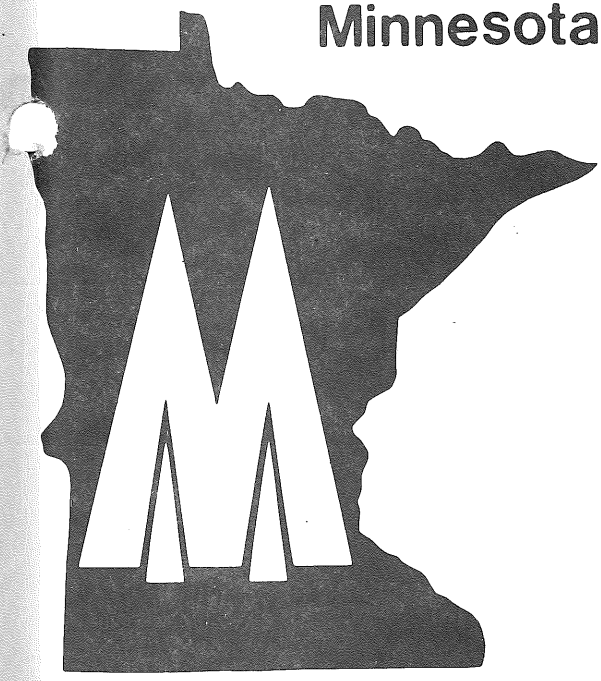

DEPARTMENT OF
NATURAL RESOURCES

Telephone Survey Sampling Methods

SCORP

Research and Policy Section
of Comprehensive Planning & Progamming

TELEPHONE SURVEY SAMPLING METHODS

Compiled

for

The Minnesota Department of Natueal Resources

Bureau of Planning

St. Paul, Minnesota

by

C.R. Michael Parent

Institute of Outdoor Recreation and Tourism

Utah State University

Logan, Utah

October, 1977

Report Number 2303

This project was carried out with funds provided by

the Legislative Commission on Minnesota Resources.

Sampling Methods

Introduction

The Minnesota SCORP planning process requires the collection of extensive primary data on both winter and summer outdoor recreation activities by the residents of different regions in Minnesota. This section of the pretest report evaluates the pros and cons of various sampling methods as they relate to sample size. The major factors in this regard are: (1) the constraints posed by the Minnesota Department of Natural Resources (DNR) data requirements for calculating and forecasting regional person participation occasions by activity, county of origin and region of participation; (2) the DNR's required precision and tolerance for error; (3) the sampling method i.e., techniques for collecting data; (4) the variability of the population as estimated by the pretest; (5) the relative importance of the various activities as determined by the political process, total participation rates, or the resource requirement, hence a cost value assessment of the data; (6) the availability of a frame from which to select potential respondents to the questionnaire; and (7) our interest in selecting a conservative sample size (one which is large relative to the minimum size required) such that meaningful evaluation of subsets of data may be pursued.

These major factors which influence the sampling process, hence a determination of sample size, are not mutually exclusive. Some are qualitative while others are quantitative. However, in the sections which follow all relevant constraints and factors are examined individually and as objectively as possible, assumptions are specified and recommendations are

offered regarding factors influencing the most appropriate sample sizes.

Census versus sample

In the best of all possible worlds one can collect data on each element of a population (census) and absent any errors in reporting the data have an accurate and complete description of some population of interest. However, there are times when taking a census is impractical or impossible. Generally, these reasons include considerations of cost, time, accuracy, and the destructive nature of the measurement. Hence, a sample is warranted.

IN the DNR case, considerations of cost, time and accuracy warrant the use and selection of a sample. Consideration of cost and time are obvious; however, it is also quite likely that the quality of data would suffer if a census were taken. It is a well documented fact in the marketing and survey research literature that as sample size increases, sampling error decreases, but non-sampling (particularly interviewer caused) error increases. In fact, where the data complexibility dictates a difficult interviewing process, sample sizes may need to be smaller than would otherwise be dictated by reason of concern for sampling error alone. Of course, this is a subjective decision made by the researcher.

