TRENT, T. T., AND O. J. RONGSTAD. 1974. Home range and survival of cottontail rabbits in southwestern Wisconsin. J. Wildl. Manage. 38:459–472.

TREXLER, J. C., AND J. TRAVIS. 1993. Nontraditional regression analysis. Ecology 74:1629–1637.

TURKOWSKI, F. J., AND L. D. MECH. 1968. Radiotracking a young male raccoon. J. Minnesota Acad. Sci. 35:33–38.

VAN HORNE, B. 1983. Density as a misleading indicator of habitat quality. J. Wildl. Manage. 47: 893–901.

VERNER, J., M. L. MORRISON, AND C. RALPH, EDITORS. 1986. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. Univ. Wisconsin Press, Madison. 470pp.

WEISBERG, S. 1985. Applied linear regression. John Wiley and Sons, New York, N.Y.

WETHERILL, G., P. DUNCOMBE, M. KENWOOD, J. KOLTERSTROM, S. R. PAUL, AND B. VOWDEN. 1986. Regression analysis with applications. Chapman and Hill, London, U.K.

WHITE, G. C., AND R. A. GARROTT. 1990. Analysis of wildlife radio-tracking data. Acad. Press, San Diego, Calif. 383 pp.

WHITMORE, R. C. 1981. Structural characteristics of grasshopper sparrow habitat. J. Wildl. Manage. 45:811–814.

WILLIAMS, B. K. 1981. Discriminant analysis in wildlife research: theory and applications. Pages 59–71 *in* D. E. Capan, ed. The use of multivariate statistics in studies of wildlife habitat. U.S. For. Serv. Gen. Tech. Rep. RM-87. Rocky Mountain For. and Range Exp. Stn., Fort Collins, Colo.

Received 24 May 1995.
Accepted 1 February 1996.
Associate Editor: McDonald.

A COMPARISON OF CONFIDENCE INTERVAL METHODS FOR HABITAT USE-AVAILABILITY STUDIES

STEVE CHERRY, Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717, USA

Abstract: Wildlife managers routinely compute sets of simultaneous confidence intervals to estimate the actual proportion of use of a set of k habitat types. Confidence intervals are determined by assuming that the counts of observed use are from k binomial populations. A set of k intervals is constructed from a large sample approximation for a confidence interval for a single binomial proportion. The simultaneous confidence level is controlled by use of the Bonferroni inequality. The coverage probability of these intervals can be less than the nominal $(1 - \alpha)$ 100% level. This paper presents results of a simulation study comparing the performance of these intervals with 3 alternatives; the usual method with a continuity correction factor, and 2 methods of computing confidence intervals for multinomial proportions. The 2 latter methods are superior and should be used in place of the binomial intervals.

J. WILDL. MANAGE. 60(3):653–658

Key words: binomial, confidence intervals, habitat studies, multinomial, resource selection, use-availability studies.

Habitat use-availability studies are common in the wildlife sciences. The primary method of determining preference or avoidance of habitats by a species has been to count the number of times individuals used a particular habitat and compare observed counts with expected counts in a Chi-square goodness-of-fit test. Expected counts are computed under a null hypothesis of no difference between true proportions of use and actual proportions available of the habitat types. Rejection of the null hypothesis typically is followed by computation of a set of $100(1 - \alpha)\%$ simultaneous confidence intervals in an effort to estimate the true proportions of use. If the proportion available of a specific habitat lies below (above) the lower (upper) limit of its associated confidence interval then the conclusion is that the species is choosing (avoiding) that type.

If there are k habitat types, then a set of k

intervals is determined.The method of interval construction is based on a large sample approximation for a confidence interval for a single binomial proportion with the overall confidence level controlled with the Bonferroni inequality. Descriptions in the wildlife literature are numerous (Neu et al. 1974, Byers et al. 1984, Alldredge and Ratti 1986). The validity of the intervals (and the preliminary goodness-of-fit test) is based on an assumption of approximate normality of the sample proportions, an assumption that requires large sample sizes. One standard rule-of-thumb for deciding when sample sizes are large enough for the goodness-of-fit test is that the expected number of observations in each habitat type under the null hypothesis is ≥ 5. Letting p_j denote the observed use proportion for the jth habitat type and N denote the total number of observations, a similar rule-of-thumb for the intervals is that Np_j and $N(1 - p_j)$ should both be ≥ 5 for $j = 1, \ldots, k$.

