MICHENER, H. and J. R. MICHENER. 1931. Variation in color of male House Finches. *Condor,* **33:** 12-19.

—— 1932. Colors induced in male House Finches by repeated feather removal. *Condor,* **34**: 253-256.

—— 1940. The molt of House Finches in the Pasadena Region, California. *Condor,* **42:** 140-153.

MOORE, R. T. 1939. A review of the House Finches of the subgenus Burrica. *Condor,* **41:** 177-205.

THOMPSON, W. I. 1960. Agonistic behavior of the House Finch, Part I: Annual cycle and display patterns. *Condor,* **62:** 245-271.

TORDOFF, H. B. 1954. A systematic study of the avian family Fringillidae based on the structure of the skull. Mus. of Zoology, Univ. of Mich., Mis. Publ. No. 81, 41 pp.

VAN ROSSEM, A. J. 1936. Birds of the Charleston Mountains, Nevada. *Pac. Coast Avif.* No. **24:** 5-65.

Received February, 1964

NOTES ON QUANTITATIVE TREATMENTS OF SUBCUTANEOUS LIPID DATA

By JACK P. HAILMAN

The observation of subcutaneous lipid deposits on migratory passerines is becoming a standard part of banding procedure. That such data are useful to students of migratory physiology is evident (e.g., see Helms and Drury, 1960 and references therein). However, the quantitative treatment of such data has been variable, and, so far as I have been able to find, invariably inadequate or inappropriate. My present purpose is to draw attention to some statistical techniques that may prove useful for analyzing these fat data.

DETERMINING THE NATURE OF THE DATA

Before applying any statistical manipulation, the bander must answer certain questions about the nature of his data, for the answers to these questions govern his choice of both descriptive statistics (e.g., central tendency and variation expressions) and "comparative" statistics (e.g., tests of differences and of correlation). The two most important questions for banding data, including weights and measurements, as well as lipid observations, seem to concern the type of measurement and the sampling distribution. When a specific test of significance is chosen, one must also consider the kind of operation to be performed (including independence or relational properties in the data), and the size of the sample. The ease of calculation may also rightly influence the choice of tests. Other problems, such as power-efficiency (including balancing Type I and II errors and stating significance levels) are beyond the scope of these notes.

(1) *Kind of measurement.* Generally, three levels of measurement are distinguished. Siegel's (1956) nominal, ordinal and interval measurements are here rephrased as:

(*a*) *Classificatory* measurement is a dichotomous scheme that assigns observations to one of two categories, for instance, "has body fat" or "lacks it." Few banding data need be so crude.

(*b*) *Ranking* measurement is a scheme that assigns each observation to a place on a single scale, such that all observations assigned "lower" places on the scale are further toward one extreme and all observations assigned "higher" places are further toward the other extreme of variability. Fat observations are ranked in this way (see below).

(*c*) *Interval* measurement also "ranks" data, but ranks them on a scale graduated in precisely stated, mensural intervals, usually of equal magnitude (I include here *ratio* measurements, which may be considered as interval scales with a true zero.) Most weights and measurements are of this nature. Thus weights of 5.0 and 4.0 grams differ by the same value as weights of 4.0 and 3.0 grams. (However, fat observations of classes 5 and 4 may not differ by the same amount as do fat observations of classes 4 and 3, even though $5 > 4 > 3$.)

(2) *Sampling distribution.* Many weights and measurements of animals have been found empirically to vary according to the "normal" or Gaussian (bell-shaped) curve. The assumption of normality underlies most descriptive and comparative statistics dealing with banding data. Other distributions undoubtedly occur, but, so far as I am aware, have not been shown in data derived from manipulation of the banded bird. (Of course, retrap-data and similar results of banding have been shown to be non-normal.) It is often the case that the experimenter does not know anything of the underlying pattern of variation, and therefore wishes to avoid assumptions about "parameters." Such a situation calls for nonparametric statistical methods.

