SOME WILDLIFE CENSUS ESTIMATES BASED ON NON-NORMAL FREQUENCY DISTRIBUTIONS

D. Fraser
Ministry of Natural Resources
Wildlife Research Section
Box 50, Maple, Ontario L0J 1E0

and J. Barbowski
Ministry of Natural Resources
Wildlife Branch, Whitney Block
Queen's Park, Toronto, Ontario M7A 1W3

Contribution No. 77-19, Fish & Wildlife Research

Abstract: Wildlife census results often give very skewed distributions which make a poor fit to the normal distribution. The question arises whether we should use normal or non-normal methods in calculating a census estimate based on such data. In this study, four sets of sampling experiments were based on wildlife census results which made a good fit to the negative binomial or Poisson distributions. Confidence limits of the mean based on these distributions were similar in width to limits based on the normal distribution, and they contained the true population mean in an approximately equal proportion of cases. This held even at sample sizes which were below the level recommended for the normal approximation to be valid. It is concluded that the non-normal techniques provided no substantial improvement in estimating the confidence limits of the mean for a variety of positively skewed results.

Wildlife census results often have very skewed frequency distributions which make statistical treatment somewhat difficult. Studies have shown that many such results correspond closely to the negative binomial distribution, the Poisson distribution, or certain of the

various "contagious" distributions for aggregated populations (Bowden et al. 1969, McConnell and Smith 1970, Stormer et al. 1977). For a given set of results, then, the best-fitting theoretical distribution can be established by a goodness-of-fit test, and appropriate variance and other statistics can be calculated on the basis of that distribution. However, it is not clear how much precision will be gained in the estimate of the mean by using the closest theoretical distribution.

For most cases of this type, the normal approximation is adequate for estimates of the mean so long as the sample size is large (Cochran 1963:38). However the large number of enquiries we receive about non-normal analysis made us feel that many wildlife biologists remain uneasy over the application of normal statistics to very skewed distributions, and that the problem warranted some empirical study.

MATERIAL AND ANALYSIS

The study is based on four examples of wildlife surveys chosen because they have positively skewed frequency distributions which correspond particularly well to one or more of the better-known non-normal distributions.

Distribution 1 is from a fecal pellet group survey of mule deer (Odocoileus hemionus) on the Starkey Experimental Forest and Range, Oregon, carried out in 1956. The results are based on a sampling unit of 65 m^2 (700 square feet), and were originally published by McConnell and Smith (1970:34). Distribution 2 is based on a 1972 pellet group survey of white-tailed deer (Odocoileus virginianus) in two winter yard areas in North and South Canonto Townships, Ontario, with a plot size of 80.0 m^2. Both of these sets of results give a good fit to the negative binomial distribution, as is common for pellet group surveys (Bowden et al. 1969, McConnell and Smith 1970, Stormer et al. 1977).

Distributions 3 and 4 are taken from the aerial moose survey conducted in parts of northeastern Ontario in the winter of 1975-76. Preliminary testing suggested that the number of moose per 25 km^2 plot generally fits some of the "contagious" distributions, (such as the double Poisson), but the routine calculation of confidence limits from these distributions is difficult. However, the number of groups of moose per plot (with group defined as one or more individuals) generally gave a good fit to the negative binomial, while the number of moose per group gave a reasonable fit to the Poisson or the negative binomial distribution. Accordingly distribution 3 is the number of groups of moose per plot in the survey of Management Units 37 to 41, while distribution 4 is the number of moose per group in Units 27, 33 and 39. The latter group of data was chosen from a number of others because the results give similar goodness of fit to the two distributions mentioned. In distribution 4 the calculations were based on the number of moose in excess of one in each group, thus yielding the more usual series of 0, 1, 2, ... items per unit instead of 1, 2, 3, ...

Table 1 shows the observed frequency distribution for each sample, together with the expected frequencies generated by the negative binomial and/or Poisson distribution, a chi-square test of goodness of fit calculated with values combined to give expected frequencies of 5 or greater, and Fisher's measure of skewness (G_1) calculated from the observed frequencies as described by Cochran (1963:41). The degrees of freedom for the chi-square test equal three fewer than the number of ratios in the test (Bliss and Fisher 1953:183). Values of $\underline{k}$ for the negative

Table 1. Observed frequency distributions for the four examples, expected frequencies by non-normal theoretical distributions, and chi-square tests of goodness of fit. G_1 is Fisher's measure of skewness; k is the negative binomial parameter.