Another necessary assumption is that of independent observations. The requirements for independent observations depend on study design. For example, the requirements in a study where 1 animal is followed over time differ from the requirements in a study where several animals are observed at 1 point in time. The availability proportions are assumed to be measured without error if the intervals are used to determine habitat avoidance/preference. The validity of inferences drawn from this method is also highly dependent on the choice of which habitat types are available (Johnson 1980). Thomas and Taylor (1990) contains a detailed discussion of the assumptions of various designs for use-availability studies, including methods based on confidence intervals.

The use of the goodness-of-fit test before construction of the intervals is not necessary for them to be valid (Byers et al. 1984). The procedure of a goodness-of-fit test followed by construction of the intervals, if the null hypothesis is rejected, is of questionable use. Results of the test can lead to rejection of the null hypothesis with none of the intervals indicating preference or avoidance, or test results may lead to a decision to fail to reject the null hypothesis with the intervals providing evidence of differential habitat selection. These inconsistencies are referred to in the statistical literature as a lack of consonance and lack of coherence. Hochberg and Tamhane (1983:44–47) discuss these concepts, describing coherence as an "essential property" of valid hierarchical multiple comparison methods.

There are other methods of producing simultaneous confidence intervals for proportions (Quesenberry and Hurst 1964, Goodman 1965, Bailey 1980) that have not been considered by researchers in the wildlife sciences. These intervals are also based on large sample properties, but are more robust and not as sensitive to small sample sizes.

My goal in this paper is to compare 4 methods of producing confidence intervals for habitat use-availability studies. The 4 methods are described first, and an example of their use is given. This section is followed by the description of a set of simulations on which the comparisons are based. Conclusions and recommendations are discussed last.

I acknowledge R. J. Boik, and 2 referees for comments on the manuscript.

Description of the Intervals

For habitat types 1, 2, ..., k let $\pi_1, \pi_2, \ldots, \pi_k$ ($\pi_j \geq 0$, $\Sigma_{j=1}^{k} \pi_{j=1}$) be the probabilities that a random observation will fall in each of the k habitat types. For a sample of N observations let n_j be the number of observations that fall in type j ($\Sigma_{j=1}^{k} n_j = N$). The maximum likelihood estimators of the probabilities π_j are

$$p_j = \frac{n_j}{N} \qquad j = 1, 2, \ldots, k.$$

A set of k large sample simultaneous binomial confidence intervals for the k parameters $\pi_1, \pi_2, \ldots, \pi_k$ has the form (Byers et al. 1984)

$$p_j^- \leq \pi_j \leq p_j^+ \qquad (j = 1, 2, \ldots, k)$$

where

$$p_j^- = p_j - z_{(1-\alpha/2k)} \sqrt{\frac{p_j(1 - p_j)}{N}},$$

$$p_j^+ = p_j + z_{(1-\alpha/2k)} \sqrt{\frac{p_j(1 - p_j)}{N}},$$

and where $z_{(1-\alpha/2k)}$ is the $(1 - \alpha/2k)$ 100th percentile of a standard normal distribution. The lower endpoint is truncated to 0 if it is negative and the upper endpoint is truncated to 1 if it exceeds 1. The interpretation of the set of intervals is that one can be at least $100(1 - \alpha)\%$

confident that all of the true (but unknown) π_j's are contained in their respective intervals. The derivation is based on the asymptotic multivariate normality of

$$\frac{(p_j - \pi_j)}{\sqrt{\frac{p_j(1 - p_j)}{N}}} \qquad j = 1, 2, \ldots, k. \qquad (1)$$

Note that if $n_j = 0$ for any j then both p_j^- and p_j^+ will be 0. Although, Byers et al. (1984) did not derive these intervals they will be referred to as the Byers intervals for convenience.

The use of a continuity correction is generally recommended for small samples sizes. The formula with the continuity correction factor is

$$p_j^- = p_j - \left(z_{(1-\alpha/2k)}\sqrt{\frac{p_j(1 - p_j)}{N}} + \frac{1}{2N}\right),$$

$$p_j^+ = p_j + \left(z_{(1-\alpha/2k)}\sqrt{\frac{p_j(1 - p_j)}{N}} + \frac{1}{2N}\right).$$

with the lower endpoint again being set equal to 0 if it is negative and the upper endpoint is 1 if it exceeds 1 (Blyth and Still 1983). These intervals will be referred to as the Byers$^+$ intervals.

