The Condor 93:200–203

COMMENTARY

REPEATED MEASURES ANALYSIS OF VARIANCE: A COMMENT ON BEAL AND KHAMIS (1990)

G. P. QUINN, *Department of Ecology and Evolutionary Biology, Monash University, Clayton, Victoria, Australia 3168.*
M. J. KEOUGH, *Department of Zoology, University of Melbourne, Parkville, Victoria, Australia 3052.*

In a recent paper, Beal and Khamis (1990) pointed out the difficulties in analyzing data sets consisting of correlated (or non-independent) observations. They recommended a repeated measures analysis of variance (see Winer 1971, Keppel 1982, Milliken and Johnson 1984), a technique long used by psychologists in analyzing repeated observations on groups of subjects (i.e., a "groups by trials" design). Beal and Khamis (1990) then stated that repeated measures analysis of variance (ANOVA) has the same assumptions as the two- or three-way ANOVA and is fairly robust to violation of those assumptions. Finally, the importance of an adequate sample size was discussed. There are a number of issues raised by Beal and Khamis (1990) that warrant further discussion, primarily because their recommendations may easily lead to the misapplication of repeated measures ANOVA. We deal with these in order of complexity.

1. Beal and Khamis (1990) unfortunately repeat the vague statements about the robustness of the ANOVA found in most textbooks. Skewness, outliers and multimodality, often combined with small and/or unequal sample sizes, are commonly encountered in ecological data sets and will violate the assumptions of normality and variance homogeneity seriously enough to affect the ANOVA F tests (Day and Quinn 1989). Exploratory data analysis and screening are essential for detecting such violations and suggesting remedies such as transformations (Hoaglin et al. 1983, James and McCulloch 1985, Tabachnick and Fidell 1989).

2. Beal and Khamis (1990) discuss a solution to correlated data sets favored by many authors, namely analzying each level of the repeated factor separately, using a univariate ANOVA or a t-test. Another common solution is to analyze the experiment as a multiway ANOVA, incorporating the main experimental factors and the repeated factor. In their examples, the analysis would be a two-way ANOVA, with the factors bird species and prey type or foraging method. Such an analysis assumes that the observations are all independent of each other, an assumption that is unlikely to be met when repeated observations are made on individual birds. Violation of this assumption is known to have serious effects on the ANOVA (Kenny and Judd 1986). Repeated measures ANOVAs can provide a partial solution to this problem when each individual is recorded under all levels of the repeated factor, but a more common situation in ornithology is when the researcher has an unknown mixture of independent and correlated observations (some individuals recorded more than once and others not). Such data sets cannot easily be analyzed by either type of ANOVA.

3. It is incorrect for Beal and Khamis (1990) to state that repeated measures ANOVA "requires the same assumptions as two- or three-way ANOVA, . . ." (p. 251); the assumptions for *univariate F* tests involving repeated factors (e.g., prey type and prey type x species in Table 3 of Beal and Khamis 1990) are more restrictive than for the usual ANOVAs (Winer 1971, Keppel 1982, Looney and Stanley 1989). Not only must the assumption of normality and homogeneity of within-group variances apply, but the covariance matrices must be equal across the grouping factor *and* the common covariance matrix must conform to a pattern known as sphericity, which partly implies that all correlations between pairs of repeated measurements are equal (O'Brien and Kaiser 1985). Inflated Type I error rates (i.e., unreliable significance levels) will result from these assumptions not being met, a situation likely for most real ecological (and ornithological) data. There are no useful preliminary tests of the sphericity assumption (Looney and Stanley 1989), so the analyses (and the reliability of the P values) used by Beal and Khamis (1990) must be considered to be doubtful.

There are two useful alternative approaches, however: multivariate analysis of variance (MANOVA) and adjustment of univariate statistics (Potvin et al. 1990). Analyzing the differences between pairs of repeated measurements as a MANOVA removes the sphericity assumption (O'Brien and Kaiser 1985) and can be done routinely by most major statistical programs, although familiarity with the assumptions and interpretation of MANOVA is necessary (Tabachnick and Fidell 1989). Conservative univariate F tests involve adjusting the degrees of freedom for the appropriate F-ratios to allow for violations in the assumptions, e.g., Greenhouse-Geisser and Huynh-Feldt corrections (Winer 1971, Keppel 1982, Looney and Stanley 1989). Our Table 1 illustrates the effects of these alternatives and/or adjustments on the raw data of Beal and Khamis (1990). Note that the Greenhouse-Geisser correction reverses the significant result in their Table 4; also note that we were not able to reproduce all their F values using Type III sums of squares (as recommended by Milliken and Johnson 1984 and others) produced by two major statistical packages (Table 1).

Two further points are worth noting. First, if the repeated measures factor is quantitative (e.g., time), profile analysis (Tabachnick and Fidell 1989) can provide an elegant interpretation of the data. Second, multiple comparison tests often used after the usual ANOVA (Day and Quinn 1989) are rarely applicable to repeated measures designs, although appropriate tests are introduced in Looney and Stanley (1989).

