Jacobson, R. M. Kaminski, B. D. Leopold, E. P. Mallory, L. R. Mewaldt, M. L. Morrison and two anonymous referees commented on earlier versions of this paper and provided many suggestions that improved it. This research was based on studies supported in part by NSF Grant SOS-SPI-8004760, The Evergreen State College, The College of Natural Resources and Department of Wildlife, Humboldt State University (LAB 1982–1984) and the Department of Forestry and Resource Management, University of California, Berkeley (LAB 1986–1989). LAB owes a special debt of thanks to Teresa Pruden for support, and to Dr. Gerald M. Allen for inspiring the initial discriminant analysis and providing insight into the use of DFA for inferring sex ratios. This is publication number J-7544 from the Mississippi Agricultural and Forestry Experiment Station.

LITERATURE CITED

BRENNAN, L. A., J. B. BUCHANAN, C. T. SCHICK, S. G. HERMAN, AND T. M. JOHNSON. 1984. Sex determination of Dunlins in winter plumage. J. Field Ornithol. 55:343–348.

———, ———, S. G. HERMAN, AND T. M. JOHNSON. 1985. Interhabitat movements of wintering Dunlins in western Washington. Murrelet 66:11–16.

BUCHANAN, J. B., L. A. BRENNAN, C. T. SCHICK, S. G. HERMAN, AND T. M. JOHNSON. 1986. Age and sex composition of wintering Dunlin populations in western Washington. Wader Study Group Bull. 46:37–41.

COCHRAN, W. G. 1977. Sampling techniques, 3rd edition. John Wiley, New York, New York. 428 pp.

CRAMP, S., AND K. E. L. SIMMONS (EDS.) 1983. Handbook of the birds of Middle Europe, the Middle East, and North Africa. Oxford Univ. Press, London, United Kingdom. 913 pp.

DESROCHERS, A. 1990. Sex determination of Black-capped Chickadees with a discriminant analysis. J. Field Ornithol. 61:79–84.

EDWARDS, T. C., JR., AND M. N. KOCHERT. 1986. Use of body weight and length of footpads as predictors of sex in Golden Eagles. J. Field Ornithol. 57:317–319.

GREEN, P. E. 1978. Analyzing multivariate data. Dryden Press, Hinsdale, Illinois. 519 pp.

GREEN, P. T. 1982. Sexing Rooks (*Corvus frugilegus*) by discriminant analysis. Ibis 124: 320–324.

HANNERS, L. H., AND S. R. PATTON. 1985. Sexing Laughing Gulls using external measurements and discriminant analysis. J. Field Ornithol. 56:158–164.

MARON, J. L., AND J. P. MYERS. 1984. A description and evaluation of two techniques for sexing wintering sanderlings. J. Field Ornithol. 55:336–342.

REESE, K. P., AND J. A. KADLEC. 1982. Determining the sex of Black-billed Magpies by external measurements. J. Field Ornithol. 53:417–418.

SCHEAFFER, R. L., W. MENDENHALL, AND L. OTT. 1986. Elementary survey sampling, 3rd edition. Duxbury Press, Boston, Massachusetts. 324 pp.

SCOLARO, J. A., M. A. HALL, AND I. M. XIMINÉZ. 1982. The Magellanic Penguin (*Spheniscus magellanicus*): sexing adults by discriminant analysis of morphometric characters. Auk 100:221–224.

SKEEL, M. A. 1982. Sex determination of adult Whimbrels. J. Field Ornithol. 53:414–416.

Received 21 Mar. 1990; accepted 10 Jan. 1991.

The COOPER ORNITHOLOGICAL SOCIETY will hold its 62nd Annual Meeting at the University of Washington in Seattle, June 22–26, 1992. The meeting will include a variety of events in addition to the scientific program, including a wildlife art exhibition, numerous social events, local excursions to the surrounding Seattle area, outdoor barbecue, pelagic birding, birding trips to Mt. Rainier and North Cascades National Parks, Mount St. Helens National Volcanic Monument, San Juan Islands, and the eastern slope of the Cascades. Dr. David A. Manuwal (Wildlife Science Group, College of Forest Resources, Univ. of Washington, Seattle, WA 98195) will be Local Arrangements Chair. Dr. Dennis Martin (Biology Department, Pacific Lutheran University, Tacoma, WA 98447) will be the Scientific Program Chair.