Visions in Leisure and Business

Volume 4 | Issue 2 Article 26

1985

Sample Size for Research in Tourism

D. C. Williams Jr.
University of Southern Mississippi

Follow this and additional works at: https://scholarworks.bgsu.edu/visions

Recommended Citation
Williams, D. C. Jr. (1985) "Sample Size for Research in Tourism," *Visions in Leisure and Business*: Vol. 4 : Iss. 2 , Article 26.
Available at: https://scholarworks.bgsu.edu/visions/vol4/iss2/26

This Article is brought to you for free and open access by the Human Movement, Sport and Leisure Studies at ScholarWorks@BGSU. It has been accepted for inclusion in Visions in Leisure and Business by an authorized editor of ScholarWorks@BGSU.

SAMPLE SIZE FOR RESEARCH IN TOURISM

BY

DR. D.C. WILLIAMS, JR., DIRECTOR

BUREAU OF BUSINESS RESEARCH
UNIVERSITY OF SOUTHERN MISSISSIPPI
SOUTHERN STATION BOX 5094
HATTIESBURG, MISSISSIPPI 39406

ABSTRACT

The amount of research conducted in the tourist industry is not as extensive as in some other industries. The problem of determining the appropriate sample size for statistical inference seems to be perplexing to many people in the industry. This article suggests that it is not as complex as often perceived. The information needed and a procedure to determine the maximum sample size needed for problems involving proportions and variables are presented. A technique for correcting the sample size needed for populations of relatively small size is also presented.

SAMPLE SIZE FOR RESEARCH IN TOURISM

INTRODUCTION

Research in tourism is sometimes considered to be in its infancy. That is somewhat natural because tourism is part of the service industry. The service sectors of an economy usually become relatively more important as the economy develops and matures.

Research is conducted to obtain information, which may be obtained from the entire population of interest or from a sample of the population. Sampling is common for two major reasons. Sample results are usually adequate and the cost of obtaining information from a sample is usually much less than from a population. A major question that usually arises when considering sampling is how many of the population should be included in the sample i.e., what sample size is needed.

The question of sample size often seems to appear complex and confusing, especially for people not trained in statistics or research. The tourist industry is no exception. The purpose of this article is to discuss this question. Hopefully, this discussion will be of interest and helpful to people in the tourist industry.

TYPES OF SAMPLES

The ability to make statements concerning characteristics of the population by evaluating the results of a sample is based upon the techniques of statistical inference. It follows that the sample must have certain characteristics for the reliability of the sample results to be measured accurately. The methods of selecting a sample are classified into two groups:

1. Probability samples. A sample is considered a probability sample if all the items in the population have an equal chance of being included in the sample. The major probability sampling methods are (a) simple random sampling, (b) systematic sampling, (c) stratified random sampling, and (d) cluster sampling.

2. Nonprobability samples. If the sample is based on the judgment of the selectors rather than randomness, it is considered to be a nonprobability sample. Nonprobability sampling methods include (a) judgment sampling, (b) quota sampling, (c) convenience sampling, and (d) panel sampling.

There is no one "best" method of drawing a sample. The objective of sampling is to obtain the needed information that reflects the characteristics of the population at a minimum of cost and effort. It is often argued that with knowledge about the population, the desired information can be obtained with a smaller sample size and with less expense by using one of the nonprobability methods. These, however, do not have the characteristics common to probability sampling, which are that all of the items in the population being studied have a chance of being selected. Thus, the reliability of the sample results cannot be measured accurately for nonprobability sampling.

A questionnaire is useful in soliciting information from a population and/or sample. It ensures that each person is asked the same questions. It does not guarantee that every respondent interprets every question in the same manner or that the response will be truthful. As Mr. Bruce W. Wolff said, some people "will give you answers they think you want to hear . . . not what they really and truly intended to do. They will say one thing, but when it comes to buying a ticket or making a selection, they will resort to different mechanicisms of selection than they have confided to the researchers." (1) However, this is a problem of research in general rather than a problem of sample size. A questionnaire does assist in minimizing biases. The primary methods of administering the questionnaire include personal interview, telephone and mail. The method selected usually depends upon the length of questionnaire, time and cost. The response rate is usually higher with personal interviews and the questionnaire can be longer and more complex. Telephone and mail questionnaires should be short and relatively simple.