Quesenberry and Hurst (1964) derived sets of simultaneous confidence intervals based on the asymptotic multivariate normality of

$$\frac{(p_j - \pi_j)}{\sqrt{\frac{\pi_j(1 - \pi_j)}{N}}} \qquad j = 1, 2, \ldots, k. \qquad (2)$$

Small sample sizes (which affect the asymptotic normality of the p_j's) will have less influence on the convergence of (2) to normality than on (1) to normality.

Goodman (1965) suggested an improvement of the Quesenberry and Hurst (1964) intervals. His modification resulted in intervals that were shorter. Bailey (1980) improved on Goodman's (1965) intervals by considering a square root transformation of the counts. His method is thus based on an assumption of normality of the transformed counts. Bailey (1980) also presented evidence that his intervals and Goodman's intervals performed better with the use of a continuity correction factor. The resulting formulas for Goodman's intervals, incorporating the continuity correction factor, are

$$p_j^- = [B + 2(n_j - 0.5) - \sqrt{B(B + 4(n_j - 0.5)(N - n_j + 0.5)/N)}] \div 2(N + B),$$

$$p_j^+ = [B + 2(n_j + 0.5) + \sqrt{B(B + 4(n_j + 0.5)(N - n_j - 0.5)/N)}] \div 2(N + B),$$

where B is the upper (α/k)100th percentile of a Chi-square distribution with 1 degree-of-freedom and $p_j^- = 0$ if $n_j = 0$ and $p_j^+ = 1$ if $n_j = N$.

The continuity corrected formulas for Bailey's intervals are

$$p_j^- = \frac{\left(\sqrt{p'_{(j-)}} - \sqrt{C(C + 1 - p'_{(j-)})}\right)^2}{(C + 1)^2},$$

$$p_j^+ = \frac{\left(\sqrt{p'_{(j+)}} + \sqrt{C(C + 1 - p'_{(j+)})}\right)^2}{(C + 1)^2},$$

where $p_{(j-)} = (n_j - 1/8)/(N + 1/8)$ and $p_{(j+)} = (n_j + 7/8)/(N + 1/8)$ and where $C = B/4N$ with B defined as above. The boundary conditions for Bailey's intervals are to set $p_j^- = 0$ if $n_j \leq (N + 1/8)\ C + 1/8$, and set $p_j^+ = 1$ if $n_j = N$.

An Example

Here I present results of calculations using the Byers, Bailey, and Goodman methods on the dataset in Byers et al. (1984:1052) consisting of 183 observations in 10 habitat types (Table 1). The results for the Byers$^+$ intervals are not shown. With the continuity correction equal to 0.0027 these intervals are a little wider than the Byers intervals. The Byers and Byers$^+$ intervals indicate that habitat types 2 and 5 are used less than expected and types 7, 8, and 9 are used more than expected. The Goodman and Bailey intervals lead to the same conclusion, although the avoidance of type 5 is borderline significant.

Byers et al. (1984) assumed that a sample size of 183 was large enough, but their example does not meet the standard rule-of-thumb. Two of the categories had observed use proportions of 0.011. Given a binomial proportion of 0.011, the probability of getting a 0 count is 0.132 (with $N = 183$), which is also the probability of getting a Byers confidence interval for that proportion with both lower and upper endpoints equal to 0. Thus, one could never be 95% confident that

Table 1. Simultaneous 95% confidence intervals for the Byers et al. (1984) data.

Habitat type	Proportion available	Proportion used	Byers intervals	Goodman intervals	Bailey intervals
1	0.237	0.169	(0.091, 0.247)	(0.104, 0.264)	(0.097, 0.257)
2	0.163	0.011	(0, 0.033)	(0.002, 0.065)	(0, 0.051)
3	0.113	0.082	(0.025, 0.139)	(0.039, 0.161)	(0.034, 0.152)
4	0.077	0.087	(0.028, 0.145)	(0.037, 0.159)	(0.039, 0.162)
5	0.058	0.011	(0, 0.033)	(0, 0.051)	(0, 0.054)
6	0.060	0.055	(0.008, 0.102)	(0.017, 0.116)	(0.019, 0.241)
7	0.072	0.153	(0.078, 0.228)	(0.085, 0.238)	(0.087, 0.247)
8	0.062	0.158	(0.082, 0.234)	(0.089, 0.244)	(0.091, 0.247)
9	0.073	0.158	(0.082, 0.234)	(0.089, 0.244)	(0.091, 0.247)
10	0.086	0.115	(0.049, 0.181)	(0.056, 0.193)	(0.058, 0.196)

the Byers intervals in Table 1 simultaneously contained their respective parameters. The results presented in the following section indicate how much confidence can be placed in the intervals (Table 1).

