Sample size determination

Influencing factors and calculation strategies for survey research

Ali A. Al-Subaihi, PhD

ABSTRACT

The paper reviews both the influencing factors and calculation strategies of sample size determination for survey research. It indicate the factors that affect the sample size determination procedure and explains how. It also provides calculation methods (including formulas) that can be applied directly and easily to estimate the sample size needed in most popular situations.

Saudi Med J 2003; Vol. 24 (4): 323-330

Descriptive research is one of the 3 broad categories of research (the other 2 categories are correlation and experimental) used to describe the characteristics of the subjects of the study. Descriptive research can be classified, in terms of how data is collected, as either survey research or observational research. Survey research is the most well known type of self-report research and widely used technique in many fields, including health administration, public administration, education, sociology, and economics.[1] A survey is an attempt to collect data from members of a population in order to determine the current status of that population with respect to one or more variables. Survey can be divided into 2 broad categories: the questionnaire and the interview. A questionnaire is usually a paper-and-pencil instrument that the respondent must complete, and the interview must be completed by the interviewer based on the respondent response. In a survey research as well as in other research types, after the research problem has been defined, the first 2 questions that concern the researchers' are: How many subjects are needed for the study, and how can they be selected? Unfortunately, the answer to these questions are not as easy as the researcher desires. There are factors which influence determining sample size and others influence determining sampling design. The researcher needs to know these factors and their effect beforehand to succeed in determining the adequate sample size. This paper attempts to highlight the factors relevant to determining the minimum sample size needed for descriptive studies and introduce some useful strategies that can be employed for the purpose of sample size determination.

Factors influencing determining sample size. Determining the adequate sample size is the most important design decision that faces the researcher.[2] The reason for this is that using too low sample size, the research will lack the precision to provide reliable answers to the questions that are under investigation. Moreover, using too large sample size, time, and resources will be wasted often for minimal gain. As stated previously, there are factors playing a vital role in determining the sample size. Knowing these factors and their effect helps the researcher to determine the sample

From the Institute of Public Administration, Riyadh, *Kingdom of Saudi Arabia*.

Address correspondence and reprint request to: Dr. Ali A. Al-Subaihi, Assistant Professor of Research Methodology and Applied Statistics, Institute of Public Administration, Riyadh 11141, *Kingdom of Saudi Arabia*. Tel. +966 (1) 4745146. Fax. +966 (1) 4792136. E-mail: subaihia@ipa.edu.sa

size needed appropriately. These factors are: the sampling design, statistical analysis, level of precision, level of confidence, degree of variability, and non-response rate. Each of these factors and their influences are as follows:

1). The sampling design. Due to the variability of characteristics among subjects in the population, the researcher typically applies scientific sampling design in the sample selection process to reduce the probability of having a biased view of the population. The researcher withdraws a sample using simple random sampling procedure or complex multistage sampling procedure that includes stratification, clustering, and unequal probabilities of selection. Or, they might withdraw a sample using one of the non-probability sampling designs such as convenience sampling, judgmental sampling, quota sampling, and snowball sampling.[3-7] The objective of the survey as well as other factors help to determine the appropriate sampling design and valid data collection methodology. In order to describe the target population adequately and make statistically valid inferences for the population using the sample survey data, the researcher should incorporate the sample design in both procedures: the sample size determination and data analysis.[8,9] That is, the researcher ought to use the sample size determination procedure that matches the sampling design which is going to be applied since the sample size required differs from one sampling design to another. To illustrate, sample sizes that are needed to estimate the monthly spending mean of a finite population (N = 800) within maximum allowable difference between the estimate and the true value [d] = $2 and 95% level of confidence are 217 when simple and 192 when stratified random designs were used, (for more details on the examples, see examples 3-5-1 and 3-5-3. Moreover, if one of the non-probability sampling designs is going to be utilized, the judgmental sample size determination procedure must be employed. The judgmental technique is normally recommended for non-probability sampling designs since the mathematical formulas that help in determining the sample size cannot be driven with lack of probability.