Given that a sample is warranted, one must choose among a variety of sampling techniques: (1) probability versus non-probability; (2) single unit versus a cluster of units; (3) unstratified versus stratified; (4) equal unit probability versus unequal probability; and (5) single stage versus multistage (Tull and Hawkins, p. 157). Our preference is a sampling process which would be a probability sample (where sampling

units are selected by chance and each has a known chance of being selected) consisting of single units (each of which would have an equal chance of inclusion in the sample) drawn in an unstratified manner (from the general population not segments having some common characteristic) in a single stage (that is drawing only one sample, notwithstanding the pretest, for each region). This preference is based on several factors: (1) many techniques of statistical analysis are based upon simple random samples and while these techniques are robust to a certain extent the conservative position is to draw a simple random sample, (2) the more complex sample designs may increase the possibility that nonsampling error will increase as will economic inefficiency with no corresponding gain in reduction of sampling error, and (3) also, there are not insignificant political considerations. That is results based upon probability samples, of which the simple random sample is best known, may be more acceptable than the non-probability alternative. The research problem then is: Can such a sample be drawn given the DNR requirements and data characteristics?

Simple random sampling versus cluster sampling

At least two problems exist for the researcher who attempts to specify a sampling procedure which is universally acceptable to all concerned: (1) there are at least two popular definitions of what constitutes a simple random sample; albeit, one is more technically correct than the other, and (2) the problems of specifying a sampling frame consistent with the ultimate objectives of the study and absent any serious deficiencies.

Simple random sampling appears to be defined from two different perspectives: micro and macro. The micro or applied view is of a process

where each element has an equal probability of being selected from an unstratified population by a single stage procedure (Tull and Hawkins, p. 159). Additionally, the simple random sample is defined as a special case probability sample where each element in the frame has an equal chance of being selected (Lansing and Morgan, p. 64-71).

While the macro (theoretical) view of the simple random sampling process does not contradict the preceeding definitions, it is more restrictive. For any given sample of size "n" from a population of size "N" every possible sample of size "n" must have the same chance of being selected (Mendenhall, et al., p. 31). At first glance, the micro and macro views may seem equal. There is a difference if the sample frame is such that some samples cannot be drawn. The DNR study presents just such a problem. If there is not an enumeration (list) of all individuals over six years of age in the population, then it is difficult to construct a frame which will allow every possible sample size "n" to have an equal chance of being selected. Such a frame would clearly be acceptable and consistent with both definitions of simple random sampling.

A second type of frame provides only an intervening set of elements so that one may associate the elements in the population with the elements in the set (Lansing and Morgan, p. 65). This is precisely the type of frame provided by a phone book, notwithstanding the traditional concerns of missing elements, foreign elements, and duplicate listings and the various types of bias they may introduce. (This is a question which will be examined separately.) If we draw a random sample from a list of phone numbers representing households and solicit information on activities from each member of the household, then all of the individuals sampled will have

had an equal, known and non zero chance of being selected in our sample. Their probability of inclusion in the sample will be equal to the probability that the phone number would be selected; hence, under our first definition, a simple random sample. Yet, not all possible samples of individuals would be achievable. Some combinations of individuals would not be possible. For example, the second of four persons represented by one phone number could not be included in a sample with the third of five persons represented by another phone number without the appearance in the sample of the other seven persons.

As the macro definition is consistent with the evolution of sampling theory, upon which many statistical techniques are based, the construction of a sampling distribution of the means based upon all possible samples of size "n" having an equal chance of selection which follows a normal distribution, the macro definition is certainly more correct. It is not unreasonable to accept the less technically correct definition or at least understand its use given the situation faced by most practitioners of selecting one and only one sample. In fact, it is difficult to find a textbook on sampling, marketing or survey research which does not offer an example or problem purported to be a simple random sample which estimates a population parameter from a household sample. The reason is: the one sampling chance the practitioner gets must be one of the possible samples with each element equally possible among among the possible samples.

If we do not accept the micro definition of simple random sampling, what kind of sample do we have if the sample is drawn from a list of phone numbers and estimats of individual activity patterns are made? As the sampling units (individuals) are enumerated by the intervening groups

(households with telephones), the sample may be considered a cluster sample.

