CLARIFICATION OF A TECHNIQUE FOR ANALYSIS OF UTILIZATION-AVAILABILITY DATA

C. RANDALL BYERS and R. K. STEINHORST, Faculty of Applied Statistics, University of Idaho, Moscow, ID 83843; and P. R. KRAUSMAN, School of Renewable Natural Resources, University of Arizona, Tucson, AZ 85721.

Neu et al. (1974) describe a statistical technique for calculating simultaneous confidence intervals for analyzing utilization-availability data. The technique is useful in determining preference of a dietary component in relation to its availability. Krausman (1978) used the technique to evaluate forage preferences of deer in relation to availability; Nelson (1979) employed the test to evaluate the hypothesis that deer used available habitat types in proportion to their occurrence. Although useful to biologists, it is not entirely clear from Neu et al. (1974) how the confidence intervals are actually computed. This paper presents a computational example of the technique.

Data collected on four male bighorn sheep (*Ovis canadensis*) in the Harquahala Mountains, Arizona, during 1980 are used to present the example. The sheep were captured and fitted with radio collars as part of a study to determine their distribution and movements. To establish "utilization," weekly aerial surveys were undertaken, and each location of a sheep was assigned to a particular vegetation type. Vegetation types were based on the dominant plants as determined from ground observations.

Data Analysis Using Chi-Square Test

The confidence interval technique of Neu et al. (1974) is often used in conjunction with a chi-square goodness-of-fit test. This chi-square test can be used to determine whether there is a significant difference between the "expected" utilization of vegetation types (based upon their availability) and the observed frequency of their usage. If a statistically significant difference is found between the utilization and availability of the vegetation types, the biologist may further investigate the data by using Bonferroni confidence intervals to determine which vegetation types are being preferred. The Bonferroni probability statements hold without reference to the initial significance of the chi-square test or whether or not it has been conducted.

The chi-square and Bonferroni procedures involve count (enumeration) data; the biologist needs to find the observed number of instances of use and the "expected" number of occurrences based upon the known availability of vegetation types in the field. For the sheep study, vegetation types were outlined on a 7.5-minute series topographic map (scale 1: 24,000), and areas were found with a planimeter (Table 1). The expected number of observations in each type was computed by multiplying the relative area of the type by the total number of sheep locations (183).

With these data the chi-square goodness-of-fit test ($\chi^2 = \sum(O_i - E_i)^2/E_i$) show significant differences between overall availability and usage ($P < 0.001$, $\chi^2 = 103.1$, df = 9). A sufficiently large sample size is needed to conduct this test to assure a reasonable approximation to the chi-square distribution. Roscoe and Byars (1971) gave a summary of necessary sample sizes for various conditions. Koehler and Larntz (1980) presented results that are useful when small expected frequencies are encountered. In this study each expected usage is greater than five, ensuring that an adequate sample was taken.

Table 1. Utilization-availability data for vegetation types in the Harquahala Mountains, western Arizona, 1980. Utilization is based on 183 locations of four male bighorn sheep using radiotelemetry techniques.

Vegetation formation (type)	Total area (ha)	Relative area (P_{io})	Expected usage ($E_i = np_{io}$)	Observed usage (O_i)[a]
1. *Cercidium microphyllum/ Encelia farinosa*	3,353	0.237	43.46	31
2. *Cercidium microphyllum–Cereus giganteus/ Encelia farinosa*	2,297	0.163	29.77	2
3. *Cercidium microphyllum–Simmondsia chinensis/ Eriogonum fasciculatum*	1,590	0.113	20.60	15
4. *Cercidium microphyllum/ Franseria dumosa*	1,087	0.077	14.09	16
5. *Cercidium microphyllum–Cereus giganteus/ Eriogonum fasciculatum*	817	0.058	10.60	2
6. *Cercidium microphyllum/ Eriogonum fasciculatum*	847	0.060	10.98	10
7. *Cereus giganteus–Canotia holacantha Eriogonum fasciculatum*	1,014	0.072	13.14	28
8. *Quercus turbinella–Canotia holacantha/Rhus ovata, Eriogonum fasciculatum*	872	0.062	11.31	29
9. *Cercidium microphyllum–Acacia greggii/Eriogonum fasciculatum–Hilaria rigida*	1,026	0.073	13.30	29
10. Others	1,215	0.086	15.76	21
Total	14,118	1.000	183.00	183

[a] $\chi^2_9 = 103.07$; $\chi^2_{9,}$ $P = 16.92$.