1. Pellet groups/plot	Observed	Negative binomial
0	466	468
1	118	109
2	34	39
3	12	16
4	5	7
5	5	3
6	2	1
7	2	1
	644	χ^2=6.164
G_1=3.238		D.F.=3
		P>0.05
		k=0.4771318

2. Pellet groups/plot	Observed	Negative binomial
0	231	235
1	159	153
2	100	100
3	72	66
4	38	43
5	32	29
6	11	19
7	10	12
8	10	8
9	4	5
10	4	4
11	2	2
12	2	2
13	2	1
14-16	4	2
	681	χ^2=7.942
G_1=2.195		D.F.=9
		P>0.50
		k=0.98073

3. Moose groups/plot	Observed	Negative binomial
0	61	61
1	71	70
2	52	56
3	42	38
4	23	24
5	11	14
6	10	8
7	4	5
8	4	2
9	1	1
10	1	1
	280	χ^2=2.017
G_1=1.216		D.F.=5
		P>0.80
		k=2.51084

4. Moose/group	Observed	Negative binomial	Poisson
1	66	74	63
2	86	72	80
3	46	44	51
4	13	21	22
5	8	9	7
6	3	3	2
7	0	1	0
8-10	3	1	0
	225	χ^2=7.037	χ^2=7.543
G_1=2.037		D.F.=3	D.F.=3
		P>0.05	P>0.05
		k=4.332	

binomial were calculated by the maximum likelihood method of Bliss and Fisher (1953:180).

In each of the four examples the observed frequency distribution was treated as a known population from which random samples were drawn with replacement by a computer. Twenty samples were drawn at each of n = 20, 40, 60, 80, 120, 240 and 480, in order to study the effect of sample size on the calculations. An additional 300 samples were drawn with $n > 25\ G_1^2$, the sample size suggested by Cochran (1963:41) as giving an approximately normal distribution of sample means, in order to examine the accuracy of the estimates in a large number of samples of uniform size. In each case the 95 percent confidence limits of the mean were calculated using the normal and non-normal statistics without the finite population correction (Cochran 1963:23).

RESULTS

The precision of the estimates of the mean (i.e., the width of the confidence limits) was related to sample size and the skewness of the parent population as shown in Fig. 1. Essentially all sets of confidence limits declined in width with increasing sample size, and were wider for the more skewed distributions. In every case the negative binomial and normal distribution statistics gave very similar results. There was some tendency for the normal confidence limits to be slightly narrower with small samples and slightly wider with large samples, but the difference rarely exceeded 1 percent of the mean averaged over the 20 cases in each set of calculations.

The accuracy of the estimates, measured as the percentage of cases in which the 95 percent confidence limits contained the true population

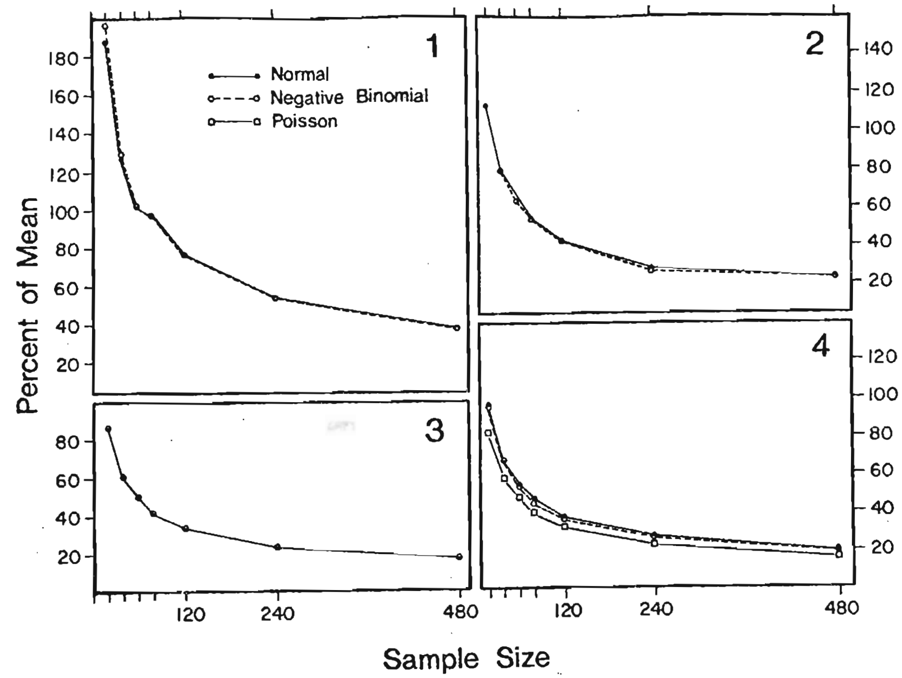

Fig. 1. The 95 percent confidence limits as a percentage of the mean, based on 20 samples at each of n=20 to 480 for the four examples. Calculations are based on the normal, negative binomial, and Poisson distributions.

mean, was almost identical for the normal and negative binomial calculations. In examples 1 to 4, for the 140 samples with n = 20 to 480, the 95 percent confidence limits contained the true mean in 126, 138, 135 and 132 of the cases by the normal calculations and in 127, 138, 134 and 133 of the cases by negative binomial calculations.