Cluster sampling is a necessay evil which is undertaken usually where consideration of cost is important (Lansing and Morgan, p. 71). Cluster sampling is also useful where no primary list of individuals in the population is available. Most examples of cluster sampling refer to the cluster in geographic terms. A city block, voting district, counties or states may be considered a cluster. Cluster sampling produces no bias. For equivalently sized samples, it may increase sampling error vis-a-vis a simple random sample; it cannot reduce sampling error. The effect of sampling error attributable to cluser sampling can be evaluated against the simple random sample by comparing the standard errors of the means of the two sampling approaches (Lansing and Morgan, p. 71, 72). Essentially, the relationship is a function of:

$$1 - \frac{\text{within-cluster variance}}{\text{total variance}} = \text{Roh}$$

such that it varies between 1 (where all the variance is between clusters and none is within clusters, hence the second term is zero) and zero (where the second term is one). Roh is multiplied by the number of elements in a cluster minus one (B-1) to estimate the relationship between the standard error of the mean of a simple random sample and the standard error of the mean of a cluster sample. For example, cluster sampling with a cluster size of 3 with all the variance between clusters and none within the clusters would, for equivalent sized samples, produce three times the standard error of a simple random sample:

$$\frac{\text{var } \bar{y}}{\text{var } \bar{y}_o} = \frac{Sa^2/a}{Sn^2/n} = DE = (1 + roh(B-1))$$

or

$$\text{var } \bar{y} = \text{var } \bar{y}_o \ (1 + roh(B-1))$$

where: $\bar{y}_o$ = mean of some variate y estimated from a simple random sample

a = number of clusters

n = number of interviews in sample

DE = design effect

B = number of elements in a cluster

roh = coefficient of interclass correlation hence,

$1 + ((1-0)(B-1)$

$1 + ((1)(3-1))$

$1 + (2) = 3 = DE$

(See Lansing and Morgan p. 70-99 for more complete discussion).

Therefore, to be efficient, a cluster sample should be designed with a small cluster size (measured by the number of elements per cluster) and hopefully a variance within each cluster which approximates the variance in the population.

A telephone sample of Minnesota residents would satisfy the preceeding requirements. For example, the average cluster produced by our pretest combined for the two regions was just over 2.6 persons per phone. The 1970 census report for Minnesota places the average family size at 3.2. However, because the pretest and final study only measure recreators over six years of age, the results of the pretest are probably an accurate

estimate of the size of household with phones excluding children under six years of age. After calculating the within cluster variance and the total variance for the pretest for any particular activity, we can ascertain the effect of clustering on the standard error of the mean. The relationship or design effect (DE) will provide an estimate of the relative sample sizes (simple random sample vs. cluster sample) to produce equivalent levels of precision. Calculations of the design effect and sample sizes will be pretested in a subsequent section.

Variability in the population and its influence on sample size

All other factors notwithstanding, e.g., level of acceptable error, tolerance, type of sample, and data requirements, the sample size is influenced by the variability of the population from which the sample is drawn. To gain an estimate of some population parameter expressed as a mean requires a larger sample when the population is quite variable than when the population is closely grouped about some characteristic. This is a fundamental characteristic of the sampling process regardless of sampling technique or the objective (variable) of interest. It holds for cluster samples, stratified samples, simple random samples and sampling to estimate average income, education, age or recreation activity occasions.

Where they are multiple question questionnaires, the largest sample size will be estimated using the question with the largest variability. However, where cost is a potential limiting factor or some information is of higher importance to the user, the most important question may be used to calculate the sample size. As the Minnesota DNR requires planning

data from respondents on several different activities and for individuals with differing socio-economic characteristics, the questionnaire solicites data with a number of questions on a variety of factors.

We examined the participation rates of respondents for the various activity types. Also, we undertook a qualitative review of the various activity types to isolate those or one which would require a substantial allocation of resources for acquisition and development considering the level of use by Minnesota residents. The purpose of this exercise was to determine a cost/value relationship for the DNR data requirements. We found fishing, picnicking, and camping to be activity types that were frequently engaged in by the respondents to the pretest. Additionally, we found a significant range in the number of activity occasions among the respondents for these activities.

It was reasoned that these activities would require a significant allocation of resources by the DNR to continue to meet resident needs for recreation. Hence, considering the cost of the sample vs. value of the data, and in terms of our estimate of population variability, we could justify a sample size dictated by any of these three activity variables. The socio-economic variables income, education, etc. also have large variances; however, precise estimates of the socio-economic variables are not required. For the purpose of this study wide ranges in these variables are sufficient for studying relationships between these demographic characteristics of the population and the activity occasions.