SPECIFICATIONS FOR SAMPLE SIZE

Statistical techniques have been developed to estimate the sample size needed in order to make statistical inferences. The sample size depends upon three factors: (1) How confident the sponsor of the study wants to be in the results of the sample, referred to as the confidence coefficient. The most commonly used are 99, 95, and 90 percent. The higher the confidence desired, the larger the sample needed. (2) The allowable error. For example, suppose a golf course developer wants to know what percent of the population in a particular market would become members of a club at a specified price, and a sample indicates 25 percent. The allowable error refers to the difference between the percentage of the sample and that would become members and the percentage of the population that would become members. The developer may accept an error of, say, .03 (3 percent). The smaller the allowable error, the larger the sample needed can be calculated. It is possible to estimate the maximum sample size.

The degree of confidence and the allowable error are value judgments of the person wanting the sample information. The standard deviation, however, is a characteristic of the population to be studied. If the deviation in the population is not known, a sample may be conducted to obtain an estimate of the deviation. Then, the sample size needed can be calculated. It is possible to estimate the maximum sample size for problems involving proportions without knowing the standard deviations by assuming the maximum variation possible. Also, it is easy to calculate a crude estimate of the standard deviation of problems involving variables if the largest and smallest units of the population, or the range, are known. Once the sample results are calculated, the adequacy of the sample size used could be determined, regardless of how the initial estimate of the sample size was determined.

SAMPLE SIZE FOR PROBLEMS OF PROPORTION

The technique of determining the sample size for problems involving variables and problems involving proportions is based upon the same probability concepts. An example of a problem involving a proportion is what percent of the golfers in a market will join a club at a given price. Questions that may be answered yes or no, questions reflecting a preference between two alternatives, etc., deal with problems of proportions. An example of a problem involving a variable is what the average income is of the population in the market. Many sample problems involve both variables and proportions. It would seem that the sample size needed for proportions would be adequate as an initial estimate of the size needed for variables. As noted earlier, once the data are collected, the adequacy of the sample size could be determined and the sample size could be increased, if needed. The focus here is on determining the adequate sample size for problems involving proportions because the maximum sample size needed can be calculated without knowing the standard deviations of the population. A formula is:

$$S_p = \sqrt{\frac{p\ 1-p)}{n}}$$

where:

Sp = the standard error of the proportion,

p = the proportion of favorable responses,

$1-p$ = the proportion of unfavorable responses,

n = the number in the sample.

Consider Cases 1 and 2.

Case 1

Assume: (1) the allowable error is .03 (3 percent), (2) the confidence coefficient desired is .95 (95% confidence), (3) the maximum variation is p = .50 and 1-p = .50, (this maximum variation is applicable to all problems of proportions),

let $D = \frac{\text{allowable error ,}}{\text{confidence coefficient}}$

which specifies the size of the standard error permitted,

$$= \frac{.03}{1.96} = .0153$$

(note that 1.96 standard errors represent the 95 percent confidence interval), by equating

$$D = S_p$$

$$D = \sqrt{\frac{P(1-p)}{n}}$$

$$D^2 = \underline{p(1-p)}$$

$$n = \frac{p(1-p)}{D^2}$$

Substituting the values for D and p into the formula gives:

$$n = \frac{.50(1-.50)}{(.0153)^2}$$

$$= \frac{.25}{.000234}$$

$$n = 1{,}068.$$

Case 2

Now assume: (1) the allowable error is .05 (5 percent), (2) the confidence coefficient desired is .90 (90 percent confidence), (3) the maximum variation, is p = .50 and 1-p = .50.

$$D = \frac{.05}{1.65} = .0303,$$

(note that 1.65 standard errors represent the 90 percent confidence interval)

and,

$$n = \frac{.25}{(.0303)^2}$$

$$n = 272$$

Thus, the range of the sample size needed based on the two sets of assumptions specified above is from 272 to 1,068.

SAMPLE SIZE CORRECTION FACTOR

These sample sizes are the maximum needed for any population size. A correction factor is appropriate for small population sizes. It is referred to as the finite correction factor. It is usually not applied unless the size of the sample is at least 5 percent of the size of the population. Let n = the initial estimate of the sample size and N = the population size. Then, if the ratio of the estimated sample size to the population size is five percent or more ($\frac{n}{N}$ = five percent or more), the correction is in order. A formula for the correction is:

$$n_c = \frac{n}{1 + \frac{n}{N}}$$

where n_c is the corrected sample size.

In Case 1 above, the initial estimated sample size of 1.068 is five percent of 21,360, (1,068 ÷ 0.05 = 21,360), so the correction factor would be appropriate for population sizes of less than 21,000 (21,360 rounded). In Case 2, the initial sampling size estimate of 272 is 5 percent of 5,400 (272 0.05 = 5,440). The correction factor would be appropriate for populations smaller than 5,400 (5,400 rounded).

