MICHIGAN DEPARTMENT OF CONSERVATION
Game Division

Report No. 2158
December 4, 1957

Some Uses of Stratified Sampling in Wildlife Investigations*

Most wildlife investigations, whether for management or research purposes, depend on samples of one kind or another. Since sampling is so important to the wildlife technician, it seems surprising that little use is made currently of efficient sampling methods. Perhaps one reason for this lies in what we might call the "experiment-station approach". Most of us get advice on statistical matters from people who are accustomed to testing for "significant differences", usually in an "analysis of variance" table. G. W. Snedecor's "Statistical Methods" is a standard citation in wildlife publications, and only the rare individual who penetrates to his last chapter, "Design and Analysis of Samplings" (4th edition), may become aware of the existence of a body of knowledge pertaining specifically to sample surveys.

A considerable amount of work has been done on sampling techniques for extensive surveys, and, while the purpose of this paper is to describe some uses of one method, stratified sampling, it is worthwhile here to list some standard sampling references. These are:

Sampling Techniques, 1953, W. G. Cochran, Wiley

Sample Survey Methods and Theory (two volumes) 1953, M. H. Hansen, W. N. Hurwitz, and W. G. Madow, Wiley

Sampling Methods for Censuses and Surveys, 1949, F. Yates, Hafner

Some Theory of Sampling, 1950, W. E. Deming, Wiley

The essential feature of stratified sampling (and of the several alternative methods) is the supposition that the investigator has some supplementary knowledge or information on the population being studied. In the case of a sample census, one usually knows where to expect high and low populations, and the sampling method should take advantage of this prior information. In stratified sampling one simply divides the area of interest into several levels (strata) of "expected" populations and obtains independent random samples in each stratum. Estimates in the several strata are combined to produce a single estimate for the entire area.

If an investigator clearly defines the strata in advance of the sampling, and selects sample units at random (by using a table of random numbers), he eliminates the element of personal bias. Unfortunately, many wildlife managers believe that prior knowledge is an acceptable basis for deliberate selection of

* Presented at the Nineteenth Midwest Wildlife Conference at Milwaukee, Wisconsin, December 8-11, 1957. A contribution from Federal Aid to Wildlife Restoration Funds, under Pittman-Robertson Project, Michigan. W-96-R.

"representative" samples. Stratified random sampling provides exactly the same approach, but with a virtually ironclad guarantee that errors in the investigator's judgement will not invalidate the survey results.

The advantage of stratified sampling is simply that it saves time, manpower, and money. It may well obtain results as reliable as those obtained from a simple random sample two, three, or more times larger. To my knowledge, the only real disadvantage is that people frequently confuse the selection of strata with the deliberate selection of sample units. This is a serious matter and requires some careful consideration in presenting results of a survey. One approach we use is shown in the attached figure. Here we have attempted to put across the idea that each stratum is a separate entity, providing an estimate entirely independent of results in the other strata.

In Michigan, we have used stratified sampling methods in three different types of extensive survey. These are:

(1) Estimating deer populations by pellet-group counts;
(2) Estimating deer herd mortality by counting carcasses on sample plots;
(3) Measuring public use of State Game Areas by counting cars.

In the first two cases, the sample unit was a plot or cluster of plots, and in the third, the unit was a portion of a public hunting area. In the case of the pellet-group counts, the primary sampling unit was defined as a square mile in which we searched a "cluster" or "course" of 8 plots.

I believe stratified sampling can be successfully applied in investigations ranging from counts of crowing pheasants to mail questionnaires. One difficulty, of course, is to define useful strata, and this can be more expensive than taking a simple random sample. As an example, we used a kind of geographic stratification for our mail surveys to estimate Michigan's game kill in 1954. The apparent precision of the results was little different from that of a simple systematic sample of all licensees, but the cost was considerably greater. However, there was a distinct advantage in that we ordinarily must wait for all license agents to return duplicate copies of licenses before we can complete the survey. In this case we were able to obtain a sample fairly rapidly by direct contact of selected license agents.

The user of stratified sampling methods must necessarily reach some decision as to the number of strata. The references cited above provide a number of useful suggestions on forming strata which need not be discussed here, except to say that our practice has been to use from 3 to 5 strata in most cases. Stratification in some of our extensive surveys has been based on the recommendations of a dozen or more people, so a major problem has been to define strata which mean essentially the same thing to each individual. A "high" deer population in one man's area may well be "medium" or "low" in some other district. Our procedure has been to set up several arbitrary classifications and ask each field-man to prepare maps on this basis. A meeting is then held to compare the maps from various areas and the number and range of strata may be adjusted as seems necessary. If a sufficiently large sample is to be taken, one can, of course, use several strata in each of a number of administrative or management units.