The unavailability of an objective method for non-probability sampling designs and the spread of the simple random design's sample size determination techniques among researchers lets most researchers use the random design's method with non-probability design. They perform it frequently to the extent that it became perfectly normal to see published survey study indicating that one of the probability sample size determination methods was employed when one of the non-probability sampling designs was actually applied. This is unfortunate since the validity of applying sample size determination technique that does not match the sampling design is questionable.

2) The statistical analysis that is planned. Another consideration with sample size is the number needed for the data analysis. If descriptive statistics (for example mean, SD, frequencies, and so forth) are going to be used, the situation is complicated as there is no existing single method (or formula) that works in any survey study. If inferential statistics (for example t-test, analysis of variance, multiple regression, so forth) are intended to be used, power analysis should be employed to determine the sample size needed. The main goal of power analysis is to allow the researcher to decide how large a sample is needed to allow statistical judgments that are accurate and reliable. The power of an inferential statistical test is the probability of rejecting a false null hypothesis, and is determined by 4 factors: sample size, level of significance, size of the population SD (σ), and magnitude of the means difference. Making an appropriate change in any of these factors increases the power of a test. The sample size has a direct relationship with the power of a test. Thus, the simplest way that researchers use to increase the power is to increase the sample size. For more details on determining sample size using the power of a test, the reader is referred to Cohen,[10] Kirk,[11] and Trochim.[1]

3) The level of precision. The sampling error is defined to be the difference between the parameter (which is a numerical quantity, such as the mean and SD, calculated using data collected from the entire population) and the statistic (which is a numerical quantity calculated using data collected from a sample). The sampling error is called precision in sampling contexts, and gives the researcher some idea relating to the accuracy of the statistical estimate. The level of precision, which also could be expressed in percentage such as ± 3%, ± 5%, ± 7%, or ± 10% (which are the commonly used values in humane studies), is the range of accuracy of estimating the true value of the parameter. And, it means that if the researcher finds that 80% of subjects in the sample have acquired a skill (or knowledge) under study with a precision level of ± 10%, the researcher might conclude that between 70% and 90% of subjects in the population have acquired the skill. The level of precision has a reverse relationship with the sample size. That is, the smaller the level of precision is predetermined, the greater sample size is needed. The reason for this is that the greater the sample size, the closer the sample is to the actual population itself. And, if the researcher takes a sample that contains the entire population, they actually have no sampling error (namely parameter = statistic). The relationship between level of precision and the sample size is not linear; however, it is curvilinear (**Figure 1a**), that is, the rate of improvement in the precision decreases as the sample size increases.[2] For example, from **Figure 1a**, one can see that the precision the researcher could get from a sample of size 250 is 50% less than the precision that they could get from a sample of size 1,000.

4) The level of confidence. The level of confidence, which is based on ideas encompassed under the central limit theorem (CLT) (for more details on CLT, the reader is referred to Glass and Hopkins),[12] is a value which indicates a specific probability that the sample contains the parameter being estimated. Such level of precision, the level of confidence is expressed in percentage such as 90%, 95%, or 99% which are commonly used values in social studies. The level of confidence means that, if a 95% is selected, 95 out of 100 samples will have the true population parameter within the range of precision specified earlier (**Figure 1b**). Though, there is always a chance that the sample which was obtained does not hold the true population value. The shaded areas in **Figure 1b** represent such samples with extreme values. This risk is reduced for 99% confidence levels and increased for 90% confidence levels.[13] The level of confidence has a positive correlation relationship with the sample size. That is, when other things being hold constant, the higher the confidence level predetermined, the larger the sample size needed (**Figure 1c**). For example, the sample sizes needed to estimate the population parameter with ± 3% precision level and 90%, 95%, and 99% confidence levels are 699, 964, and 1556.

