Detecting Differences in Fish Diets

David A. Somerton

Honolulu Laboratory, Southwest Fisheries Science Center
National Marine Fisheries Service, NOAA, 2570 Dole Street
Honolulu, Hawaii 96822-2396

Statistical comparison of the diet of a predator between areas or time-periods allows one to distinguish true dietary differences from sampling variability and may lead to a better understanding of a species' feeding habits. Despite the utility of statistical testing, few procedures appropriate for dietary comparisons have been developed. Perhaps one impediment to the development of a general approach to dietary comparisons is the wide variety of ways in which diets have been expressed and the lack of consensus about which is best. For example, diets expressed as the numeric or gravimetric proportions of the total food consumed will require different approaches to statistical comparison than those expressed either as the proportion of the samples containing each of the various prey types, or as the index of relative importance of each prey type (Pinkas et al. 1971).

For cases in which diets are expressed in terms of gravimetric proportions, Crow (1979) and Ellison (1979) have recommended statistical tests of between-sample differences based on multivariate analysis of variance (MANOVA). Validity of such tests, however, requires that the prey proportions have a multivariate normal distribution and that the variance-covariance structure of the prey proportions is identical among samples (Morrison 1976). Recognizing that dietary data are unlikely to have these properties, Crow (1979) further recommended using MANOVA that incorporates non-parametric procedures. Herein, this recommendation is followed, and a new approach for testing differences in diets using non-parametric MANOVA is examined. This approach combines the usual measure of between-sample differences employed in parametric MANOVA (i.e., Hotelling's T^2 statistic; Morrison 1976) and a non-parametric procedure (i.e., a randomization test; Edgington 1987) to determine the significance of T^2. The method is then applied to determine whether the diet of pelagic armorhead *Pseudopentaceros wheeleri* from the Southeast Hancock Seamount changed between two sampling periods.

Materials and methods

Testing for between-sample differences is accomplished in three steps: (1) calculating for each sample the gravimetric dietary proportions and their variances and covariances, (2) calculating a measure of the statistical difference between samples, and (3) determining the statistical significance of the measure. The gravimetric proportion of the diet contributed by prey category j(p_j) can be estimated as the total weight of prey category j in all stomach samples divided by the total weight of all prey categories combined (Hyslop 1980). Algebraically this is expressed as

$$p_j = \frac{\Sigma_k w_{jk}}{\Sigma_j \Sigma_k w_{jk}} = \frac{w_{j.}}{w_{..}} \qquad (1)$$

where w_{jk} is the weight of prey category j for individual k, $w_{j.}$ is the total weight of prey category j summed across all individuals, and $w_{..}$ is the total weight of all prey. Each p_j is transformed to x_j, where $x_j = \arcsin \sqrt{p_j}$, so that it conforms more closely to a normal random variable (Sokol and Rohlf 1969). Because x_j is estimated as a pooled proportion rather than the average of the proportions for individual fish, the variance of x_j and the covariance between x_j and x_i cannot be calculated directly and instead are approximated by using the delta method (Seber 1973). In the following, $\mathbf{x}_i$ indicates the vector of x_j for sample i, and $\mathbf{S}_i$ indicates the matrix of variance and covariance estimates for $\mathbf{x}_i$.

The measure of statistical difference used is the Hotelling's T^2 statistic, a multivariate extension of the t-statistic (Morrison 1976). In matrix notation, this statistic is expressed as

$$T^2 = (\mathbf{x}_1 - \mathbf{x}_2)' \, \mathbf{S}.^{-1} \, (\mathbf{x}_1 - \mathbf{x}_2), \qquad (2)$$

where $\mathbf{S}.^{-1}$ is the inverse of the pooled estimate of the variance-covariance matrix (Morrison 1976). $\mathbf{S}.$ is approximated, assuming reasonably large sample sizes (>50 stomachs with prey per sample), as

$$\mathbf{S}. = \frac{2(N_1\mathbf{S}_1 + N_2\mathbf{S}_2)}{N_1 + N_2}, \qquad (3)$$

where N_1 and N_2 are the sizes of the two samples.

Once a value of T^2 is computed, its significance is determined from an empirical probability distribution of T^2 computed by using a technique known as randomization (Edgington 1987). Computation of the empirical probability distribution using this technique proceeds as follows: (1) stomach content data from both time or area samples are

Manuscript accepted 12 October 1990.
Fishery Bulletin, U.S. 89:167–169 (1991).

Table 1
Probability levels associated with the randomization tests of the individual hypotheses. Because these are *a posteriori* tests, significance levels were adjusted according to the Bonferroni inequality to 0.25α ($^*P \leq 0.0125$, $^{**}P \leq 0.0025$).

	Mean proportion			
	Sample 1	Sample 2	t-value	Probability level
Tunicates	0.32	0.69	−9.1	<0.001**
Crustaceans	0.27	0.13	5.0	0.011*
Fish	0.26	0.10	3.7	0.046
"Others"	0.15	0.08	1.9	0.294

combined, (2) the combined data are randomly sorted into two new samples equal in size to the originals, and (3) a value of T^2 is calculated for the two samples. Steps 2 and 3 are repeated iteratively (the present study uses 5000 iterations), and the probability distribution of the randomized values of T^2 is then calculated. Next, the significance level of obtaining the original T^2 is estimated by determining the proportion of the randomized T^2 values that, ignoring signs, is equal to or greater than the original.