For illustration, consider a survey of golfers. Assume there are 3,992 golfers in the market. The adjustment in Case 1 is:

$$n_c = \frac{n}{1 + \frac{n}{N}} = \frac{1,068}{1+1,068} \quad \frac{1,068}{1+.268} = \frac{1,068}{1,268} = 842$$

In Case 2 the corrected size is

$$n_c = \frac{272}{\frac{272}{1+3,992}} = \frac{272}{1+.0679} = \frac{272}{1.0679} = 255$$

Thus, the adjusted sample size would range from 255 under the assumption of Case 2 to 842 under the assumptions of Case 1.

SAMPLE SIZE FOR PROBLEMS OF VARIABLES

An example of calculating the sample size for a problem involving a variable is as follows. Consider the formula:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

where:

$s_{\bar{x}}$ = the standard error of the sample mean,

s = the standard deviation of the sample,

n = the sample size

The above formula is equal to

$$n = \left(\frac{s}{s_{\bar{x}}}\right)^2$$

For illustration, assume the owner of a tourist attraction wants to know the mean income of families in a given market with (1) an acceptable error of $1,000 at (2) a confidence of .90. The standard deviation of income in the population is not known but data available indicate the lowest income is $5,000 and the highest is $65,000. A crude approximation procedure estimating the standard deviation is, for problems involving variables where the standard deviation is not known is:

$$\frac{\text{Highest value in the population - Lowest value in the population}}{\text{six}}$$

Using the information specified above, an estimate of the standard deviation is:

$$s = \frac{\$65,000-\$5,000}{6} - \frac{\$60,000}{6} = \$10,000.$$

Similar to the illustration of problems involving proportions, let

$$D = \frac{\text{Allowable error}}{\text{Confidence Coefficient}}$$

$$= \frac{\$1,000}{1.65} = 606,$$

(Note again that 1.65 standard errors represent the 90 percent confidence interval)

by equating

$$D = s_{\bar{}}$$

$$D = \frac{s}{\sqrt{n}}$$

which, as noted above, is equivalent to

$$n = \left(\frac{s}{D}\right)^2$$

substituting the values for D and s in the formula

$$n = \left(\frac{10,000}{606}\right)^2 = (16.5)^2 = 272.$$

As with problems involving proportions, a correction factor is appropriate if the sample size is at least 5 percent of the size of the population. Thus, a corrected sample size would be appropriate for population sizes of less than 5,400 (272 ÷ .05 = 5,440 and rounded to 5,400). The corrected sample size for problems involving variables is calculated the same as presented above for the problem involving proportions.

SUMMARY AND IMPLICATIONS

In summary, sampling is a useful tool to obtain information desired by people in the tourist industry. However, determination of the appropriate sample size appears to be complex to many people. This article outlines and illustrates the steps that may be used to estimate the sample size. The maximum size calculated for problems involving proportions apply to all problems involving proportions with the same allowable error and degree of confidence specified. Thus, the maximum size calculated in this example could be used. Also, the maximum size may be adjusted by the formula presented for use in cases where the population size is relatively small. The sample size needed depends upon the value judgment of the research sponsor relative to the degree of confidence desired in the results and the error allowed, and the variation of the unit of measurement in the population, rather than the type of industry.

The purpose of sampling is to obtain quality data for decisions. If a probability sampling methods and an adequate sample size are used, the statistical reliability of the sample results can be measured. Two common problems that produce results of questionable quality are use of a sample that is too small and improper measurement and/or failure to measure the appropriate variables. The latter is a problem of survey design, which, although very important, was not the major focus of this article.

Care should be taken by public entities as well as private enterprises to ensure that quality results of surveys are achieved. For example, if a state agency supports a project that is approved based on faulty data, taxpayers in general are adversely affected. Also, individuals and firms often undertake business activities related to the state sponsored project based upon data released by the state. Faulty state data may result in financial disaster and bankruptcy of such individals and firms.

The article is intended neither to cover all situations nor to explain the sampling concepts in detail. It is intended, however, to suggest that determining the appropriate sample size is not as complex as it is often perceived. Readers interested in a more detailed explanation of sampling are referred to a statistic book dealing with the subject. Most introductory statistics books cover the subject matter. An example is <u>Statistical Techniques in Business and Economics</u>, Revised Edition, 1970, by Robert D. Mason, published by Richard D. Irwin, Inc., Homewood, Illinois.

REFERENCES

1. D. C. Williams, <u>Travel Research: Its Impact on the Travel Marketing Process</u>, Travel and Tourism Research Association Fourteenth Annual Conference, Banff, Alberta, June 12-15, p. 40, 1983.