TABLE 1. Comparative analyses of the data sets used by Beal and Khamis (1990). Raw data, obtained from their Tables 1 & 2, were analyzed using two major statistical packages, including the package (SAS) used by Beal and Khamis (1990). The table compares the published analyses with those produced using SYSTAT and SAS-PC (Type III sums of squares method; SYSTAT also uses this method). Where appropriate, we show the probabilities associated with the application of the Greenhouse-Geisser (G-G) correction for violation of the sphericity assumption in a univariate repeated measures, and the calculation of the Pillai trace statistic, a multivariate statistic that is often preferred to its univariate equivalent.

	Beal and Khamis (1990)			SYSTAT/SAS			G-G	Pillai
Source of variation	df	*F*	*P*	MS	*F*	*P*	*P*	*P*
TABLE 3 of Beal and Khamis (1990)—Prey types								
Species	1	2.07	0.2000	17.25	2.00	0.2071		
Bird within species	6			8.63				
Prey type	2	1.92	0.1890	11.51	1.63	0.2371	0.2475	0.3571
Prey type × species	2	3.53	0.0624	25.40	3.59	0.0600	0.0897	0.1024
Bird within species by prey type	12			7.08				
Diptera	1		0.0512		4.18	0.0870		
Trichoptera	1		0.1098		3.56	0.1082		
TABLE 4 of Beal and Khamis (1990)—Foraging methods								
Species	1	1.74	0.2444	12,111	1.74	0.2445		
Bird within species	5			6,965				
Foraging method	2	4.91	0.0195	16,276	4.73	0.0358	0.0717	0.0112
Foraging × species	2	0.87	0.4499	2,981	0.87	0.4495	0.4062	0.1387
Bird within species by foraging method	10			3,438				

4. The discussion of sample size in Beal and Khamis (1990) is timely, particularly given the problems posed by the small sample size of the data sets they chose to illustrate repeated measures ANOVA. Designing ecological experiments to attain required statistical power is well known, although still rarely applied (see reviews by Underwood 1981, Peterman 1990). Such calculations are particularly important for the repeated measures situation, as reliable tests of null hypotheses (MANOVA, adjusted repeated measures ANOVA) will have fewer than anticipated degrees of freedom. Computer-based power analysis programs are readily available and some include repeated measures designs (Goldstein 1989).

In conclusion, Beal and Khamis (1990) have provided ornithologists with a useful discussion about the problems of correlated observations in the data sets and have provided a possible solution, in repeated measures ANOVA, which may become popular for some designs. Although univariate repeated measures ANOVA is a powerful tool, its assumptions are generally more restrictive and less easily tested than those of standard ANOVA. There are less restrictive alternative tests available, and the limitations and alternatives should be recognized before researchers naively process their data with statistical software.

We thank Alan Lill for useful discussion.

LITERATURE CITED

BEAL, K. G., AND H. J. KHAMIS. 1990. Statistical analysis of a problem data set: correlated observations. Condor 92:248–251.

DAY, R. W., AND G. P. QUINN. 1989. Comparisons of treatments after an analysis of variance in ecology. Ecol. Monogr. 59:433–463.

GOLDSTEIN, R. 1989. Power and sample size via MS/PC-DOS computers. Am. Stat. 43:253–260.

HOAGLIN, D. C., F. MOSTELLER, AND J. W. TUKEY. 1983. Understanding robust and exploratory data analysis. John Wiley and Sons, New York.

JAMES, F. C., AND C. E. MCCULLOUGH. 1985. Data analysis and the design of experiments in ornithology, p. 1–63. *In* R. F. Johnston [ed.], Current ornithology. Vol. 2. Plenum Press, New York.

KENNY, D. A., AND C. M. JUDD. 1986. Consequences of violating the independence assumption in analysis of variance. Psychol. Bull. 99:422–431.

KEPPEL, G. 1982. Design and analysis. A researcher's handbook. 2nd ed. Prentice-Hall, Englewood Cliffs, New Jersey.

LOONEY, S. W., AND W. B. STANLEY. 1989. Exploratory repeated measures analysis for two or more groups. Am. Stat. 43:220–225.

MILLIKEN, G. A., AND D. E. JOHNSON. 1984. Analysis of messy data. Vol. 1. Van Nostrand Reinhold Co., New York.

O'BRIEN, R. G., AND M. K. KAISER. 1985. MANOVA method for analysing repeated measures designs: an extensive primer. Psychol. Bull. 97:316–333.

PETERMAN, R. M. 1990. Statistical power analysis can improve fisheries research and management. Can. J. Fish. Aquat. Sci. 47:2–15.

POTVIN, C., M. J. LECHOWICZ, AND S. TARDIF. 1990. The statistical analysis of ecophysiological response curves obtained from experiments involving repeated measures. Ecology 71:1389–1400.