In the following analysis of sample size, we chose to concentrate on the camping activity for several reasons. Camping, unlike other activity types, does not have an adequate substitute of another activity type at the time the activity is sought by a recreator. The camping activity

certainly requires among the largest site development resource allocation. Since it is an over-night activity, peak use can't be shifted to other times of the day as in the picnicking activity. And, finally, due to the questionnaire length and information requirements, a mail questionnaire is contemplated for gathering data on fishing; hence, fishing will be covered adequately and in more depth than camping. In our pretest we found that approximately 11.1 percent of the sample reported camping with a range of 0-4 activities. As camping was treated as a 24-hour participation period, the total hours committed to this activity type are high. In the following section, data from the pretest are used to calculate the sample size based upon the variability of the pretest sample on the issue of camping.

Calculation of sample size

We proceed to calculate the required sample size by making several decisions and a few assumptions. First, each telephone number is considered a cluster. Based upon our pretest, the Minnesota 1970 census data and our interest in members of the household over 6 years of age, we assume the average cluster size to be approximately 2.6 Also, we use the pretest to estimate the population variance. The estimate of σ_c^2, the cluster population variance, is s_c^2 or .32. And, finally, we wish to place a bound (B) about the estimate representing the maximum error we are willing to accept. In this case, we assume the DNR requires a 95 percent confidence level. Therefore, the approximate sample size required to estimate "μ" with a bound, B, on the error of estimation is found by:

$$n = \frac{N\sigma_c^2}{ND+\sigma_c^2}$$

where σ_c^2 is estimated by s_c^2

and,

$$D = \frac{B^2\bar{M}^2}{4}$$

and,

$$B = 2\sqrt{V_{(\bar{y})}} = .03$$

where: N = the number of clusters in the population

$\bar{M}$ = average cluster size

n = number of clusters in a sample by a simple random sample.

$$B = .03$$

$$D = \frac{(.03)^2\ (2.6)^2}{4} = \frac{(.0009)\ (.6.76)}{4} = .00152$$

$$n = \frac{567{,}000\ (.32)}{567{,}000\ (.00152) + .32} = \frac{181{,}400}{862.16}$$

$$n = 210.4$$

(Mendenhall, et al., pp. 121-136)

We can also calculate the design effect to determine the relative efficiency of the cluster sample vs. a simple random sample of the same size as previously discussed. In this case:

$$1 - \frac{\text{within cluster variance}}{\text{total variance}} = \text{roh}$$

$$1 - \frac{.11}{.14} = \text{roh}$$

$$1 - .79 = \text{roh}$$

$$.21 = \text{roh}$$

Design Effect = 1 + roh (B-1)

Design Effect = 1 + .21 (2.6-1)

Design Effect = 1.34

Consequently, we can conclude, as a result of the relationship between the within cluster to total cluster variance and average size of each cluster, for equivalent sample sizes the cluster sample is less efficient than simple random sampling. It produces approximately 34 percent more sampling error than a simple random sample would.

We can also calculate the sample size for a simple random sample for further comparison and evaluation of the sample size issue. We can estimate σ^2 by taking one fourth of the range (Mendenhall, et al., p. 41). Hence, $\sigma^2 \approx (.25)^2 = .0625$.

Therefore, we can substitute into the equation:

$$n = \frac{N\sigma^2}{(N-1)D+\sigma^2}$$ (Mendenhall, et al. p. 40)

where,

$$D = \frac{B^2}{4}$$

or,

$$D = \frac{(.03)^2}{4} = .000225$$

and,

$$n = \frac{1{,}814{,}000(.0625)}{1{,}813{,}999(.000225)+.0625} = \frac{113{,}375}{408.2123}$$

$$n = 277.74$$

To recapitulate, we have chosen an activity which is of significant interest to the Minnesota DNR, camping. From the pretest results, we have estimated σ^2 by calculating s^2; substituting .03 as a bound about our estimate of $\bar{y}$ assumes a .95 confidence interval. Treating the sample as a cluster sample, drawing each cluster using a simple random sample, we calculated the sample size necessary to estimate the mean of the population. It should be noted that a cluster sample of 210 would result in approximately 546 total responses (210 times 2.6 elements per cluster).

