Wiley StatsRef:
Statistics Reference Online

Survey Sampling Procedures

By *Kris K. Moore*

Keywords: *population, sample, frame, questionnaire design, sampling methods, sample size*

Abstract: Survey sampling provides procedures to estimate characteristics of a population based on a sample. Questionnaire design, collection of sample information, sampling methods, and sample size are discussed.

Survey sampling provides a method of obtaining information about a population based on a sample selected from the population. A population is the totality of elements or individuals about which one wants to obtain information. A sample is a portion of the population that has been selected randomly. A sample characteristic is used to estimate the population characteristic. The cost of surveying all the elements in a population can be high, and the time for surveying each member of the population can be enormous. Sampling can provide a savings in cost and time.

Parameters or characteristics of the population are estimated from statistics that are characteristics of the sample. The estimates may differ by only ±2.5% from the actual population values. For example, in 2004, public opinion polls in the United States obtained information about approximately 300 million people from a sample of 1500 people. If in the sample of 1500, say, 600 favor the president's performance, the sample statistic is $600/1500 = 0.4$. The statistic 0.4 or 40% differs by only 2.5% from the population parameter 19 out of 20 times a sample of this size is taken. Thus, this sample would indicate that the interval 37.5 to 42.5% estimates the lower and upper bound of an interval that contains the percentage who favor the president's performance. About 95 out of 100 times a sample of this size (1500) would produce an interval that contains the population value.

Instrument of Survey Sampling

A questionnaire is usually designed to obtain information about a population from sampled values (*see* **Survey Questionnaire Design**). The questionnaire should be as brief as possible, preferably not more than two or three pages. The questionnaire should have a sponsor that is well known and respected by the sampled individuals. For example, if the population is a group of teachers, the teachers will be more likely to respond if the survey is sponsored by a recognizable and respected teacher organization. Also,

Baylor University, Waco, TX, USA

This article was originally published online in 2005 in Encyclopedia of Statistics in Behavioral Science, © John Wiley & Sons, Ltd and republished in Wiley StatsRef: Statistics Reference Online, 2014.

Copyright © 2005 John Wiley & Sons, Ltd. All rights reserved.

money incentives such as 50¢or a $1.00 included with the mailing improve the response rate on mailed questionnaires.

The questionnaire should begin with easy-to-answer questions; more difficult questions should be placed toward the end of the questionnaire. Demographic questions should be placed toward the beginning of the questionnaire because these questions are easy to answer. Care should be taken in constructing questions so that the questions are not offensive. For example rather than asking, 'How old are you' a choice of appropriate categories is less offensive. For example, 'Are you 20 or younger?' 'Are you 20 to 30?' and so on is less offensive. A few open-ended questions should be included to allow the respondent to clearly describe his or her position on issues not included in the body of the questionnaire.

Questionnaires should be tested in a small pilot survey before the survey is implemented to be sure the desired information is selected and the questions are easily understood. The pilot survey should include about 30 respondents.

Type of Question

Again, some open-ended questions should be used to allow the respondent freedom in defining his or her concerns. The Likert scale of measurement should be used on the majority of the questions. For example,

Strongly disagree	Mildly disagree	Disagree	Neutral	Agree	Mildly agree	Strongly agree
1	2	3	4	5	6	7

The Likert scale is easily converted to a form that computers can process.

The Frame

The sample in survey sampling is randomly selected from a frame or list of elements in the population. A random selection means each element of the population has an equal chance of being selected. Numbers can be assigned to each member of the population 1, 2, . . ., N, where N is the total number in the population. Then random numbers can be used to select the sample elements.

Often a list of the population elements does not exist. Other methods will work for this type of problem and are discussed in the Selection Methods section.

Collection of the Sample Information

The basic methods of obtaining responses in survey sampling are personal interview, mail questionnaire, **telephone interview**, or electronic responses via computers. **Personal interviews** are accurate but very expensive and difficult if the population includes a large geographic area. The person conducting the interview must be careful to be neutral and must not solicit preferred responses.

Mail questionnaires (*see* **Mail Surveys**) are inexpensive but the response rate is usually low, sometimes less than 10%. Incentives such as enclosing a dollar or offering the chance to win a prize increase response rates to 20 to 30%. Additional responses can be obtained by second and third mailings to those who failed to respond. Responses from first, second, and third mailings should be compared to see if trends are evident from one mailing to the other. For example, people who feel strongly about an issue may be more likely to respond to the first mailing than people who are neutral about an issue.

Copyright © 2005 John Wiley & Sons, Ltd. All rights reserved.

Telephone interviews are becoming more difficult to obtain because the general public has grown tired of tele-marketing and the aggravation of telephone calls at inconvenient times. In the United States, approximately 95% of the general population has telephones and of these, 10 to 15% of telephone numbers are unlisted. Random digit dialing allows unlisted numbers to be included in the sample by using the prefix for an area to be sampled, for example 756-XXXX. The XXXX is a four digit random number that is selected from a list of random numbers. This procedure randomly selects telephone numbers in 756 exchange.

