unpaired. The two paired Mallards might have been unpaired when they first returned in late March or early April and found mates before they were first observed in early May.

I thank B. R. Bacon, R. A. Kahl, T. R. Grunewald, P. J. Fassbender, B. K. Lueth, and T. W. Trapp for field assistance. I also thank L. R. Petersen, A. D. Afton, and A. H. Brush for critical review of the manuscript. Partial funding for this study was provided by the Federal Aid to Wildlife Restoration under Pittman-Robertson Wis. Proj. W-141-R.

LITERATURE CITED

ANDERSON, D. R. 1975. Population of the Mallard. V. Temporal and geographic estimates of survival, recovery, and harvest rates. U.S. Fish and Wildl. Serv. Res. Publ. No. 125.

ANDERSON, M. G., R. D. SAYLOR, AND A. D. AFTON. 1980. A decoy trap for diving ducks. J. Wildl. Manage. 44:217–219.

BLOHM, R. J. 1978. Migrational homing of male Gadwalls to breeding grounds. Auk 95:763–766.

COOCH, G. 1953. Techniques for mass capture of flightless Blue and Snow Geese. J. Wildl. Manage. 17:460–465.

CUMMINGS, G. E., AND O. H. HEWITT. 1964. Capturing waterfowl and marsh birds at night with light and sound. J. Wildl. Manage. 29:120–126.

DOTY, H. A., AND R. J. GREENWOOD. 1974. Improved nasal-saddle marker for Mallards. J. Wildl. Manage. 38:938–939.

DWYER, T. J., S. R. DERRICKSON, AND D. S. GILMER. 1973. Migrational homing by a pair of Mallards. Auk 90:687.

HOCHBAUM, H. 1944. The Canvasback on a prairie marsh. American Wildlife Institute, Washington, DC.

HUNT, G. S., AND K. J. DAHLKA. 1953. Live trapping of diving ducks. J. Wildl. Manage. 17:92–97.

JOHNSON, D. H., AND J. W. GRIER. 1988. Determinants of breeding distributions of ducks. Wildl. Monogr. No. 100.

LINCOLN, F. C. 1934. The operation of homing instinct. Bird-Banding 5:149–155.

LINCOLN, F. C. 1939. The migration of American birds. Doubleday, Doran, NY.

POSTON, H. J. 1974. Home range and breeding biology of the Shoveler. Can. Wildl. Serv. Rep. Ser. No. 25.

ROHWER, F. C., AND M. G. ANDERSON. 1988. Female-based philopatry, monogamy, and the timing of pair formation in migratory waterfowl, p. 187–221. *In* R. F. Johnston [ed.], Current ornithology. Vol. 5. Plenum Press, New York.

SOWLS, L. K. 1955. Prairie ducks. Wildlife Management Institute, Washington, DC.

TITMAN, R. D. 1983. Spacing and three-bird flights of Mallards breeding in pothole habitat. Can. J. Zool. 61:839–847.

The Condor 92:248–251

STATISTICAL ANALYSIS OF A PROBLEM DATA SET: CORRELATED OBSERVATIONS[1]

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

Key words: Statistics; repeated measures; hypothesis testing.

Most statistical test procedures require that a random sample—a set of independent observations—be used for analysis. Observations are independent if the probability of observing one event does not affect the probability of observing another event. A few forms of analysis, such as the paired *t*-test, are designed specifically for correlated or dependent observations. Tests such as the two sample *t*-test, Pearson's chi-square test, the likelihood ratio test or *G*-test, routine analysis of variance, and regression analysis are not valid, however, if the assumption of independent observations is violated. The explanation of the importance of this assumption may be found in mathematical statistics texts (e.g., Hogg and Craig 1978).

A series of behavioral responses from a given bird comprises a set of correlated or dependent observations. Biologists who present data sets consisting of 20, 30, or more observations of behavioral responses from only a few animals may have very little information about the response of the population as a whole. An examination of the 1987 issues of *The Auk, The Condor,* and *The Wilson Bulletin* revealed at least 15 instances in which Pearson's chi-square analysis or the *G*-test (likelihood ratio test) was used for a data set in which the observations were clearly correlated (not independent). This indicates that (1) misuse of Pearson's

---

[1] Received 25 May 1989. Final acceptance 19 September 1989.