(3) *Kind of operation.* Most "comparative" statistics are designed to describe how two sets of measurements vary together (correlation, regression, contingency) or to test whether some observed difference is due merely to chance variation. Only the latter is of immediate interest here. The difference being tested may be between the observed sample of data and some universal hypothesis (such as "equal numbers of male and female"). Alternatively, the difference may be between two samples of data gathered ("do males weigh more than females?"). The data in such comparisons may be independent (e.g., weights of Massachusetts birds versus weights of birds in North Carolina), or may be related (e.g., the weights of individual birds at banding and recovery on migration).

THE NATURE OF LIPID OBSERVATIONS

Similar schemes for ranking subcutaneous lipid deposits on migratory passerine birds have been offered by McCabe (1943), used also by Connell, Odum and Kale (1960); by Wolfson (1954), used by Lawrence (1958) and the Patuxent Bird-Banding Office (1960) for "Operation Recovery;" by Graber and Graber (1962); and by Helms (1959), revised by Helms and Drury (1960) and used by Helms (1963). The schemes vary in the number of distinguishable steps (ranks), this number being a function both of the observer and the species studied. However, these systems have in common easily-observed steps of increasing amounts of fat, which however, cannot be considered equal intervals. That the last is true is indicated by Baird's data in Connell *et al* (1960), and by the detailed comparisons of weights by fat class in a number of emberizine species studied by Helms (1963; Helms and Drury, 1960).

It follows from the demonstration that fat data are ranked measurements, that the sampling distribution underlying the variation is unlikely to be a well-known one. These two facts suggest immediately that appropriate statistics for the treatment of fat data are nonparametric methods and descriptions that assume a level of measurement no higher than ranking (ordinal).

DESCRIPTIVE STATISTICS FOR LIPID DATA

Central Tendency

Some observers (e.g., Lawrence, 1960; Nisbet *et al*, 1963), perhaps recognizing the difficulties in describing central tendencies, have side-stepped the question and used no measure; others (e.g., Graber and Graber, 1962; Helms, 1959, 1963; Helms and Drury, 1960), perhaps with equal recognition of the difficulties, have nonetheless calculated inappropriate means in preference to leaving the data unanalyzed. The *mean* (or numerical average) is found by summing all values, and dividing by the sample size; its accuracy as a measure of central tendency in data is strictly contingent upon the assumption of equal intervals of measurement, or with an appropriate correction for inequality of the intervals. This assumption is not met by lipid data.

The correct measure of central tendency for rank-ordered data is the *median*. The median is the value (e.g. fat class) of the middle observation. Thus the median of a group of three observations, one each in fat classes 1, 3 and 4, is fat class 3. The median of an even number of observations is the mid-point between the two middle observations; for example, with one observation each in fat classes 1, 3, 4 and 5, the median is 3.5.

The *mode*, the fat class containing the largest number of observations in the sample, is another, less useful (for ranked measurements) measure of central tendency. It is usually almost accidental for the mode to be a good measure of central tendency. It is requisite that the real distribution (not merely the theoretical) shall be very nearly symmetrical.

Variation

Most observers have avoided stating any measure of variation for fat data; Graber and Graber's (1962) use of the "t" test implies that standard deviations were used to measure variation of their data. The same violation of assumptions about the level of measurement, as well as the additional burden of normality assumptions, render standard deviations, like means, inappropriate.

One alternative is the use of frequency distributions, as shown by Lawrence's (1958) graphs. Another, perhaps better, method is the use of percentiles (or quartiles); for these the reader is referred to standard statistics texts such as Snedecor (1956). My own inclination would be to avoid variation measures of fat altogether, since variation per se is rarely the object of study. In parametric data statements on variation are useful in making rapid tests of significance, especially by well-known graphic methods, but tests of nonparametric data cannot be so readily made from percentiles.

TESTS OF SIGNIFICANCE FOR LIPID DATA

Independent Data

Graber and Graber (1960) used Student's 't" test, which is based upon assumptions of normal distribution of interval measurements. Again, this is clearly inappropriate. Helms (1958) avoided using any method, but was forced to conclude that the difference between two fat samples "may be regarded as significant" without the aid of a probability statement. This situation may be rectified by use of either the Median Test or the Mann-Whitney U Test (e.g., Siegel, 1956).