The similarity held at all sample sizes. Since n ranged from 20 to 480, each example had some estimates with n less than 25 G_1^2 and some with n larger. Combining the four examples, there were 300 estimates with n < 25 G_1^2, and the 95 percent confidence limits contained the true mean in 93.7 and 94.3 percent of cases by the normal and negative binomial calculations respectively. With n > 25 G_1^2, there were 260 estimates, and the confidence limits contained the mean in 96.2 and 95.8 percent of cases respectively.

In example 4 the Poisson calculations gave narrower confidence limits than the other two sets at every sample size, and the true population mean was in the range in 130 of the 140 cases.

The additional sets of 300 samples with n > 25 G_1^2 are summarized in Table 2. The normal and negative binomial calculations gave very similar results, with the true population mean generally falling within the 95 percent confidence limits in a little less than 95 percent of cases. The normal confidence limits were typically a little wider and contained the true mean slightly more often, but all differences were small. In example 4 the Poisson confidence limits were narrower than the other two, and contained the true mean less often. In the great majority of the errors, the true mean exceeded the upper confidence limit, especially with the normal calculations.

Table 2. Analysis of 300 random samples from each population with sample size of n > 25 G_1^2 in each case.

Example	n	Distribution	95% confidence interval as percent of mean	True mean within 95% confidence limits (percent)	True mean less than 95% c.l. (percent)	True mean greater than 95% c.l. (percent)
1	264	Normal	51.6	95.3	0.3	4.3
		Neg. binomial	50.2	94.7	1.0	4.3
2	124	Normal	45.2	91.7	1.3	7.0
		Neg. binomial	44.4	91.3	1.7	7.0
3	38	Normal	59.2	95.0	1.3	3.7
		Neg. binomial	59.4	94.3	2.0	3.7
4	104	Normal	40.2	94.0	1.3	4.7
		Neg. binomial	39.2	94.3	1.3	4.3
		Poisson	34.4	92.3	1.7	6.0

DISCUSSION

Of the four examples tested, three gave a good fit to the negative binomial and one gave a moderately good fit to both the negative binomial and Poisson distributions. Nonetheless, the normal and negative binomial confidence limits were very similar, bore about the same relation to the size and skewness of the sample, and contained the true population mean in an approximately equal percentage of cases. In the one example which warranted Poisson calculations, the confidence limits were somewhat narrower but did not contain the true mean as often, presumably because the fit to the Poisson distribution was imperfect.

These findings will come as no surprise to statisticians. The validity of the normal approximation can be shown by statistical theory. These examples simply illustrate the principle with some actual results. In addition they show, for the range of results studied, that the normal and negative binomial calculations give very similar results, even at sample sizes below the level recommended for the normal approximation to be valid.

Cochran (1963:40) points out that with positively skewed distributions the true population mean will generally fall within the 95 percent confidence limits in less than 95 percent of cases, and that sampling error will produce underestimation of the true mean more often than overestimation in the remaining small percentage of cases. Both of these observations apply to the above results. However, the true mean was within the confidence limits in only slightly less than 95 percent of cases in each of the examples.

In recent years, considerable effort has been given to fitting wildlife census results to various non-normal distributions. This

information can be used in a number of ways as summarized by McConnell and Smith (1970:29), and non-normal calculations may well improve certain types of statistical analysis. However, for straightforward census estimates based simply on the mean and its confidence interval, the familiar normal distribution statistics provide virtually as precise an estimate as calculations based on some of the best-fitting non-normal distributions, over a variety of conditions.

ACKNOWLEDGEMENTS

We are grateful to Drs. J.E. Paloheimo and F. Raymond for useful discussion, to M. Crête for valuable comments on the manuscript, to Mr. R. Hepburn for the white-tailed deer results, and to our many co-workers who collected the aerial moose survey material.

REFERENCES

Bliss, C.J. and R.A. Fisher. 1953. Fitting the negative binomial distribution to biological data and note on the efficient fitting of the negative binomial. Biometrics 9(2): 176-200.

Bowden, D.C., A.E. Anderson and D.E. Medin. 1969. Frequency distributions of mule deer fecal group counts. Journal of Wildlife Management 33(4): 895-905.

Cochran, W.G. 1963. Sampling Techniques, Second edition. John Wiley and Sons, New York. 413 pp.

McConnell, B.R. and J.G. Smith. 1970. Frequency distributions of deer and elk pellet groups. Journal of Wildlife Management 34(1): 29-36.

Stormer, F.A., T.W. Hoekstra, C.M. White and C.M. Kirkpatrick. 1977. Frequency distribution of deer pellet groups in southern Indiana. Journal of Wildlife Management 41(4): 779-782.