LITERATURE CITED

ABLE, K. P. 1982. Skylight polarization at dusk influences migratory orientation in birds. Nature 299: 550–551.

ABLE, K. P. 1989. Skylight polarization patterns and the orientation of migratory birds. J. Exp. Biol. 141:241–256.

BALDA, R., AND A. C. KAMIL. 1989. A comparative study of cache recovery by three corvid species. Anim. Behav. 38:486–495.

BRINES, M., AND J. GOULD. 1982. Skylight polarization patterns and animal orientation. J. Exp. Biol. 96:69–91.

HELBIG, A. J., AND W. WILTSCHKO. 1989. The skylight polarization at dusk affects the orientation behavior of Blackcaps, *Sylvia atricapilla*. Naturwissenschaften 76:227–229.

KAMIL, A. C., AND R. P. BALDA. 1985. Cache recovery and spatial memory in Clark's Nutcracker (*Nucifraga columbiana*). J. Exp. Psychol.: Anim. Behav. Processes 11:95–111.

MOORE, F. R., AND J. B. PHILLIPS. 1988. Sunset, skylight polarization and the migration of Yellow-rumped Warblers, *Dendroica coronata*. Anim. Behav. 36:1770–1778.

SCHMIDT-KOENIG, K. 1958. Der Einfluß experimentell veränderter Zeitschätzung auf das Heimfindevermögen von Brieftauben. Naturwissenschaften 45:47.

WALLRAFF, H. G. 1974. Das Navigationssystem der Vögel. Schriftenreihe "Kybernetik," R. Oldenbourg Verlag, München, Wien.

WEHNER, R. 1976. Polarized light navigation by insects. Scient. Amer. 235:106–115.

WILTSCHKO, W., AND R. P. BALDA. 1989. Sun compass orientation in seed-caching Scrub Jays (*Aphelocoma coerulescens*). J. Comp. Physiol. 164: 717–721.

WILTSCHKO, W., AND R. WILTSCHKO. 1982. The role of outward journey information in the orientation of homing pigeons, p. 239–252. *In* F. Papi and H. G. Wallraff [eds.], Avian navigation. Proc. in Life Sciences, Springer-Verlag, Berlin.

WILTSCHKO, W., AND R. WILTSCHKO. 1987. Cognitive maps and navigation in homing pigeons, p. 201–216. *In* P. Ellen and C. Thinus-Blanc [eds.], Cognitive processes and spatial orientation in animal and man. Martinus Nijhoff, Dordrecht, The Netherlands.

The Condor 93:1023–1025

A PROBLEM IN STATISTICAL ANALYSIS: SIMULTANEOUS INFERENCE[1]

KATHLEEN G. BEAL
Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435

HARRY J. KHAMIS
Department of Mathematics and Statistics and Department of Community Health, School of Medicine, Wright State University, Dayton, OH 45435

Key words: Statistics; simultaneous inference; Bonferroni method.

An analysis of variance (ANOVA) problem is often used as a general example of simultaneous inference since decisions about main and interaction effects are made concurrently. Furthermore, the various multiple-comparison techniques (e.g., Fisher's LSD Procedure, Duncan's New Multiple Range Method) are well-known methods of simultaneous inference. We wish to bring to the attention of ornithologists a common statistical error involving simultaneous inferences or conclusions based on two or more tests of hypotheses. Simultaneous inferences are drawn from a family of conceptually related hypotheses. These related hypotheses usually emanate from groups of observations that are collected by an individual researcher or research team (Miller 1981).

As an example, consider a researcher who quantifies several response variables in a single study. Finch (1991) used a number of reproductive measures (e.g., mean laying date, rates of nest failure, initial clutch size, fledgling rates) in a study of the effect of three levels of flooding on timing of reproduction and productivity of House Wrens (*Troglodytes aedon*). Because Finch reaches a set of simultaneous conclusions about the relationships among reproductive measures and flooding, they constitute a family of hypotheses.

We set α, the probability of a type I error (rejection of the null hypothesis when it is true), at a conventional value such as 0.05. Therefore, the probability of accepting the null hypothesis when true, $(1 - \alpha)$, is quite high for a single test of hypotheses (see, for example,

[1] Received 1 July 1991. Final acceptance 22 July 1991.

Zar 1984:44 or Ott 1984:110). When we perform P tests of hypotheses on P response variables, the probability of accepting all null hypotheses when true is $(1 - \alpha)^P$ if the tests are independent. For example, if $P = 3$ then the probability of accepting all three null hypotheses if true is $(1 - 0.05)^3$ or 0.8574 (Johnson and Wichern 1988:187). The combined probability of a type I error for all hypotheses, called the *experimentwise type I error rate* and denoted by α', is then $\alpha' = (1 - 0.8574)$ or 0.1426, not 0.05. The experimentwise type I error rate increases with the number of simultaneous tests of hypotheses; for five tests it is 0.2262, for 10 tests it is 0.4013, etc. So, for $\alpha = 0.05$ and a large number of tests, the chances of rejecting a true null hypothesis at least once, α', becomes much larger than 0.05. According to a formula called the *Bonferroni inequality,* the experimentwise type I error rate, α', is always less than or equal to $(P \cdot \alpha)$ regardless of the correlation structure among the P tests; that is

$$\alpha' \leq P \cdot \alpha.$$

Consequently, in order to insure that the experimentwise type I error rate, α', does not exceed, say 0.05, simply use a value of α equal to $0.05/P$ (see Morrison 1990:32–33).