5) The degree of variability. In some studies, the object is to estimate the percentage (p) (proportion) of subjects in the population having some attribute. For example, the researcher may wish to estimate the proportion of females in a nurse population, the proportion of divorced women in a workingwoman population, or the proportion of teenagers in a smoker population. In this situation, the variable of interest is an indicator variable namely:

$$x_i = \begin{cases} 1 \text{ if the subject i has the attributed} \\ 0 \text{ if the subjects i does not have it} \end{cases}$$

where x_i denote the variable of interest. The proportion of subjects in the population having some attribute refers to the distribution of attributes in the population. The degree of variability in the attributes being measured equals p (1-p) and has a direct relationship with the sample size. That is, the more the degree of variability of the distribution of attributes in the population, the larger the sample size is required to obtain a given level of precision. The less variable population, the smaller the sample size. For example, sample sizes that are needed to estimate true proportions p = 0.20 and 0.50 with 3% precision level and 95% confidence level are 683 and 1067 subjects (**Figure 1d**). From **Figure 1d**, one may notice that a proportion of 0.50 (or equivalently written 50%) requires the largest sample size since it indicates a greater variability than other proportion values. Thus, p = 0.50 is usually used to determine a conservative sample size when the true variability of the population attribute is unknown.

6) The non-response rate. It has been noted that the sample size needed is referred to the number of valid responses not the number of subjects. In other words, the sample size that the researcher had is not the number of subjects who were selected to participate in the research. Rather, it is the number of subjects who responded correctly to the survey. The difference between the 2 numbers is called a non-response error, which is common in survey research. This is unfortunate due to a high non-response rate might lead to biased results.[19] Common sources of non-response are: refusals, unable to answer, and not found. The consequences of a high non-response error vary. As non-response rate increases, the possibility for having a biased sample increases. This is because the obtained responses of a probability sample may no longer be representative of the target population. In addition, the non-response might reduce a probability sample to a convenience sample and consequently, the conclusions are weaker.[14] In an effort to obtain enough data for analysis, the researcher should increase the sample size needed by a certain percentage to compensate for non-response. This percentage varies according to who is surveyed. The non-response rate, for example, among busy people (such as politicians, businessmen, general managers and so forth) is higher than among normal people. Thus, the researcher who wants to survey busy persons must keep that into consideration and selects more subjects to ensure having an adequate sample size. Since the non-response rate varies according to who is surveyed, the researcher is advised to consult previous studies in the research arena to determine the percentage wanted for non-response adjustment.

Strategies for determining sample size. There are several approaches to determine the sample size. These include using a census for small populations, imitating a sample size of similar studies, using Internet sample size calculator, using published tables, and using formulas. Each strategy is discussed:

A) Using a census for small populations. A survey research that attempts to acquire data from each and every member in the population is called a census survey, and one approach to determine the sample size is to deal with the entire population and use it as the sample. If the population that the researcher wishes to study is small (for example <200), the researcher should measure the interest variables for every subject in the population. The reason for this is that a census eliminates sampling error, and virtually the entire population would have to be sampled in small populations to achieve a desirable level of precision.

B) Using a sample size of a similar study. Another approach to determine the size of the sample needed for a particular study is to use the same sample size as those of studies similar to the one under plan. However, the researcher must note that without reviewing the procedures employed in these studies, they may run the risk of repeating similar errors that were made in determining the sample size for another study. Inspite of that, reviewing the literature in the study's discipline along with reviewing the procedures employed can reduce the possibility of repeating same errors. The literature review can provide guidance on sample sizes that are usually used.

C) Using Internet sample size calculator. A third way to determine sample size is to utilize one of the Internet sample size calculators, which provide the