The Median test specifically distinguishes between two independent samples that differ in median, the usual kind of information required from lipid data. The more "powerful" Mann-Whitney U Test is certainly the best alternative to the parametric "t" test whenever equal interval measurement or normal distribution requirements cannot be met. There is available in the Mann-Whitney Test a correction for ties (more than one observation per class) which, when used, makes this test nearly the equal of the "t" test for large samples. Since lipid data will almost always involve ties, such a provision is very helpful. However, my recommendation is to compute the test without bothering with the tie-correction, unless the difference being tested is small. Used without correction, the test is "conservative," giving a higher probability of chance difference than would be found with the correction. Thus, if the difference is significant without the correction, it will be (more) significant with the correction.

Related Data

I cannot find that anyone has tested the difference between related sets of lipid data. The appropriate tests for doing so are the Sign Test and the Wilcoxon matched-pairs signed-ranks test, of which the latter is the more powerful. Both tests assume that the underlying variable (e.g., fat) is continuous, but that the observer can only measure certain classes along this continuum—this assumption is certainly met by lipid data. In addition, the latter test requires ranked (ordinal) differences between the pairs of data studied. In other words, the difference between "fat 1" of bird A and "fat 3" of bird B must be of the same order as the difference of "fat 1" of bird A at time (1) and the "fat 3" of bird A at time (2). This assumption is also met by lipid data (providing the bander does not alter his criteria for judgment of fat classes with time).

AN EXAMPLE

The study by Helms (1958) of fat on Song Sparrows (*Melospiza melodia*) on two successive days provides an excellent example. Dr. Helms has been kind enough to supply me with the original data from that study for the following analysis. The hypothesis to be tested was that birds banded prior to a night of migration (5 April 1958) had more fat than birds caught the next day. Weights from these two independent samples (actually, two individual birds were caught on both days) proved significantly different by ordinary parametric methods. However, it is important to the hypothesis of "accumulation of fat prior to a migratory flight" that fat measurements also be compared directly.

Helm's fat data, utilizing a scale of 12 classes, are shown in Table I. The *mode* for 5 April is fat 2, that for 6 April, fat 1, although it is evident that thus taking merely the most popular value hardly describes the central tendency accurately. Helms reports the *mean*

TABLE I. ORIGINAL DATA FROM STUDY OF HELMS (1959)

Fat class	Number of individuals found on: 5 April	6 April
0	0	3
1−	0	1
1	0	11
1+	2	6
2−	0	1
2	10	9
2+	9	2
3−	7	2
3	6	1
3+	4	0
4−	1	0
4	2	0

fat to be 2.35 and 1.40 for 5 and 6 April, respectively. My calculations for the more appropriate *medians* give very similar values of 2+ and 1+. (The sample sizes here used are smaller than those used for weights in the original paper because many birds were not classed for fat, even though they were weighed.)

I have tested the difference between these medians by both the Median and the Mann-Whitney U tests. Both tests yield a very small probability ($p < .001$) that the observed difference is due to chance alone. Furthermore, calculation of the latter test with and without corrections for ties yields the same result. The data have been provided so that the reader may explore these methods himself (see Siegel, 1956 or any text on nonparametric statistics for the method of calculation). Note that the statistical decision (significant difference) accords well with Helms's intuitive decision.

OTHER COMMENTS

There are a wealth of other tests which may be used in testing significance. Some of these are put forth in Siegel's (1956) readable text, while others appear only in more technical books or in the literature. For a nonparametric analysis-of-variance situation, I have found Wilson's (1956) method to be handy, although perhaps not very powerful. For correlations, I have utilized Kendall's rank correlation coefficient (see Siegel, 1956) with behavioral data. It should prove useful for lipid data as well, since it can further be generalized to a partial correlation, where the association of fat with, say, weight might depend on a third variable which is held constant.