TABACHNICK, B. G., AND L. S. FIDELL. 1989. Using multivariate statistics. 2nd ed. Harper and Row, New York.

UNDERWOOD, A. J. 1981. Techniques of analysis of variance in experimental marine biology and ecology. Oceanogr. Mar. Biol. Annu. Rev. 19:513–605.

WINER, B. J. 1971. Statistical principles in experimental design. 2nd ed. McGraw-Hill, New York.

REPLY TO QUINN AND KEOUGH

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

Our goal in Beal and Khamis (1990) was to bring to the attention of the ornithological research community a common, serious statistical problem, namely, treatment of a correlated data set as if it consisted of independent observations. We chose a real data set, rather than a contrived one, that presented additional challenges (such as small sample size).

Quinn and Keough have brought up several points concerning our paper. We respond to comments 1 and 3 and comment on their Table 1.

Comment 1. Quinn and Keough state that we repeat the vague claims of the robustness of analysis of variance (ANOVA) procedures that are found in many textbooks. Robustness of ANOVA procedures is a controversial issue among statisticians. Some advocate the use of ANOVA when moderate deviations from the assumptions occur (e.g., Montgomery 1984, p. 87, 91). Zar (1984, p. 170) carefully provides primary references for the robustness of ANOVA, concluding that "... analysis of variance may typically be depended upon unless the data deviate severely from the underlying assumptions." The statement that we used in our paper is somewhat milder than these, more in agreement with Quinn and Keough's own statements in Comment 1. We agree with Quinn and Keough that exploratory data analysis is always advisable. But when the data set is very small, as in this case, normality checks will yield little useful information (Montgomery 1984, p. 86), and tests for homogeneity of variances are unreliable (Zar 1984, p. 183).

Comment 3. Quinn and Keough correct a misstatement made concerning the assumptions needed for validity of the repeated measures procedure. We had intended to state that the standard ANOVA assumptions are necessary but not sufficient conditions for the repeated measures procedure. The additional sphericity assumptions mentioned by Quinn and Keough, one form of which is referred to as the Huynh-Feldt conditions (Huynh and Feldt 1970), should have been stated in our paper for informational purposes. However, from a practical point of view, we felt that a discussion of these conditions would contribute little to the point of the paper because they are difficult to test for, especially in a small data set of the type we analyzed, and because of their theoretical complexity. Also, Huynh and Feldt (1980) indicate that some departure from sphericity may not substantially change the nature of the traditional F tests in repeated measures designs (see Read et al. 1988, p. 605–606).

Quinn and Keough suggest two possible alternative forms of analysis. One of these is use of multivariate analysis of variance (MANOVA). We avoided a discussion of this technique because it is our feeling that univariate techniques should be used if possible in order to avoid the additional complexities associated with multivariate procedures. The Greenhouse-Geisser (GG) correction that is used to adjust for violations of the sphericity assumption recommended as the second alternative has been shown in simulation studies to be ultraconservative (Ott 1988, p. 800).

Concerning Table 1. There are a number of reasons why Quinn and Keough's numbers in Table 1 differ somewhat from ours:

1. The rates recorded in Tables 1 and 2 of our paper use one decimal place accuracy, as recommended by a reviewer; however, our analyses were based on three decimal place accuracy. Assuming that Quinn and Keough used the rates as recorded in our paper, their ANOVA table results are somewhat less accurate than ours.
2. Our computations were carried out on an IBM 3083 computer using SAS Version 5; Type III sums of squares were used in computing the *F*-ratios.
3. In Table 4, the test for foraging method was conducted *after the nonsignificant interaction term* (foraging × species) *was dropped from the model,* a very common practice when working with ANOVA models; apparently Quinn and Keough did not drop the interaction term from their model when testing for foraging method. Had they dropped this interaction term, the "reversal" of the significant result that they mention in their comments concerning the GG adjustment might not take place. Alternatively, this reversal may be in part due to the ultra-conservativeness of the adjustment—note that the GG *P*-value is somewhat higher than the other two techniques for which the interaction term was not dropped (SYSTAT/SAS and Pillai).

In conclusion, we note that the *P*-values from our ANOVA tables are in general agreement with those given by the other methods presented in Table 1 of Quinn and Keough, with the above comments in mind. In particular, the same general conclusions would be made regardless of which technique is used. Of course, given the sample size and possible violations of assumptions, the *P*-values must be treated as approximations and, as stated in our paper, care must be used in interpreting the ANOVA table.

LITERATURE CITED

BEAL, K. G., AND H. J. KHAMIS. 1990. Statistical analysis of a problem data set: correlated observations. Condor 92:248–251.

HUYNH, H., AND L. S. FELDT. 1970. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. J. Am. Stat. Assoc. 65:1582–1589.

HUYNH, H., AND L. S. FELDT. 1980. Performance of traditional F tests in repeated measures designs