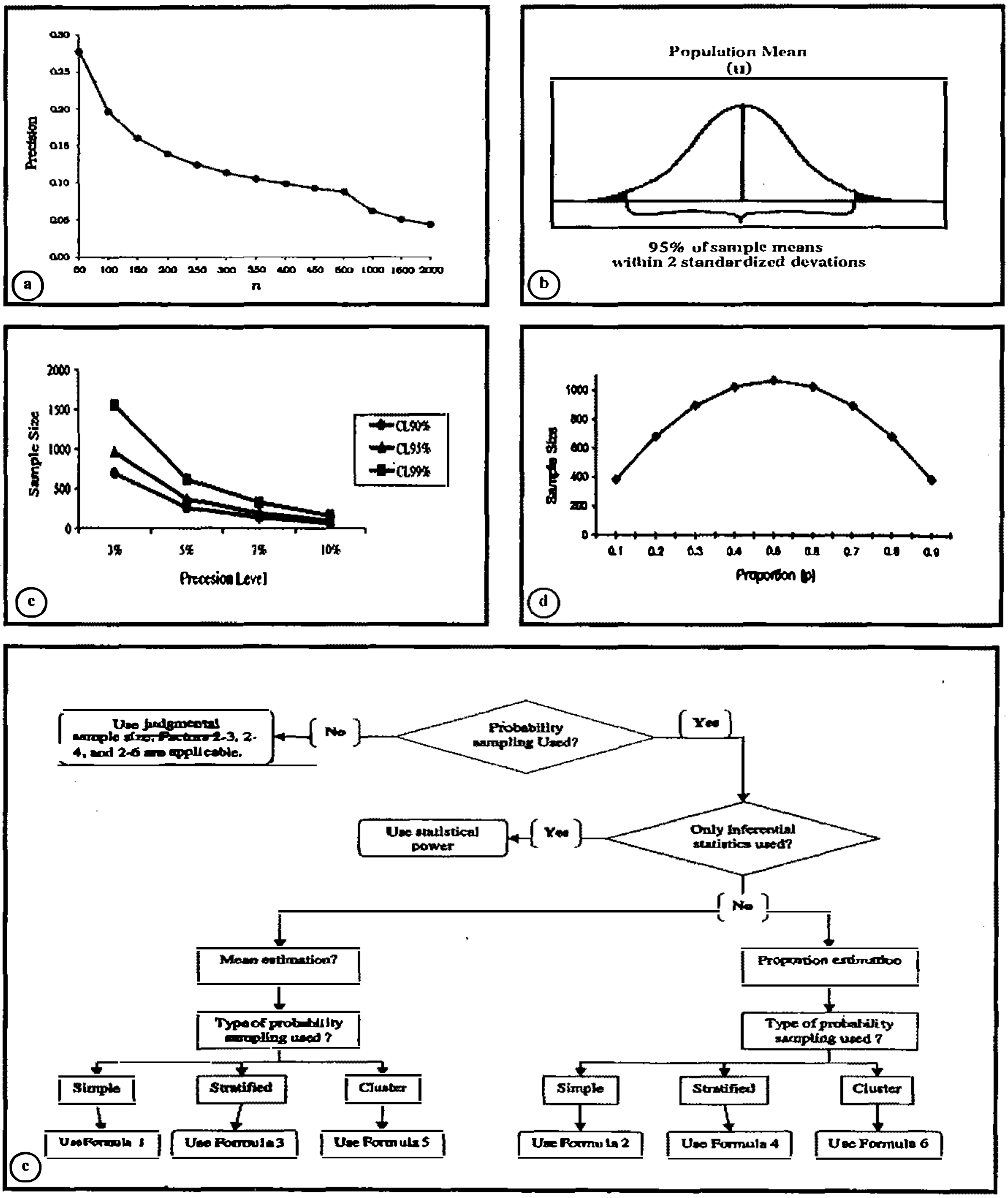

Figure 1 - A chart showing a) The relationship between the precision and sample size required to estimate a mean of normally distributed population with $\sigma = 1$. b) The 95% of sample means within 2 SD. c) The relationship between the level of confidence and sample size. d) The relationship between the degree of variability and sample size. e) Sample size determination flowchart. CL - confidence level

sample size for a given set of criteria. Sites such as: http://www.surveysystem.com/sscalchtm http:/ebook.stat.ucla.edu/calculators/sampsize.phtm and http://www.azplansite.com/samplesize.htm (and many others) provide an interactive way to determine the sample sizes that would be necessary for given combinations of precision, confidence levels, and variability. These sites and others similar are designed to provide sample size that: reflects the number of obtained responses, and not necessarily equals the number of surveys should be mailed or interviews must be planned (for more details about the difference between the two numbers, see a non-response rate factor); presumes the attributes being measured are distributed normally (or nearly so) with estimated proportion p = 0.5. Note that if the normality assumption cannot be met, then using the entire population should be considered. Assumes the simple random sampling design is going to be employed.

