MODULE 2

SAMPLE SIZE CONSIDERATIONS

Prof Oluwadiya K S
www.oluwadiya.com
College of Medicine,
Ekiti State University, Ado-Ekiti, Nigeria

A sample is a percentage of the total population in statistics. And the aim is to use the data from a sample to make inferences about the population as a whole. Determining sample size is one of the most ticklish problems in research. What should be the size "n" of the sample. If the sample size is too small, it may not achieve the objectives of the study, and if it is too large, we may incur huge costs and waste resources. Ethics, economics, time and other constraints dictates that we keep our sample size optimal. In essence, the sample must be just the right size: neither too large or too small.

What are the factors that may affect the size of a sample?

The hormogeneousness of the population: This is the uniformity of the population in relation to the variable of interest. If the population is homogeneous, a small sample will serve the purpose. But if heterogenous, then a large sample would be required.

i. **The type of study**: If a survey is to be carried out for descriptive purposes (e.g. prevalence of a disease or characteristics), the sample size is based on the required precision of the prevalence estimate and margin of error. On the other hand, if it is comparative study, the sample size is based on how detailed a comparison is desired and power calculations.
ii. **Grouping**: For adequate statistical analysis, each group in the study must have sufficient participants. Therefore, the larger the number of groups in a study, the larger the overall sample size of the study.
iii. **Level of precision**: If a high level of precision is required, then a relatively larger sample would be required.
iv. **The effect size**: Detecting very small differences requires larger samples than detecting large differences.
v. **Analytical technique**: More sophisticated statistic techniques require larger sample sizes than simple techniques. That is why if you are using many statistic methods for your study, you must calculate the sample size for each technique. You will use the largest calculated sample size for the study.For example, if you propose to find difference between group means, chi square as well as Multiple regression. You will calculate the sample size requirements for all three statistics and use the largest as the sample size for your study.

The bottomline:

There is no magic number or threshold. A sample size of 35 is a large sample for some laboratory studies while 20,000 may be too small for studying very rare diseases.

What is the source of Error in samples

Every sample has an weakness because when you use a certain sample of the population, uncertainty creeps in to your statistics. If you can survey only a part of the true population, you can never be 100% sure that the statistics derived from such sample are complete and accurate

representation of the population. This uncertainty is called sampling error and is usually measured by the confidence interval.

TWO QUESTIONS TO CONSIDER BEFORE THE ACTUAL SAMPLE SIZE CALCULATION:

Question 1: Is the study descriptive or comparative?

Descriptive studies include surveys to assess prevalence and other characteristics that have as their main objective the estimation of rates, proportions and means in a population. For example, a survey may be undertaken to assess the prevalence of malnutrition among children in Igbo-Ora. The main results to be reported might be the percentage of children who are malnourished, or the mean weight of children. Sample size determination for descriptive studies such as this is based on confidence intervals (i.e. the level of precision) required in providing estimates of the rates, proportions and means.

Comparative studies are studies where a comparison between two or more groups is the key analysis. These studies include case-control studies and randomized clinical trials. The main aim here is to determine whether there are statistically significant differences between groups with respect to the key outcome variable. Sample size determination for comparative studies is based on hypothesis tests and power. The aim is to calculate the sample that is large enough to find differences among groups when differences in actual fact, exist.

Please note that the same research might include the two types of studies.

Question 2: Is the main outcome measurement continuous or categorical? Continuous variables are usually summarized with mean or medians while categoric variables are usually summarized with proportions.

These two questions are all you need to answer before calculating the sample size for descriptive studies.

CAVEAT!
Many formulae are available for calculating sample size. The examples I am using here are to highlight some of the issues involved in calculating sample size for different study designs.

DESCRIPTIVE STUDIES

If the answer to the first question is descriptive study, you should then determine if the main outcome measure is continuous or categorical.

For categorical outcomes, the following formula may be used:

$$\boldsymbol{Sample\ Size} = \frac{(Z_\alpha)^2 p(1-p)}{d^2} \qquad (1)$$

Where

Z_α = Standard Normal Variate of the Type I Error to use. For a Type I Error of 0.05, Z is 1.96, and for a Type I error of 0.01, this is 2.58. (0.05 is most commonly used).

p = Expected proportion in the poulation with the characteristic or disease based on previous studies.

d = Absolute error of precision. This will be decided by the researcher.

Example:

A researcher wants to estimate the proportion in a population with hypertension. A previous study has shown that about 5.5% of the population in the city are hypertensive. The sample size will be calculated with a precision of 3.5% and a type I error of 5%.

$$\boldsymbol{Sample\ Size} = \frac{1.96^2 x\ 0.055\ x\ (1 - 0.055)}{0.035^2}$$

= 162.99

Plus 10% attrition rate, the total sample size is 179.2. This can be approximated to 180.

NOTE: Please see equation (3) for a modification of this equation.