The nonparametric methods that I have nominated as being useful for treatment of lipid measurements can likewise be applied to other banding data of similar sorts. Thus Blake's (1957) scale of gape color between red and yellow in Purple Finches (*Carpodacus purpureus*) might be treated in the same way. Some authors (e.g., Nisbet *et al*, 1963) have made remarks which suggest that the boundaries between fat classes are somewhat subjective, such that two birds with identical fat might be assigned different ranks by the same observer. The very large problem of interobserver differences also exists in such data. It would seem most reasonable in this case to demand a rather high level of significance for differences (say .01 in lieu of the usual .05) to counteract this effect. However, such problems lead quickly to exploration of power-efficiencies and balancing of errors, subjects outside the present aim of this communication.

Finally, it should be noted that all these methods can be used on the more traditional kinds of banding data (weights, wing-lengths, etc.). The methods only require *at least* ranking measurement, and no assumptions about parameters. Such an extension of nonparametric methods has at least three advantages. (1) Some traditional data may not, in fact be normally distributed; they may be log-normal, bimodal (peaks for males and females, or by age), or even more complex. Thus, usual parametric methods are somewhat inappropriate. (2) In general, nonparametric methods are much easier to calculate, yet in most cases yield identical information. Finally (3), because they are easier to calculate, fewer mistakes are likely. Some statisticians believe that scientists may more often make the correct statistical decision by using easy, less powerful methods than by utilizing the most powerful tests and thereby introducing damaging errors in calculation.

SUMMARY

Ranking schemes for subcutaneous lipid deposits on migrating birds demand nonparametric statistical analysis. The appropriate measure of central tendency is the *median* (not mean), and of variation is the *percentile* (not the standard deviation). The appropriate test of significance for two sets of independent data is the *Mann-Whitney U Test* (not Student's "t" test), and for related data is the *Wilcoxon Matched-pairs Signed-ranks Test*; other nonparametric methods are also appropriate, but not as powerful. Similar treatment of other banding data is discussed.

BIBLIOGRAPHY

BLAKE, C. H. 1957. Notes on juvenal Purple Finches. *Bird-Banding* **28**: 29-40.

CONNELL, C. E., E. P. ODUM and H. KALE. 1960. Fat-free weights of birds. *Auk* **77**: 1-9.

GRABER, R. R. and J. W. GRABER. 1962. Weight characteristics of birds killed in nocturnal migration. *Wilson Bull.* **74**: 74-88.

HELMS, C. W. 1959. Song and Tree Sparrow weight and fat before and after a night of migration. *Wilson Bull.* **71**: 244-253.

HELMS, C. W. 1963. Tentative field estimates of metabolism in buntings. *Auk* **80**: 318-334.

HELMS, C. W. and W. H. DRURY, JR. 1960. Winter and migratory weight and fat field studies on some North American buntings. *Bird-Banding* **31**: 1-40.

LAWRENCE, L. DE K. 1958. On regional movements and body weight of Black-capped Chickadees in winter. *Auk* **75**: 415-443.

McCABE, T. T. 1943. An aspect of collectors' technique. *Auk* **60**: 550-558.

NISBET, I. C. T., W. H. DRURY, JR. and J. BAIRD. 1963. Weight-loss during migration. *Bird-Banding* **34**: 107-159.

ODUM, E. P. 1960. Premigratory hyperphagia in birds. *Amer. J. Clinical Nutrition* **8**: 621-629.

PATUXENT BIRD-BANDING OFFICE. 1960. Memo: Operation Recovery, 1960. 5 pp. mimeo. 19 Aug. 1960.

SIEGEL, S. 1956. *Nonparametric Statistics for the Behavioral Sciences.* McGraw-Hill, New York. 312pp.

SNEDECOR, G. W. 1956. *Statistical Methods.* Iowa State College Press, Ames. 534 pp.

WILSON, K. V. 1956. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.* **53**: 96-101.

WOLFSON, ALBERT. 1954. Weight and fat deposition in relation to spring migration in transient White-throated Sparrows. *Auk* **71**: 413-434.

Department of Zoology, Duke University, Durham, N. C.
(Present address: Institute of Animal Behavior, 31 Fulton St., Newark, N. J.)

Received March, 1964