The preceeding calculations were based upon data from the Minneapolis-St. Paul region. Sample size may change from region to region; hence, we calculated the sample size for the Rochester region as follows:

$$n = \frac{N\sigma_c^2}{ND+\sigma_c^2} = \frac{(26,283)(.32)}{(26,283)(.00152)+.32} = 208.9$$

If the variability (σ^2) of the population in any region is greater than .32, then a larger sample size would be dictated and vice versa. Or, if the region is a key user of recreation facilities, then even though σ^2 does not differ from other regions, one could justify a large sample for that region. However, one should note that the number of clusters, all other things being equal, does not influence sample size.

At this point, we should note, again, our preference for large versus small samples consistant with considerations of cost. Also, as the participation rates within the population are quite small for some activities based upon the results of our pretest, larger sample sizes, sample sizes greater than 210 clusters per region, are suggested. Further, when pursuing the winter questionnaire and subsequent summer questionnaire, we suggest making an estimate of σ^2 from the first week's results for certain key variables to confirm the sample size per region necessary for making estimates regarding population parameters.

Telephone directory vs. random number sample selection

In the recent survey and marketing research literature much has been written about random digit dialing (RDD) as a means of reducing frame

error for phone sampling. However, our experience over time, substantiated by the pretest, as reported in this analysis, indicated the RDD process would be time consuming and costly relative to random selection from the phone books. The issue at hand is whether the benefits of RDD are worth the added cost.

As indicated in the results section (Table 2, page 11), there were problems with random number dialing even when known residential exchanges were used. Random dialing resulted in a 28 percent contact rate while numbers from telephone directories resulted in a 66.4 percent rate of contact, thus, the directory selection method was shown to be least costly in terms of calls and more convenient. The question has been posed though, as to the reliability of telephone directory samples in terms of representation of the entire population.

There is evidence which indicates this is not a problem in Minnesota. Bell Telephone Company statistics place Minnesota near the top of the list of states in terms of the percentage of households with telephones with a rate of more than 98 percent. The percentage of households with telephones in Minnesota is equal to or greater than 44 other states. In a paper entitled "Is Random Digit Dialing Really Necessary?" published in the August, 1977 volume of the Journal of Marketing Research[a], Clyde L. Rich, Supervising Statistician of the Division of Market Research of the Pacific Telephone company states:

---

[a] Rich, C. L., "Is Random Digit Dialing Really Necessary?" Journal of Marketing Research, Vol XIV, August 1977, Chicago, Illinois, pp. 300-305.

> There are significant differences in some demographic characteristics between published and unpublished[b] telephone subscribers. However, because many of the differences are much smaller than the pub population, samples drawn from telephone directories have virtually the same demographic characteristics as samples which include nonpubs.

Rich used the Los Angeles and San Francisco Metropolitan Regions as well as the entire Pacific Telephone System as his test cases where the percentage of non-pubs is the highest in the U.S. Even in this region containing two large metropolitan areas where there was an average of over 27 percent nonpublished numbers did his findings hold true. In the Minnesota metropolitan areas of Minneapolis, St. Paul, Duluth, and Rochester where there are only 9.5 percent unpublished numbers[c] a stronger case could be built for the accuracy of samples drawn from telephone directories. The percentage of unpublished numbers is even less for the remainder of Minnesota.

Rich's data serve to dispell other accepted ideas. He found that large families, not small families, are more likely to have nonpublished numbers, clerical/service workers are more likely to have nonpublished numbers than professionals, and low income families are more likely to have nonpublished numbers than high income families.

---

[b] Unlisted, unpublished, and newly connected telephone numbers.

[c] Northwestern Bell Telphone Company, Minneapolis, Minnesota.

Literature Cited

Lansing, J. B. and J. M. Morgan. Economic Survey Methods. Survey Research Center, University of Michigan, Ann Arbor, 1974.

Mendenhall, W., L. Ott, and R. Schaffer. Elementary Survey Sampling. Wadsworth Publishing Company, Belmont, Ca., 1971.

Rich, C. L. "Is Random Digit Dialing Really Necessary?" Journal of Marketing Research, Vol. XIV, August 1977, Chicago, Illinois, pp. 300-305.

Tull, D. and D. Hawkins. Marketing Research. Macmillan and Company, New York, 1976.