D) Using published tables. A forth way to determine sample size is to utilize published sample size tables, which are available in, almost, every research methodology and sampling textbooks. These tables are designed exactly in the same way that the Internet calculators are, however for fixed and predetermined combinations of precision, confidence levels, and variability. These features and the spread of the Internet usage among researchers limited the tables' utilization value. Thus, tables are not going to be provided here.

E) Using mathematical formulas. The traditional method to determine the sample size needed is to use directly the mathematical formulas. These formulas cover most of the probability sampling designs and can be used in the study aimed to estimate either the population mean or proportion. Indeed, they are the ones used in the Internet sample size calculators and the published tables. The most frequently used formulas are: (i) 3-5-1 simple random sampling (Mean). The simple random sampling is a method of selecting n (number) subjects out of the N (the population size) such that everyone of the

$$C^N_n = \frac{N!}{n!\,(N-n)!}$$

distinct samples has an equal chance of being drawn.[1] And, the formula that is used to determine the sample size for a study employs a simple random sampling design in order to estimate the population mean is by using formula 1-a:

$$n = \frac{N\,S^2}{(N-1)\,\frac{d^2}{Z^2} + S^2}$$

where n is the sample size needed, Z is the inverse of the standard normal cumulative distribution that correspond to the level of confidence, σ^2 is the variance of an attribute in the population which is usually estimated using a pilot sample variance (S^2), and d^2 is the maximum allowable difference between the estimate and the true value. The Z values that correspond to the frequently used confidence levels are shown in **Table 1.** When the population size is infinite or unknown, the sample size is estimated by using formula 1-b:

$$n = \frac{Z^2\ \sigma^2}{d^2} \approx \frac{Z^2 S^2}{d^2}$$

Survey samplers interested in estimating μ rarely have an estimate of σ readily available to use in either formula 1- a or formula 1- b for determining n. Thus, rather than guessing its value, the researcher must conduct a pilot sample to estimate the population SD σ. Typically, the pilot sample is a relatively small sample (namely n = 10, or 20).[13]

A pilot study (example 3-5-1) was conducted to estimate the population variance (σ^2) of monthly spending in health care among families. The sample variance (S^2) was found to be $232. What is the sample size needed in the study to estimate the population mean (μ) within d = $2 of the true mean with 95% confidence if the simple random sampling is planned to be used and N = 8,000? Solution to the sample size needed is:

$$n = \frac{N\,S^2}{(N-1)\,\frac{d^2}{Z^2} + S^2} = \frac{8{,}000\ (232)}{(8{,}000-1)\,\frac{(2)^2}{3.841} + (232)} = 216.77 \approx 217$$

(ii) Simple random sampling (proportion). When a simple random sampling design is employed to estimate the proportion of the population, the following formula can be used to determine the sample size: (formula 2- a)

Table 1 - The Z values that correspond to the frequently used confidence level.

Level of confidence(%)	Z	Z^2
90	1.645	2.706
95	1.960	3.841
99	2.576	6.635

Z - the inverse of the standard normal cumulative distribution that correspond to the level of confidence

$$n = \frac{N\,p(1-p)}{(N-1)\dfrac{e^2}{Z^2} + p(1-p)}$$

where p is the estimated variability of an attribute of interest in the population, e is the precision level, and the remaining variables are as defined in formula 1-a. Also, when the population size is infinite or unknown, the sample size is estimated by using formula 2- b:

$$n = \frac{Z^2\,p(1-p)}{e^2}$$

If the actual variability value of an attribute of interest in the population is unknown, p in formula 2- b is replaced by 0.5 to become: (formula 2-c)

$$n = \frac{0.25\,Z^2}{e^2}$$

In example 3-5-2, suppose that the nurse population in the Kingdom of Saudi Arabia is 5000. What is the sample size needed for a study to estimate the proportion of women in the nurse population with 5% precision and 95% confidence if the simple random sampling design is going to be used? Since there is no prior information on the degree of variability, formula 2-c will be used to estimate the sample size required:

$$n = \frac{0.25\,Z^2}{e^2} = \frac{0.25\,(3.841)^2}{(0.05)^2} = 384.1 \approx 385 \text{ subjects}$$