A further problem in designing a stratified sampling plan concerns the allocation of sample units to strata. That is, suppose one is to conduct a sample survey involving counts on sample plots. When the strata have been selected and delineated on suitable maps, one must decide how many plots to use and how they should be distributed or "allocated" among the several strata. The desirable total sample size will necessarily depend on some advance knowledge (or estimate) of the variation likely to be encountered in the survey, but often it will depend mainly on the available man-power.

Two methods of allocation are commonly used. These are "proportional" allocation in which the sample is distributed among the strata in proportion to the area of each stratum, and "optimum" allocation in which the sample is allocated on the basis of both area and variability expected in each stratum ("area" may also mean "number of possible sampling units"). Proportional allocation is simplest to use and explain, but the "optimum" approach seems to me worth attempting in wildlife investigations because we usually have very different variances in the several strata.

If results of a previous extensive survey are available for the area of interest, one can usually extract satisfactory variance estimates for optimum allocation. In the usual situation, the available data frequently pertain approximately to one stratum, often that of the highest expected mean. In this case, or in the case where only the expected average values for the strata are available, the above-mentioned references suggest an allocation based on a constant coefficient of variation, i.e., that the stratum standard deviations be regarded as proportional to the stratum means. If the available data pertain to one of the higher strata, this suggestion may, in wildlife investigations, lead to undersampling the lower strata since our experience has been that the coefficient of variation in lower strata (lower means) is usually larger than that in higher strata.

In pellet-group counts it seems reasonable to use a standard deviation (s) approximately equal to the expected stratum mean (m) for allocation in strata where the expected mean is 10 pellet groups or more per sample unit. In strata having expected means of less than 10 pellet groups, the following approximation is suggested:

$$s = m \left(\frac{2m + 1}{m}\right)^{\frac{1}{2}}$$

The data on which this approximation is based are from counts of pellet-groups on sample units of eight 1/50 acre plots. Probably the approximation will apply to sample units of different sizes but we have not tested this assumption. The approximation is based on the equation for variance of a Negative Binomial distribution:

$$s^2 = m + \frac{m^2}{k}$$

We estimated the "index", k, of the distribution from the above relationship for 23 different pellet-group count strata. The average value of k for these 23 strata is very close to unity, giving the approximation:

$$s = m\left(\frac{m + 1}{m}\right)^{\frac{1}{2}} \doteq m$$

However, six out of the seven strata having means of less than 10 pellet-groups per sample unit yielded values of k very nearly equal to $\frac{1}{2}$, so we suggest the first approximation for protection against undersampling the lower strata.

In the case of dead deer counts, two different situations are encountered. In our experience, when mortality is due to causes other than starvation, the approximation $s = m^{\frac{1}{2}}$ (Poisson distribution) seems appropriate. When starvation losses occur, we find stratum standard deviations are larger than $m^{\frac{1}{2}}$.

An average value of k = 2 as obtained from 14 different dead deer search strata gives the approximation:

$$s = m\left(\frac{m + 2}{2\ m}\right)^{\frac{1}{2}}$$

I do not recommend the above approximations as being at all precise. They may serve to give some notion of what to expect if no other data are available. In most wildlife investigations, I believe one can get better estimates by allocating samples on the assumption of stratum standard deviations proportional to the means rather than by simply allocating proportional to the area (or sample units) in the stratum. Some further imporvement seems likely if the sampling rate in the lower strata is increased by an approximation such as is given above or by arbitrary addition of a few sample units.

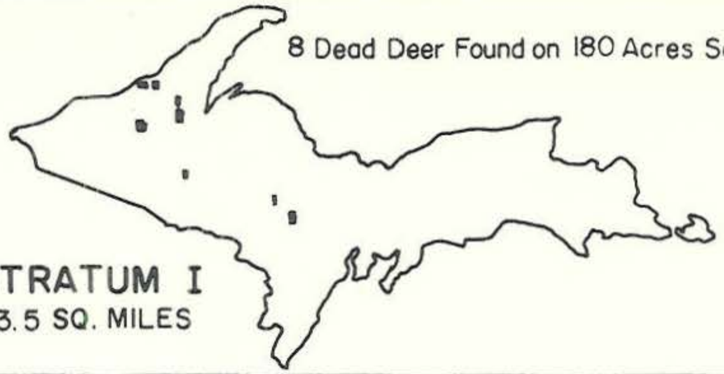
In summary, stratified sampling methods may permit a very considerable reduction of effort required in various wildlife investigations. The statistical techniques of survey sampling are rather different from those generally encountered by the wildlife technician, but the gains from using such methods certainly justify the expenditure of a great deal of effort in the sampling designs.