TABLE 1. Individual Louisiana and Northern waterthrush captures (and rates of capture, prey taken per hour) for five prey classes.[a]

	Diptera	Trichoptera	Ephemeroptera	Oligochaeta	Isopoda	Total
Louisiana Waterthrushes	2	6	1	0	2	11
	(0.7)	(2.0)	(0.3)	(0.0)	(0.7)	
	4	10	6	8	0	28
	(1.4)	(3.5)	(2.1)	(2.8)	(0.0)	
	7	8	1	0	0	16
	(2.4)	(2.8)	(0.4)	(0.0)	(0.0)	
Total	13	24	8	8	2	55
Northern Waterthrushes	17	6	4	2	3	32
	(8.5)	(3.0)	(2.0)	(1.0)	(1.5)	
	25	2	0	4	0	31
	(14.7)	(1.2)	(0.0)	(2.4)	(0.0)	
	4	2	0	0	0	6
	(2.7)	(1.3)	(0.0)	(0.0)	(0.0)	
	7	3	11	0	1	22
	(3.8)	(1.6)	(6.0)	(0.0)	(0.5)	
	13	0	1	0	1	15
	(6.9)	(0.0)	(0.5)	(0.0)	(0.5)	
Total	66	13	16	6	5	106

[a] Data from Craig 1987 and unpubl.

chi-square and G-test procedures is common, and (2) researchers, reviewers, and editors alike are not aware of the problem.

Ornithologists often study small populations of birds or, if the population is large, are not able to gain access to large numbers of birds due to physical, budgetary, or time constraints. The problem of small sample size is addressed eloquently by Morrison (1988), and his recommendations concerning adequate sample size are supported.

Our goal is to reanalyze a correlated data set for which the original analysis (Pearson's chi-square analysis) is not appropriate. We consider two alternative forms of analysis that are more appropriate for the data.

Our example concerns observations of prey selection and foraging technique of Louisiana (*Seiurus motacilla*) and Northern (*S. noveboracensis*) waterthrushes (Craig 1987). We choose this study because we feel it has features that are typical of many ornithological studies. Specifically, the species involved are found in small numbers, the field season is limited, and many observations can readily be made on each bird.

The author reported that the 55 total prey items consumed by Louisiana Waterthrushes consisted of 13 Diptera, 24 Trichoptera, 8 Ephemeroptera, 8 Oligochaeta, and 2 Isopoda. For Northern Waterthrushes the tally of 106 total prey items consisted of 66 Diptera, 13 Trichoptera, 16 Ephemeroptera, 6 Oligochaeta, and 5 Isopoda. Three foraging methods were observed. Louisiana Waterthrushes were observed picking 435 times, leaf pulling 419 times, and hovering 7 times. For Northern Waterthrushes, the counts were 945, 812, and 12, respectively.

Using Pearson's chi-square test statistics, the author concluded that (1) Louisiana Waterthrushes ate Trichoptera larvae more frequently than did Northern Waterthrushes, (2) Northern Waterthrushes ate more Diptera larvae than did Louisiana Waterthrushes, and (3) the two species did not differ in their use of three foraging methods.

These conclusions, based on the Pearson's chi-square test for homogeneity of proportions, would be valid if a total of 55 Louisiana Waterthrushes and 106 Northern Waterthrushes in the instance of prey selection, for example, had been observed. That is, the theory underlying the Pearson's chi-square test is based on the assumption that the data in the contingency table represent *independent* observations. However, this was not the case. The 55 observed prey selections by Louisiana Waterthrushes were generated by only three birds, and the 106 observations for Northern Waterthrushes by five birds. Because of the correlation among observations (a single bird was observed repeatedly), the chi-square test is rendered invalid. Similar remarks can be made regarding the observations of foraging methods, for which only seven birds were involved.