(iii) Stratified random sampling (Mean). In stratified sampling, a sample is drawn from each stratum of the population of N subjects, which is assumed to be divided into subpopulations (or strata) according to a specific characteristic such as age, gender, race, and so forth. When a simple random sample is taken in each stratum, the whole procedure is described as stratified random sampling.[1] To determine the sample size in order to estimate the population mean when the stratified random sampling design, the formula 3 is used:

$$n = \frac{\sum_{i=1}^{L}\left(\dfrac{N_i^2 S_i^2}{w_i}\right)}{N^2\dfrac{d^2}{Z^2} + \sum_{i=1}^{L}(N_i S_i^2)}$$

where L is the total number of strata, N_i is the size of stratum(i), S_i is the estimated variance of the attribute in the stratum(i), w_i is the estimated proportion of N_i to N, and the rest variables are previously defined.[15]

In example 3-5-3, the researcher wants to estimate the monthly spending mean of households in health care in a particular city, where households are divided into 3 strata: high, medium, and low income. It is expected that the mean spending varies between the 3 strata. The population size of the city is 8,000 and the size of high income household are 2,000, medium 1,000, and low income households are 5,000. A pilot study was conducted to estimate the variance in each stratum and found to be \$225, \$300, \$170. What are the sample sizes needed to estimate the population mean (μ) within d = \$2 with 95% confidence? the computation should be: N = 8,000, Ni = 2,000; 1,000; 5,000 and S^2_i = 225; 300; 170.

$$w_i = \frac{N_i}{N} = 0.25;\ 0.125;\ 0.625$$

Plugging values in formula 3, yields

$$n = \frac{\sum_{i=1}^{L}\left(\dfrac{N_i^2 S_i^2}{w_i}\right)}{N^2\dfrac{d^2}{Z^2} + \sum_{i=1}^{L}(N_i S_i^2)}$$

$$n = \frac{\left[\dfrac{(2000)^2(225)}{0.25} + \dfrac{(1000)^2(300)}{0.125} + \dfrac{(5000)^2(170)}{0.625}\right]}{\left[(8000)^2 \times \dfrac{2^2}{(1.96)^2}\right] + \left[(2000 \times 225) + (1000 \times 300) + (5000 \times 170)\right]}$$

n = 191.98 ≈ 192. Applying the strata weights, the researcher should sample: $n_1 = w_1 n$ = 0.25 (192) = 48; $n_2 = w_2 n$ = 0.125 (192) = 19.2 ≈ 20; $n_3 = w_3 n$ = 0.625 (192) = 124.8 ≈ 125 households from stratum 1, stratum 2, and stratum 3.

(iv) Stratified random sampling (proportion). The formula that gives the sample size needed to estimate the population proportion when the stratified random sampling technique used is formula 4:

$$n = \frac{\sum_{i=1}^{L}\left(\dfrac{N_i^2\,p_i(1-p_i)}{w_i}\right)}{\dfrac{N^2 e^2}{Z^2} + \sum_{i=1}^{L} N_i p_i (1-p_i)}$$

where p_i is the subpopulation proportion for stratum(i), and the remaining variables were defined previously.[15]

In example 3-5-4, refer to example 3-5-3. Find the sample sizes n_1, n_2, and n_3 needed to estimate the population proportion with 5% precision and 95%

confidence. Since there is no information on the strata proportions, the conservative values (p_1, p_2, p_3 = 0.5) should be used.