Lee Eberhardt

LLE/nlp

RESULTS OF UPPER PENINSULA DEAD DEER SURVEY - 1956

8 Dead Deer Found on 180 Acres Searched = 28.4 per Sq. Mile

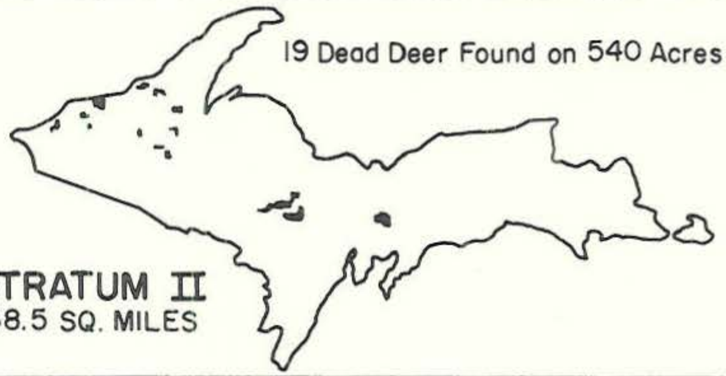


28.4 x 33.5 Sq. Mi. = 950

STRATUM I
33.5 SQ. MILES

---

19 Dead Deer Found on 540 Acres Searched = 22.5 per Sq. Mile

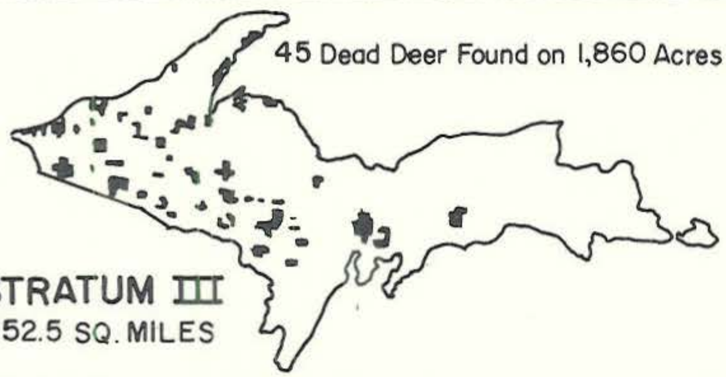


22.5 x 138.5 Sq. Mi. = 3,100

STRATUM II
138.5 SQ. MILES

---

45 Dead Deer Found on 1,860 Acres Searched = 15.5 per Sq. Mile

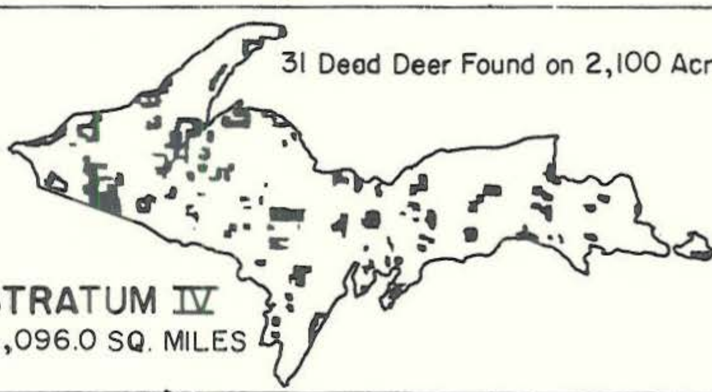


15.5 x 752.5 Sq. Mi. = 11,650

STRATUM III
752.5 SQ. MILES

---

31 Dead Deer Found on 2,100 Acres Searched = 9.4 per Sq. Mile

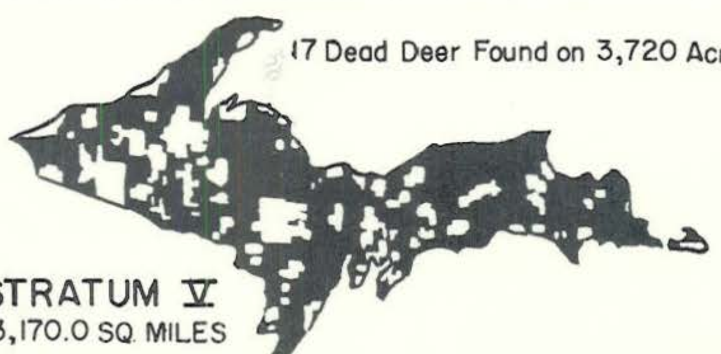


9.4 x 2,096.0 Sq. Mi. = 19,800

STRATUM IV
2,096.0 SQ. MILES

---

17 Dead Deer Found on 3,720 Acres Searched = 2.9 per Sq. Mile



2.9 x 13,170.0 Sq. Mi. = 38,500

STRATUM V
13,170.0 SQ. MILES

TOTAL DEAD DEER in the UPPER PENINSULA 74,000