RESULTS

The simulations in this study are similar to those in Alldredge and Ratti (1986). In that paper, the authors considered sets of 4, 7, 10, and 15 habitat types, with varying availability and use proportions. The values chosen for this paper are the same as those indicated in their Table 2 (Alldredge and Ratti 1986:161), and reproduced here in Table 2. In addition, a fifth set of 10 habitat types with availability and use proportions equal to those in the example in Byers et al. (1984) was also considered. Those proportions are also given in Table 2. For each habitat type combination 1,000 simulations were generated for sample sizes of 150, 500, and 1,000 observations, and sets of the 4 simultaneous 95% confidence intervals were determined for each simulated dataset. In addition, 1,000 simulations were generated for sample sizes of 50 observations for the 4 and 7 habitat type combinations. In each case, the number of times at least 1 of the true use proportions was not contained in its interval was recorded. Also, in those cases where the use and availability proportions were different, the number of times a proportion in the available category was contained in an interval was recorded (indicating no preference or avoidance). These 2 types of errors are analogous to the Type I and Type II error rates in hypothesis testing. Finally, the average length of the intervals was determined. All simulations were performed with Splus.

Alldredge and Ratti (1986) assessed the performance of the Byers intervals based on the mistaken detection of a difference in proportional use when habitats were actually used according to their availability. There were only 4 possibilities for this type of error to occur, and that was for those categories with equal use-availability proportions (Table 2). However, these intervals should be thought of as estimators of population parameters. The method of constructing the Byers intervals supposedly ensures that they simultaneously cover the true π_i's with a specified level of confidence if the assumptions are met. The method can provide good protection against the error rate of Alldredge and Ratti (1986) while yielding simultaneous coverage

Table 2. Availability and use proportions of simulated habitat types. The combinations with 4, 7, 15, and the first set of 10 types are from Alldredge and Ratti (1986). The second set of 10 is from Byers et al. (1984).

No. of types	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
% selected	5	15	40	40											
% available	45	15	20	20											
% selected	9	20	30	8	8	20	5								
% available	3	1	6	15	15	20	40								
% selected	2	2	10	20	5	30	3	5	20	3					
% available	1	1	1	1	3	3	10	15	20	45					
% selected	16.9	1.1	8.2	8.7	1.1	5.5	15.3	15.8	15.8	11.5					
% available	23.7	16.3	11.3	7.7	5.8	6	7.2	6.2	7.3	8.6					
% selected	12	12	4	5	5	2	2	3	20	6	2	3	4	15	5
% available	3	3	1	3	12	8	3	2	20	4	1	4	6	1	29

rates that are considerably below the nominal $(1 - \alpha)100\%$ confidence level when the assumptions are violated. Such intervals are not providing good estimates of the true use proportions.

The percentage of the 1,000 simulated confidence intervals with at least 1 interval failing to capture the true use proportions is given in Table 3. The Byers and Byers⁺ intervals performed poorly with error rates that were unacceptable, particularly for the 10 and 15 habitat type categories and for small sample sizes. Note that Table 3 indicates that the error rates for the set of simulations where the Byers et al. (1984) use proportions were considered the true use proportions with $n = 150$ were 45.3 and 40.5% for the Byers and Byers⁺ intervals and 4.5 and 2.4% for the Goodman and Bailey intervals. Thus, the level of confidence that one can have in the intervals presented in Byers et al. (1984) is considerably less than 95%.

The Goodman and Bailey intervals generally had error rates lower than the nominal 5%. None of the error rates ever exceeded 6% for these latter 2 intervals. Thus, the Goodman and Bailey intervals are performing as they should, with the Bailey intervals appearing to perform best overall. All 4 of the methods performed well with respect to the error rates specified in Alldredge and Ratti (1986).The error rates for the Bailey and Goodman intervals never exceeded 1.5%, and were generally around 0.5%.

If the assumptions are met, the Goodman intervals should be shorter than the binomial intervals (Casella and Berger 1990:442). Bailey (1980) showed that his intervals tend to be shorter than the Goodman intervals. The actual lengths (not reported here) were consistent with theory. As expected, all the methods produce shorter intervals as sample sizes increase. The length of the intervals increases as the number of habitat types increases.