Researchers seem to be largely unaware of the problem regarding simultaneous conclusions drawn from multiple tests of hypotheses concerning two or more response variables. Introductory statistical texts and statistical references (such as Sokal and Rohlf 1981 and Snedecor and Cochran 1989) used most frequently by ornithologists restrict discussion of the experimentwise type I error rate to multiple comparison techniques for ANOVA. An examination of the 1989 issues of *The Auk, The Condor,* and *The Wilson Bulletin* revealed at least forty instances in which multiple univariate tests were used for a data set without regard to inflated values of the experimentwise type I error rate.

We describe a univariate procedure that addresses this problem (simultaneous inference) which is simpler than the multivariate procedures, but is relatively unknown among researchers. In fact, this univariate alternative, called the *Bonferroni method,* is actually superior to multivariate methods when the number of hypotheses is small (see Johnson and Wichern 1988: 188–190).

For a single test of hypothesis using a pooled sample t-test (as described, for example, by Ott 1984, df = 10, and $\alpha = 0.05$), the null hypothesis is rejected in favor of a two-sided alternative if t is greater than the critical value of 2.228 or less than that of -2.228. If five sets of hypotheses are tested simultaneously the calculation of the test statistics is typical of multiple t-tests but interpretation of the observed level of significance (P value) is adjusted according to the Bonferroni method, a method that accounts for the inflation of the experimentwise type I error incurred by simultaneous tests. The Bonferroni method requires us to divide the probability of a type I error, here 0.05, by the number of tests and to draw our conclusions based on that new level of significance, namely $\alpha' = 0.05/5 = 0.01$. The critical values of t are now ± 3.169, more extreme than those for the single test of hypothesis. This technique allows us to use simple tests of hypotheses and to control our experimentwise type I error rate so that it does not exceed 0.05. Johnson and Wichern (1988) discuss theory and development of the Bonferroni method in multivariate analysis, however some of their formulas are in error (Johnson, pers. comm. with H.K.; corrigenda can be obtained from H.K.).

The use of the Bonferroni method, namely using an alpha level of 0.01 for each t-test in this example, insures that the probability of incorrectly rejecting any one or more of the five null hypotheses is bounded by 0.05. If the Bonferroni method is not used, then each t-test would be conducted at the 0.05 level of significance leading to an experimentwise type I error rate as high as 0.2262. So, null hypotheses might be rejected in good faith when they should not be.

The primary disadvantage of the Bonferroni method is that it may be more conservative than a multivariate procedure, that is, the actual experimentwise type I error rate may be somewhat less than α. Rice (1989) discusses this problem and provides an adjustment, called the sequential Bonferroni technique, which increases the power of the Bonferroni procedure. For discrete data, Tarone (1990) has provided a modified Bonferroni method.

Finally, the Bonferroni method is most effective when P, the number of tests, is small. As P increases, $\alpha' = \alpha/P$ decreases. Since the Bonferroni method is conservative when P is large, the probability of the type II error (failing to reject the null hypothesis when it is false) becomes a concern. Generally, researchers are urged to carefully consider the number of tests of hypotheses during the design phase of their research. Many researchers appear to feel that the more tests of hypotheses the better. Eliminating redundant or unnecessary tests of hypotheses and carefully choosing the sets of tests to be Bonferroni adjusted reduces P, prevents α' from becoming smaller than necessary, and bolsters the power of the tests.

The Bonferroni method provides a solution to a serious statistical problem, namely control of the experimentwise type I error rate for simultaneous inferences. This solution is simple computationally and in terms of interpretation, and it is superior to other more complex multivariate techniques when the number of hypotheses is small. The Bonferroni method can be applied to any test procedure or confidence interval and is not restricted to the use of t-tests (see Morrison 1990 for a variety of applications). It is a simple way of compensating for the multiplicity of significance tests and requires no assumptions other than those necessary for the validity of the individual significance tests (Miller 1981:8).

LITERATURE CITED

Finch, D. M. 1991. House Wrens adjust laying dates and clutch size in relation to annual flooding. Wilson Bull. 103:25–43.

Johnson, Richard A., and Dean W. Wichern. 1988. Applied multivariate statistical analysis. Prentice-Hall, Englewood Cliffs, NJ.

Miller, R. G., Jr. 1981. Simultaneous statistical inference. Springer-Verlag, New York.

MORRISON, DONALD F. 1990. Multivariate statistical methods. McGraw-Hill, New York.

OTT, LYMAN. 1984. An introduction to statistical methods and data analysis. Duxbury Press, Boston, MA.

RICE, WILLIAM R. 1989. Analyzing tables of statistical tests. Evolution 43:223–224.

SNEDECOR, G. W., AND W. G. COCHRAN. 1989. Statistical methods. Iowa State Univ. Press, Ames, IA.

SOKAL, R. R., AND R. J. ROHLF. 1981. Biometry. W. H. Freeman and Co., New York.

TARONE, R. E. 1990. A modified Bonferroni method for discrete data. Biometrics 46:515–522.

ZAR, JERROLD H. 1984. Biostatistical analysis. Prentice-Hall, Englewood Cliffs, NJ.