$$n = \frac{\sum_{i=1}^{L}\left(\frac{N_i^2 p_i (1-p_i)}{w_i}\right)}{\frac{N^2 e^2}{Z^2} + \sum_{i=1}^{L} N_i p_i (1-p_i)}$$

$$n = \frac{\left[\frac{(2000)^2 (0.25)}{0.25} + \frac{(1000)^2 (0.25)}{0.125} + \frac{(5000)2\,(0.25)}{0.625}\right]}{\left[(8000)^2 \times \frac{(0.05)^2}{(1.96)^2}\right] + \left[(2000 \times 0.25) + (1000 \times 0.25) + (5000 \times 0.25)\right]}$$

n = 375.15 ≈ 376. Applying the strata weights, the researcher should sample $n_1 = w_1 n$ = 0.25 (376) = 94; $n_2 = w_2 n$ = 0.125 (376) = 47; $n_3 = w_3 n$ = 0.625 (376) = 235 households from stratum 1, stratum 2, and stratum 3.

(v) Cluster sampling design (Mean). Cluster sampling is a sampling technique where the entire population is divided, usually geographically, into clusters, and a random sample of these clusters, not individuals, is selected. The number of clusters needed to estimate the population mean when the cluster sampling employed can be determined by (Formula 5):

$$n = \frac{N S_c^2}{\frac{N d^2 \bar{M}^2}{Z^2} + S_c^2}$$

where n is number of cluster in a simple random sample, M is the average size of the clusters in the population,

$$S_c^2 = \frac{\sum_{i=1}^{L}(x_i - \bar{x} m_i)^2}{n-1}$$

is the estimated variance of the cluster total, x_i is the total of all subjects in cluster$_{(i)}$. m_i is the size of cluster$_{(i)}$,

$$\bar{x} = \frac{\sum_{i=1}^{L} x_i}{\sum_{i=1}^{L} m_i}$$ is the estimated population mean.

In example 3-5-5, a health insurance company wanted to estimate the mean number of insurance subscriptions per household in a town containing 1,500 households. The town is divided into 200 non-overlapping geographic areas. What is the sample size (namely, number of clusters) needed to estimate the mean number of insurance subscriptions (μ) within d = 1 of the true value with 95% confidence if the estimated variance of the cluster total (S_c^2) is 100? N = 200; M = 1,500; d = 1; S_c^2 = 100; and $\bar{M}$ = M divided N = 7.5. Then the sample size needed is:

$$n = \frac{N S_c^2}{\frac{N d^2 M^2}{Z^2} + S_c^2} = \frac{200\,(100)}{\frac{200\,(1)^2 (7.5)^2}{(1.96)^2} + 100} = 6.6 \approx 7$$

(vi) Cluster sampling design (Proportion). The sample size required to estimate the proportion of the population when the cluster sampling is used can be estimated by using formula 6:

$$n = \frac{N S_c^2}{\frac{N e^2 \bar{M}^2}{Z^2} + S_c^2}$$

where $$S_c^2 = \frac{\sum_{i=1}^{L}(\alpha_i - p m_i)^2}{n-1}$$

is the estimated variance of the number α of successes in α_i cluster, α_i is number of subjects in cluster$_{(i)}$ having the attribute of interest,

$$p = \frac{\sum_{i=1}^{L} \alpha_i}{\sum_{i=1}^{L} m_i}$$ is the proportion.

In example 3-5-6, refer to the Example 3-5-5. Find the sample size (namely, number of clusters) needed to estimate the population proportion (p) with 5% precision and 95% confidence, if the estimated variance of the cluster total (S_c^2) is 5. The sample size is:

$$n = \frac{N S_c^2}{\frac{N e^2 \bar{M}^2}{Z^2} + S_c^2} = \frac{200\,(5)}{\frac{200\,(0.05)^2 (7.5)^2}{(1.96)} + 5} = 81.16 \approx 82$$

The following flowchart helps guiding the researcher to the suitable method that can be used to determine the sample size needed for a survey research.