The percentage of times an interval contained the availability proportion (for those cases where availability and use proportions differed) was comparable for the 4 methods. The error rate decreased with increasing sample size and increased with an increase in the number of different habitat types.

Table 3. Percentage of times at least 1 of the 95% confidence intervals failed to capture the true proportions in 1,000 simulations.

No. of habitats	Sample size	Byers intervals	Byers⁺ intervals	Goodman intervals	Bailey intervals
4	50	15.3	10.5	1.9	1.9
4	150	9.1	4.8	4.2	4.4
4	500	5.3	3.6	3.8	3.7
4	1,000	4.9	4.9	3.7	4.0
7	50	33.5	21.1	2.9	2.4
7	150	8.0	8.6	2.4	2.9
7	500	5.6	4.3	4.2	2.8
7	1,000	5.6	5.0	4.2	4.4
10	150	28.3	28.3	3.8	2.9
10	500	15.4	9.9	3.7	4.1
10	1,000	8.1	6.6	4.3	3.8
10	150	45.3	40.5	4.5	2.5
10	500	12.3	8.1	3.8	4.1
10	1,000	8.0	7.3	4.1	4.1
15	150	33.7	32.8	4.5	1.7
15	500	16.9	10.2	5.5	4.5
15	1,000	9.3	8.7	3.8	4.0

DISCUSSION

The results presented here show that the Byers and Byers⁺ intervals do not perform as well as the Bailey and Goodman intervals. One obvious advantage of the binomial intervals is their simpler form and ease of use. However, with the availability of computing power and sophisticated software the more complicated Goodman and Bailey intervals can be generated easily. The Bailey intervals had the most consistent performance. The nominal 5% error rate was never exceeded by these intervals, and they did not achieve this superior performance at the cost of longer intervals.

The poor performance of the Byers and Byers⁺ intervals is surprising. Even with sample sizes of 500 and 1,000, the error rates were unacceptable for larger numbers of habitats with small proportions of use. The standard rules of thumb ($Np_i \geq$) do not work as claimed in such cases. Blyth and Still (1983) and Blyth (1986) recommended sample sizes in the range of 1,000 to 5,000 when p was close to 0 or 1 before the use of Byers and Byers⁺ could be justified for binomial proportions. For the Bailey and Goodman intervals the standard rule-of-thumb works well, and may be conservative.

As a rule, the Bailey intervals are not difficult to compute and provide the best combination of low error rates and interval length. Their use should be encouraged in wildlife habitat use-availability studies.

LITERATURE CITED

ALLDREDGE, J. R., AND J. T. RATTI. 1986. Comparison of some statistical techniques for analysis of resource selection. J. Wildl. Manage. 50:157–164.

BAILEY, B. J. R. 1980. Large sample simultaneous confidence intervals for the multinomial probabilities based on transformation of the cell frequencies. Technometrics 22:583–589.

BLYTH, C. R. 1986. Approximate binomial confidence limits. J. Am. Stat. Assoc. 81:843–855; correction 84:636.

———, AND H. A. STILL. 1983. Binomial confidence intervals. J. Am. Stat. Assoc. 78:108–116.

BYERS, C. R., R. K. STEINHORST, AND P. R. KRAUSMAN. 1984. Clarification of a technique for analysis of utilization-availability data. J. Wildl. Manage. 48:1050–1053.

CASELLA, G., AND R. L. BERGER. 1990. Statistical inference. Wadsworth and Brooks/Cole, Pacific Grove, Calif. 650pp.

GOODMAN, L. A. 1965. On simultaneous confidence intervals for multinomial proportions. Technometrics 7:247–254.

HOCHBERG, Y., AND A. C. TAMHANE. 1987. Multiple comparison procedures. John Wiley and Sons, New York, N.Y. 450pp.

JOHNSON, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. Ecology 61:65–71.

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.

QUESENBERRY, C. P., AND D. C. HURST. 1964. Large sample confidence intervals for multinomial proportions. Technometrics 6:191–195.

THOMAS, D. L., AND E. J. TAYLOR. 1990. Study designs and tests for comparing resource use and availability. J. Wildl. Manage. 54:322–330.

Received 12 May 1995.
Accepted 26 February 1996.
Associate Editor: McDonald.