Electronic methods of sampling are the most recently developed procedures (*see* **Internet Research Methods**). Internet users are sampled and incentives are used to produce a high response rate. Samples are easily selected at low cost via the computer.

Selection Methods

The **simple random sample** in which each element has an equal chance of selection is the most frequently used selection method when a frame or list of elements exists (see [2]). A systematic sample in which every kth element is selected is often easier to obtain than a random sample. If N is the number in the population and n the sample size, then $k = N/n$, where k is rounded to the nearest whole number. The starting point 1 to k is randomly selected. If the sampled elements are increasing in magnitude, for example, inventory ordered by value from low-cost items to high-cost items, a systematic sample is better than a random sample. If the elements to be sampled are periodic, for example sales Monday through Saturday, then a random sample is better than a systematic sample because a systematic sample could select the same day each time. If the population is completely random, then systematic and random sample produce equivalent results.

If a population can be divided into groups of similar elements called strata, then a stratified random sample is appropriate (*see* **Stratification**). Random samples are selected from each stratum, which insures that the diversity of the population is represented in the sample and an estimate is also obtained for each stratum.

If no frame exists, a **cluster sample** is possible. For example, to estimate the number of deer on 100 acres, the 100 acres can be divided into one-acre plots on a map. Then a random sample of the one-acre plots can be selected and the number of deer counted for each selected plot. Suppose five acres are selected and a total of 10 deer are found. Then the estimate for the 100 acres would be 2 deer per acre and 200 for the 100 acres.

Sample Size

An approximate estimate of the sample size can be determined from the following two equations (see [1]). If you are estimating an average value for a quantitative variable, (1) can be used.

$$n = \left(\frac{2\sigma}{B}\right)^2, \tag{1}$$

where n is the sample size, σ is the population standard deviation, and B is the bound on the error. The bound on the error is the maximum differences between the true value and the stated value with probability. 9544. An estimate of σ may be obtained in three ways: (a) estimate σ from historical studies, (b) estimate σ from (range of values /6) $\cong\sigma$, and (c) obtain a pilot sample of 30 or more elements and estimate σ by calculating $\hat{\sigma}$ (the sample standard deviation), (see 2).

Copyright © 2005 John Wiley & Sons, Ltd. All rights reserved.

For example, if you wanted to estimate miles per gallon, mpg, for a large population of automobiles within 2 mpg, then B would equal 2. A pilot sample can be taken and $\hat{\sigma}$ is used to estimate σ where $\hat{\sigma}$ is found from equation

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}}, \tag{2}$$

where X_i is the sample value and $\overline{X}$ is the sample mean. For the above example, suppose $\sigma \cong 4$. Then using (1)

$$n = \left(\frac{2(4)}{2}\right)^2 = 16. \tag{3}$$

A sample of size 16 would produce a sample mean mpg that differs from the population mean mpg by 2 mpg or less.

If a population proportion (percentage) is to be estimated, the sample size is found from (4).

$$n = \left(\frac{2}{B}\right)^2 p(1-p), \tag{4}$$

where B is the bound on the error of the estimate, n is the sample size, and, p is the population proportion. The population proportion can be estimated three ways: (a) use historical values of p, (b) obtain a pilot sample of 30 or more elements and estimate p, and (c) use $p = .5$ because this produces the widest interval.

An example of determining the sample size necessary to estimate the population proportion is given below when no information exists about the value of p. Suppose an estimate of the proportion of voters that favor a candidate is to be estimated within $\pm 3\%$ points. If there is no information from previous research, we select $p = .5$. Then $B = .03$ and n is determined from (4)

$$n = \left(\frac{2}{.03}\right)^2 .5.5 = 1111.11 \text{ or } 1112. \tag{5}$$

Consider another example. If a low percentage is to be estimated like the proportion of defective light bulbs that is known to be 5% or less, then use $p = .05$ to estimate the proportion of defectives. If we want to estimate the percentage within $\pm 1\%$, we use $B = .01$. Then from (4)

$$n = \left(\frac{2}{.01}\right)^2 .05(.95) = 1900. \tag{6}$$

In summary, survey sampling procedures allow estimates of characteristics of populations (parameters) from characteristics of a sample (statistics). The procedure of survey sampling saves time and cost, and the accuracy of estimates is relatively high.

Related Articles

Survey Sampling
- **Survey Questionnaire Design**
- **Survey Quality and Survey Ethics**
- **Survey Error**

Copyright © 2005 John Wiley & Sons, Ltd. All rights reserved.

References

[1] Keller, G. & Warrack, B. (**2003**). *Statistics for Management and Economics*, 6th edition, Thomson Learning Academic Resource Center.
[2] Scheaffer, R. L., Mendenhall, I. I. I. W. & Ott, L. O. (**1996**). *Elementary Survey Sampling*, 5th edition, Duxbury Press.

Copyright © 2005 John Wiley & Sons, Ltd. All rights reserved.