Table 1 is constructed from Craig's raw data and identifies prey capture counts and rates for the eight birds involved in this study. Similarly, Table 2 contains the same information for use of the three foraging methods. We are still interested in determining whether there is a significant difference in prey capture preferences and use of foraging methods between the two species of waterthrush. However, we must be careful in our analysis to take into account the correlational structure of the observations; i.e., we must be aware that each bird gives rise to more than one observation. Since the amount of time devoted to observing each bird varied, our analyses will be based on the *rates* of

TABLE 2. Individual Louisiana and Northern waterthrush use (and rates of use, behaviors per hour) of three foraging methods.[a]

	Pick	Leaf pull	Hawk	Total
Louisiana Waterthrushes	47 (15.9)	2 (0.7)	2 (0.7)	51
	176 (60.7)	36 (12.4)	4 (1.4)	216
	212 (73.5)	381 (132.1)	1 (0.4)	594
Total	435	419	7	861
Northern Waterthrushes	280 (140.0)	41 (20.5)	3 (1.5)	324
	325 (191.2)	504 (296.5)	9 (5.3)	838
	174 (94.1)	47 (25.4)	0 (0.0)	221
	166 (88.1)	220 (116.8)	0 (0.0)	386
Total	945	812	12	1,769

[a] Data from Craig 1987 and unpubl.

TABLE 3. Repeated measures ANOVA table and pertinent tests of hypotheses for use of three prey categories by waterthrushes (format similar to Sokal and Rohlf 1981).

Source of variation	df	*F*	*P*
Species	1	2.07	0.2000
Bird within species	6	—	—
Prey type	2	1.92	0.1890
Prey type × species	2	3.53	0.0624
Bird within species × prey type	12	—	—

Hypotheses tested	Significance probability
Species and prey type interaction	$P = 0.0624$
(1) Mean rate of consumption of Diptera	$P = 0.0512$
Northern Waterthrushes: 7.31 per hr Louisiana Waterthrushes: 1.50 per hr	
(2) Mean rate of consumption of Trichoptera	$P = 0.1098$
Northern Waterthrushes: 1.43 per hr Louisiana Waterthrushes: 2.75 per hr	

activities (behaviors per hour) rather than counts. Also, because the birds consumed few Ephemeroptera, Oligochaeta, and Isopoda we combined these prey items into one category, thereby reducing the number of prey categories from five to three.

One possible method of statistical analysis is the Mann-Whitney *U*-test (see, e.g., Sokal and Rohlf 1981). For example, to determine whether rates of consumption of Diptera differed for the two waterthrush species, we ranked the capture rate for each bird and proceeded with the test in the usual way. We discovered that the Louisiana Waterthrushes receive the lowest ranks and the Northern Waterthrushes the highest ranks. The test must be repeated for each prey category and foraging method of interest. We chose not to continue this analysis because the multiple tests of the hypothesis involving the same eight waterthrushes are not independent and the experimentwise type I error rate (the likelihood of incorrectly rejecting the null hypothesis in at least one of the tests) inflates. One way to adjust for the inflation of the experimentwise type I error rate is to reduce the nominal level of significance (which is usually 0.05) by dividing it by the number of tests that are conducted (see a discussion of the Bonferroni method, for example in Neter and Wasserman 1974).

The second form of analysis we explored involved the SAS repeated measures ANOVA technique (SAS 1985). A lucid explanation of repeated measures ANOVA is contained in Milliken and Johnson (1984). This analysis takes into account the repeated measurements from a single bird in determining whether there is a significant difference between the two species in terms of prey capture rate and whether those differences depend upon the kind of prey (two-factor interaction effect).

One repeated measure is the rate of prey capture; the second repeated measure of interest is rate of foraging method use. A model representation and important statistical results are found in Tables 3 and 4.

The first conclusion is that there tends to be an interaction between species and prey type ($P = 0.0624$). While the *P*-value is not less than 0.05, it is very close (marginally significant) and indicates that the interaction effect warrants further study. The mean rate of consumption of Diptera was greater for Northern Waterthrushes (7.31 Diptera per hour) than for Louisiana Waterthrushes (1.50 Diptera per hour; $P = 0.0512$). This result is consistent with the author's original conclusion.

Our analysis allows us to conclude that the mean rates of consumption of Trichoptera for the two species do not differ significantly ($P = 0.1098$). This conclusion differs from that of the author's original conclusion. Finally, our results suggest that there is no interaction between foraging method and species ($P = 0.4499$), and the use of the three foraging methods does not differ between the two species ($P = 0.2444$). This conclusion is consistent with the original analysis. However, we are able to conclude that waterthrushes in general favor the leaf pull and the pick, using hawking rarely. This is a predictable result given the energy costs of hawking, especially for species not proficient at hovering (Welty 1982, p. 539).