Simultaneous Confidence Intervals

If the biologist was specifically interested in making a statement about the true proportion of utilization (p) of a single, *preselected* vegetation type, a confidence interval for this proportion would likely be constructed using the formula:

$$\bar{p} - Z_{\alpha/2}\sqrt{\bar{p}(1 - \bar{p})/n} \le p \le \bar{p} + Z_{\alpha/2}\sqrt{\bar{p}(1 - \bar{p})/n}$$

where $Z_{\alpha/2}$ is the upper standard normal table value corresponding to a probability tail area of $\alpha/2$. Thus, if we were to construct an interval estimate for the true proportion of utilization for *Cercidium microphyllum/Encelia farinosa*, the appropriate 95% interval estimate would be:

$$0.169 - 1.96\sqrt{0.169(0.831)/183} \le p \le 0.169 + 1.96\sqrt{0.169(0.831)/183}$$

$$0.115 \le p \le 0.223$$

where $\bar{p} = 31/183$.

However, seldom does the biologist just wish to investigate a single vegetation type, rather the interest is in the entire set of

Table 2. Simultaneous confidence intervals using the Bonferroni approach for utilization of vegetation types, P_i.

Vegetation type	Expected proportion of usage P_{io}	Actual proportion of usage P_i	Bonferroni intervals for P_i
1	0.237	0.169	$0.091 \le P_1 \le 0.247$
2	0.163	0.011	$0 \le P_2 \le 0.033$*
3	0.113	0.082	$0.025 \le P_3 \le 0.139$
4	0.077	0.087	$0.028 \le P_4 \le 0.145$
5	0.058	0.011	$0 \le P_5 \le 0.033$*
6	0.060	0.055	$0.008 \le P_6 \le 0.102$
7	0.072	0.153	$0.078 \le P_7 \le 0.228$*
8	0.062	0.158	$0.082 \le P_8 \le 0.234$*
9	0.073	0.158	$0.082 \le P_9 \le 0.234$*
10	0.086	0.115	$0.049 \le P_{10} \le 0.181$

* Indicates a difference at the 0.05 level of significance.

types. When this is the case, as it most often is, the biologist must modify the above procedure and construct simultaneous confidence intervals. Bonferroni's inequality (Miller 1966:216–217, Neu et al. 1974) provides a solution. One can be at least $100(1 - \alpha)\%$ confident that the intervals contain their respective true proportions, p_i:

$$\bar{p}_i - Z_{\alpha/2k}\sqrt{\bar{p}_i(1 - \bar{p}_i)/n} \le p_i \le \bar{p}_i + Z_{\alpha/2k}\sqrt{\bar{p}_i(1 - \bar{p}_i)/n}$$

where $Z_{\alpha/2k}$ is the upper standard normal table value corresponding to a probability tail area of $\alpha/2k$; k is the number of categories tested.

A set of simultaneous confidence intervals was constructed for the true proportion of utilization (p_i) of each of the 10 vegetation types (Table 2). Where the expected proportion of usage, P_{io}, does not lie within the interval, we conclude the expected and actual utilization are significantly different. Here P_{io} represents the expected relative utilization and corresponds to the relative area of the vegetation type. In this case, with $\alpha = 0.05$ and k equal to 10 categories, $Z_{\alpha/2k} = Z_{0.0025} = 2.807$. In contrast to our previous interval for *Cercidium microphyllum/Encelia farinosa*, we obtained the following Bonferroni interval for p_1:

$$0.169 - 2.807\sqrt{0.169(0.831)/183} \le p_1 \le 0.169 + 2.807\sqrt{0.169(0.831)/183}$$

$$0.091 \le p_1 \le 0.247.$$

The Bonferroni intervals (Table 2) show that forage Types 2 and 5 are utilized less than would be expected by chance, whereas Types 7, 8, and 9 are used more than would be expected by chance.