If between-sample equality of the diet is rejected, it is then appropriate to test for the equality of individual prey categories to determine which categories contribute most to the difference in diet. In this case, the measure of statistical difference used is the univariate t-statistic. In matrix notation, the vector of t-statistics (**t**) can be computed as

$$\mathbf{t} = \mathbf{D}.^{-1}(\mathbf{x}_1 - \mathbf{x}_2), \quad (4)$$

where $\mathbf{D}.^{-1}$ is the inverse of a matrix formed from the diagonal elements of $\mathbf{S}.$. As before, the significance of each t-statistic is determined from an empirical probability distribution computed for each prey category using randomization. Computation of these probability distributions is identical to that described for the multivariate case, except that a matrix of univariate t-statistics (Eq. 4) is used in the calculation instead of the T^2-statistic (Eq. 2). These individual tests are *a posteriori* tests and require some adjustment of the error rate considered to be significant. Using the Bonferroni inequality (Morrison 1976, Miller 1981), an appropriate adjustment is to assume significance at $\alpha \cdot n^{-1}$.

The above procedures have been incorporated into the computer program DIETTEST, which is designed to run on IBM-compatible microcomputers. This program is available from the author.

As an example of the application of this method, it has been used to test for dietary differences between two samples of pelagic armorhead: 55 fish collected in June 1985, and 101 fish collected in August 1988 (only fish with prey in their stomachs were used in the analysis). Both samples were obtained from the Southeast Hancock Seamount (lat. 30°N, long. 180°W) on the Hawaiian Ridge. Stomach contents were sorted to the lowest taxonomic category possible, then blotted and weighed to the nearest milligram. To simplify the analysis, various prey items were pooled into four major prey categories: tunicates, crustaceans, fish, and "others."

Results and discussion

When the method was applied to the two armorhead samples, the test of the simultaneous equality of all dietary proportions between samples was highly significant ($P<0.001$), indicating that the diet of pelagic armorhead had changed between the two sampling periods. Because of this, tests were also made for individual prey categories. Two of the four prey categories, tunicates and crustaceans, differed significantly between samples (Table 1) and therefore appeared to be responsible for the overall difference in diet.

The measure of between-sample difference, T^2, employed in the proposed test was selected primarily because the absolute differences between samples are scaled by the within-sample variances, a particularly desirable feature when dealing with highly variable quantities such as fish diets. This choice, however, imposes a constraint on the proposed statistical test; that is, the method can only be applied to cases in which the two diet samples lack mutually exclusive components. This constraint arises because computation of T^2 requires inversion of the variance-covariance matrix (Eq. 2) which is singular and therefore not invertible when a prey category is completely absent from one of the samples. Although this constraint may not be severe when the diet of a single predator is being examined for spatial or temporal variation, especially if one is willing to accept the pooling of prey to relatively high taxa, the proposed method is likely to be of limited value for comparing the diets of different predators. For such between-predator comparisons, a more appropriate test could be developed by utilizing some measure of diet overlap (Caillet and Barry 1979) which is not affected by mutually exclusive prey categories, instead of T^2 as a measure of between-sample difference.

Acknowledgments

I thank George Boehlert, Tim Gerrodette, John Hoenig, Russell Kappenman, Jeff Polovina, and two anonymous

referees for reviewing the manuscript and offering helpful suggestions. In addition, I thank Donald Kobayashi for his programming assistance and Mike Seki for initially posing the problem.

Citations

Caillet, G.M., and J.P. Barry
1979 Comparison of food array overlap measures useful in fish feeding habit analysis. *In* Lipovsky, S.J., and C.A. Simenstad (eds.), Gutshop '78 fish food habits studies; Proceedings of the second Pacific Northwest technical workshop, October 10–13, 1978, Maple Valley, Washington, p. 67–79. Wash. Sea Grant Publ., Univ. Wash., Seattle.

Crow, M.E.
1979 Multivariate statistical analysis of stomach contents. *In* Lipovsky, S.J., and C.A. Simenstad (eds.), Gutshop '78 fish food habits studies; Proceedings of the second Pacific Northwest technical workshop, October 10–13, 1978, Maple Valley, Washington, p. 87–96. Wash. Sea Grant Publ., Univ. Wash., Seattle.

Edgington, E.S.
1987 Randomization tests, 2d ed. Marcel Dekker, NY, 341 p.

Ellison, J.P.
1979 The use of discriminate analysis in the study of fish food habits. *In* Lipovsky, S.J., and C.A. Simenstad (eds.), Gutshop '78 fish food habits studies; Proceedings of the second Pacific Northwest technical workshop, October 10–13, 1978, Maple Valley, Washington, p. 80–86. Wash. Sea Grant Publ., Univ. Wash., Seattle.

Hyslop, E.J.
1980 Stomach contents analysis—a review of methods and their application. J. Fish. Biol. 17:411–429.

Miller, R.G. Jr.
1981 Simultaneous statistical inference, 2d ed. Springer-Verlag, NY, 299 p.

Morrison, D.F.
1976 Multivariate statistical methods, 2d ed. McGraw-Hill, NY, 415 p.

Pinkas, L., M.S. Oliphant, and I.L.K. Iverson
1971 Food habits of albacore, bluefin tuna, and bonito in California water. Calif. Dep. Fish Game. Fish. Bull. 152, 105 p.

Seber, G.A.F.
1973 The estimation of animal abundance and related parameters. Hafner Press, NY, 506 p.

Sokol, R.R., and F.J. Rohlf
1969 Biometry. W.H. Freeman, San Francisco, 776 p.