The repeated measures ANOVA has several advantages over the author's original analysis. It takes into account repeated observations from the same bird; it is based on the activity rate, so that it is not biased by varying behavioral observation time periods; because it is an ANOVA, it simultaneously tests for significance of effects; finally, it contains a test for the species by prey type and species by foraging method interactions.

Ornithological researchers may wish to consider re-

TABLE 4. Repeated measures ANOVA table and pertinent tests of an hypothesis for use of three foraging methods by waterthrushes (format similar to Sokal and Rohlf 1981).

Source of variation	df	*F*	*P*
Species	1	1.74	0.2444
Bird within species	5	—	—
Foraging method	2	4.91	0.0195
Foraging method × species	2	0.87	0.4499
Bird within species × foraging method	10	—	—

Hypothesis tested	Significance probability
Foraging method Mean rate of use of method Hawk: 1.31 per hr Pick: 94.79 per hr Leaf pull: 86.35 per hr	$P = 0.0327$

peated measures ANOVA when planning a study involving multiple observations on marked or identifiable birds. This procedure requires the same assumptions as two- or three-way ANOVA, specifically that the samples are independent and random, that the underlying populations are normally distributed, and that the population variances are equal (see Sokal and Rohlf 1981, chapter 13). That ANOVA is fairly robust (i.e., valid given minor violations of the distributional assumptions) is another one of its merits. Repeated measures ANOVA is available on the most commonly used computer packages (SAS and SPSS-X) though care must be used in interpreting the ANOVA table, especially when the effects under consideration are random and not fixed (see Milliken and Johnson 1984).

Our comments are directed toward the study involving identifiable individuals or periods of continuous observation of individuals. Researchers taking repeated samples from an area do not know if observations of the same individuals recur. Researchers may take a number of precautions to avoid unwitting collection of such correlated data. By choosing a large population for study the observer reduces the probability of multiple observations of a few individuals. Random sampling from various points in the study area decreases the likelihood that the observer inadvertently concentrates efforts on a small part of the population. Collecting observations from various sites—ideally visiting each site once—or allowing an appropriate amount of time to pass between visits to one site are additional considerations for the researcher. Finally, an estimate of population size provides a good indicator of whether or not repeated measures of the same individuals are likely.

In our analyses we dealt with a correlated data set based on small sample size. We have demonstrated that treating a correlated data set as a set of independent observations may lead to incorrect conclusions. Repeated measures analysis is a powerful and useful tool. However, it is most reliable—as are all statistical procedures—when sample size is adequate. The researcher should make every effort to properly determine the optimal sample size. Khamis (1988) discusses the general sample size problem and provides sample size formulas for many of the elementary statistical procedures. Also see Sokal and Rohlf (1981), Morrison (1988), and Snedecor and Cochran (1989). If the realized sample size falls short of this ideal number then care should be exercised in choosing and interpreting statistical procedures.

We are indebted to Robert Craig for sharing his original data set with us and for helpful suggestions from reviewer Michael L. Morrison.

LITERATURE CITED

CRAIG, R. 1987. Divergent prey selection in two species of waterthrushes (Seiurus). Auk 104:180–187.

HOGG, R. V., AND A. T. CRAIG. 1978. Introduction to mathematical statistics. Macmillan Publishing Co., New York.

KHAMIS, H. J. 1988. Statistics refresher II: choice of sample size. J. Diagn. Med. Sonography 4:176–183.

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

MORRISON, M. L. 1988. On sample size and reliable information. Condor 90:275–278.

NETER, J., AND W. WASSERMAN. 1974. Applied linear statistical models. Richard D. Irwin, Homewood, IL.

SAS INSTITUTE. 1985. SAS user's guide: statistics. Version 5. SAS Institute, Cary, NC.

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

SOKAL, R. R., AND F. J. ROHLF. 1981. Biometry. 2nd ed. W. H. Freeman and Co., San Francisco.

WELTY, J. C. 1982. The life of birds. Saunders Publishing, New York.