Note here that the level of significance, α, refers to the entire set of intervals. When one is interested in only a single interval for a single, preselected vegetation type, a narrower interval can be obtained. However, in this case we were interested in the entire set of vegetation-type categories and, hence, the experiment-wise or family-wise intervals are required. The resulting interval estimates are termed a $100(1 - \alpha)\%$ "family" of confidence intervals.

As a final note, the researcher must be aware that the effectiveness of this test as with many other statistical techniques depends largely on the manner in which the utilization data are collected. Data collection procedures must be such that animals that are studied have access to and opportunity to be collected (observed) in the various availability categories. The applicability of the procedures depends on the sheep moving independently of each other. The temporal spacing of the observations must be such that they are not autocorrelated.

LITERATURE CITED

KOEHLER, K. J., AND K. LARNTZ. 1980. An empirical investigation of goodness-of-fit statistics for sparse multinomials. J. Am. Stat. Assoc. 75:336–344.

KRAUSMAN, P. R. 1978. Forage relationships between two deer species in Big Bend National Park, Texas. J. Wildl. Manage. 42:101–107.

MILLER, R. G. 1966. Simultaneous statistical inferences. McGraw-Hill Book Co., New York, N.Y. 272pp.

NELSON, M. E. 1979. Home range location of white-tailed deer. U.S. For. Serv. Res. Pap. NC-173. 10pp.

NEU, C. W., C. R. BYERS, AND J. M. PEEK. 1974. A technique for analysis of utilization-availability data. J. Wildl. Manage. 38:541–545.

ROSCOE, J. T., AND J. A. BYARS. 1971. An inspection of the restraints with respect to sample size commonly imposed on the use of the chi-square statistic. J. Am. Stat. Assoc. 66:755–759.

Received 24 October 1983.
Accepted 15 November 1983.

THE JOLLY-SEBER METHOD APPLIED TO AGE-STRATIFIED POPULATIONS

S. LYNNE STOKES,[1] U.S. Fish and Wildlife Service, Migratory Bird and Habitat Research Laboratory, Laurel, MD 20708.

Capture–recapture experiments have long been used by wildlife biologists interested in estimating animal population size and survival rates. Jolly (1965) and Seber (1965) were first to find maximum likelihood estimators for these parameters from stochastic models for an open population. Their likelihood functions, although not identical, are both built as products of multinomial densities and their point estimators are identical. Robson (1969) reproduced the Jolly-Seber estimators by using a multihypergeometric likelihood function and generalized it by allowing an animal's capture history to affect its survival rate. Pollock (1975) extended this work by allowing an animal's capture history to affect capture rate as well. More recently, Pollock (1981) considered a different type of heterogeneity within the population, i.e., that which arises when capture and survival probabilities are age-specific. His model follows the multihypergeometric approach, and he assumes that there is one capture period per year for k years, where a year is defined as the period of time an animal remains in a single age-class. Furthermore, he assumes that there are $l + 1$ distinguishable age-classes, each allowed to have a different capture rate in the i^{th} sample and a different survival rate from the i^{th} to $(i + 1)^{th}$ sample. He finds maximum likelihood estimators (MLE) of survival rates and population sizes for all age groups, except that it is not possible to estimate the number of young animals in the population for any sampling period.

Estimation of population size for young for all sampling periods except the first of a year, as well as for adults, is possible if more than one sampling period per year is allowed. In this paper I provide MLE's for population size and survival rates for young and adult (two age-classes) for this case, and their asymptotic variances. The likelihood function used is of the Jolly-Seber (multinomial) type rather than the hypergeometric.

The multinomial model requires the assumption that each individual in a given subgroup of the population has the same probability of survival to the next capture period, but the number actually surviving is a random variable rather than a constant. Likewise, the number of surviving individuals in a subgroup that is captured at any time is also a random variable. By contrast, in the hypergeometric models,

[1] Present address: Center for Statistical Texas, Austin, TX 78712.