In conclusion, the first question that comes to the researcher's mind, after finishing writing the research question, is "How many subjects are needed?". The question is essential and difficult; its importance comes from the desire not to survey either a too low or a too large sample size, and the difficulty comes from the non-existence of a direct answer. There are factors that play an important role in the process of determining the sample size, and in order for the researcher to determine the sample size required adequately, they must answer some supportive questions. These supportive questions are: Is one of the probability sampling designs planned to be used to select the subjects? If the answer is No (namely, one of the non-probability sampling designs is going to be used), the researcher needs to stop here and consult experts or previous studies in the field to determine the satisfactory sample size. That is because there is no objective sample size determination method that can be driven from non-probability sampling designs. If the answer is yes, the researcher needs to answer the following question. Is inferential statistics only going to be used to analyze the data? If yes, the researcher has to stop here and utilizes the statistical power analysis to determine the sufficient sample size. The Sample Power software that is developed by SPSS or any statistical software performs power analysis is recommended to be utilized here. If No (namely, descriptive statistics will be used), the researcher ought to respond to the next question. What is planned to be estimated? (Mean or Proportion). If the population mean will be estimated, the researcher needs to choose the related formula from formulas one, 3, or 5 according to the probability sampling design that is planned to be used. The predetermined precision and confidence levels are then plugged in. If the population proportion is intended to be estimated, the researcher has to choose formula 2, 4, or 6 also according to the probability sampling design that is going to be used. Again, the precision level, level of confidence, and the degree of variability in the attributes being measured are plugged into the analogous proportion formula.

Finally, the researcher needs to take into consideration who is surveyed to adjust for non-response rate. The non-response rate adjustment should always be applied regardless of what sampling design will be used or statistical analysis is planned. The sample size required is the number of valid responses not the number of subjects selected to participate in the study.

References

1. Trochim WM. Research Methods Knowledge Base. 2nd ed. Available from URL: http://trochim.human.cornell.edu/kb/index.htm.
2. McClave JT, Benson PG, Sincich T. Statistics for Business and Economics. 8th ed. Upper Saddle River (NJ): Prentice Hall International Inc; 2001.
3. Hansen, MH, Hurwitz WN, Madow WG. Sample Survey Methods and Theory. Volumes I and II. New York (NY): John Wiley & Sons Inc; 1953.
4. Kalton G. Introduction to Survey Sampling. SAGE University Paper series on Quantitative Applications in the Social Sciences. Series no. 07-035, Beverly Hills (CA), London (UK): SAGE Publications Inc; 1983.
5. Cochran WG. Sampling Techniques. 3rd ed. New York (NY): John Wiley & Sons Inc; 1977.
6. Kish L. Survey Sampling. New York (NY): John Wiley & Sons. Inc; 1965.
7. Lunsford T, Lunsford B. The Research Sample. Part I: Sampling. *Journal of Prosthetics and Orthotics* 1995; 7: 105-112.
8. Israel GD. Sampling the Evidence of Extension Program Impact, Program Evaluation and Organizational Development, IFAS, PEOD-5. Florida (FL): University of Florida; 1992-A
9. SAS Institute. "SAS/STAT User's Guide", Version 6. 4th ed. Cary (NC): SAS Institute Inc; 1989.
10. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale(NJ): Erlbaum; 1988.
11. Kirk RE. Experimental Design. 3rd ed. Pacific Grove (CA): Brooks/Cole Publishing Company; 1995.
12. Glass GV, Hopkins BR. Basic Statistics for the Behavioral Sciences. 3rd ed. Boston (MA): Allyn and Bacon; 1996.
13. Israel GD. Determining Sample Size, Program Evaluation and Organizational Development, IFAS. PEOD-6. Florida (FL): University of Florida; 1992-B
14. Israel GD. Sampling Issues: Non-response, Program Evaluation and Organizational Development, IFAS. PEOD-9. Florida (FL): University of Florida; 1992-C.
15. Sincich T. Business Statistics By Examples. 5th ed. Upper Saddle River (NJ): Prentice Hall International Inc; 1996.