For Continuous outcome variable, You may use the following formula:

$$\boldsymbol{Sample\ Size} = \frac{(Z_\alpha)^2 SD^2}{d^2} \qquad (2)$$

Where:

Z_α = Is the Standard Normal Variate as mentioned in the previous section

SD = Standard Deviation as determined by previous studies.

d = Absolute error of precision. This will be decided by the researcher.

Example:

What if the researcher in the previous section, is interested in knowing the mean blood pressure of the population. The sample size is to be determined with a 5% type I error, and precision of 5mmHG and Standard Deviation of 20mmHG based on a previous study.

$$\boldsymbol{Sample\ Size} = \frac{1.96^2\ x\ 20^2}{5^2}$$

= 61.45 (Approximate to 62)

NOTE: $(Z_\alpha)^2$, the Standard Normal Variate squared is 3.84. So, some authors approximate this to 4; therefore, equation 2 becomes

$$\boldsymbol{Sample\ Size} = \frac{4SD^2}{d^2} \qquad (3)$$

This makes the calculation easier with marginal increase in sample size. For example, using this formula, the sample size will be 64, instead of 62 using the previous formula. The same modification also can be applied to equation (1) above.

COMPARATIVE STUDIES

For comparative studies, you need to answer four more questions before you can calculate their sample sizes.

Question 1: **What is the acceptable significance level?** O.05 (5%) is the commonly used. The significance level is also called alpha, and it is the probability of making a type I error in hypothesis testing.

Question 2: How large a power do you want your study to have? Power is the probability of not making a type II error i.e. correctly rejecting the null hypothesis. Type II error (β) is the probability of accepting the null hypothesis when in fact, it is false (Power = 1-β). Power is so central to sample size calculations that sample size calculations are often referred to as power calculations. Conventionally, 0.8 (80%) is used.

Question 3: What is the population standard deviation? This is usually obtained from previous studies.

Question 4: What is the smallest effect that the researcher wants to detect? What is the magnitude of the smallest difference that the researcher wants to detect between the two (or more) groups? If you are comparing two drugs for hypertension, the researcher might be willing to wage that the new drug will reduce the blood pressure by 10mmHg compared to the standard drug. He will therefore use 10mmHg as the effect.

Once you have answered the four questions, you should then determine if the main outcome measure is continuous or categorical before you can finally calculate your sample size.

For categorical outcomes, the following formula may be used:

$$\boldsymbol{Sample\ Size} = \frac{2(z_\alpha + Z_\beta)^2 \mathrm{P}(1-\mathrm{P})}{(p_1 - p_2)^2} \qquad (4)$$

Where:

z_α = 1.96 at 5% type I error and 2.58 at 1% type 1 error

Z_β = O.842 at 80% power, and 1.28for 90% power.

P = Pooled prevalence = (prevalence in case + prevalence in control)/2

$p_1 - p_2$= Effect size = Difference in proportion of events in the two groups

Example

A researcher wants to compare a new drug for treating breast cancer. With the standard drug, 25% of the patients die within one year. The researcher feels that if the new drug improves the survival so that only 15% dies within a year, the finding can be considered clinically significant.

P = (0.25+0.15)/2 = 0.20

$p_1 - p_2$ = 0.15-0.25 = - 0.1

$$\boldsymbol{Sample\ Size} = \frac{2(1.96 + 0.84)^2\ 0.20(1-0.20)}{-0.10^2}$$

= 250.88 (251).

The researcher needs 294 per group.

For continuous outcomes, the following formula may be used:

$$\boldsymbol{Sample\ Size} = \frac{2SD^2(z_\alpha + Z_\beta)^2}{d^2} \qquad (5)$$

Where:

SD = Standard deviation from previous studies or pilot study

z_α = 1.96 at 5% type I error and 2.58 at 1% type 1 error

Z_β = O.842 at 80% power, and 1.28for 90% power.

D = Effect size = Difference between mean values.

Example

A researcher wants to compare the effect of a new antihypertensive drug to the standard. He feels that if the new drug reduces the mean blood pressure by 15mmHg, then it should be considered clinically significant. A previous study showed that the standard deviation was 25mmHg. If sets the level of significance at 0.05 and power at 0.80, then the sample size will be:

$$\boldsymbol{Sample\ Size} = \frac{2(25)^2(1.96 + 0.84)^2}{10^2}$$

= 98 subjects per group.

The Finite Population Correction factor

When the sample is more than 10% of the population from which it is taken, it is possible to use the Finite Population Correction factor (FPC). This will reduce the sample size required. The formula is

$$n_a = \frac{n_r}{1 + \frac{(n_r - 1)}{N}} \qquad (6)$$

Where

n_a = the adjusted sample size,

n_r = the original required sample size and

N = population size

Example

Having calculated 98 as the sample for the previous section, let's use the FPC to determine the adjusted sample size. Let's assume the population from which the sample was taken is 800.

$$n_a = \frac{98}{1 + \frac{(98 - 1)}{800}}$$

= 87.4 (Approximately 88)

Caveat: Be very cautious in using FPC because it reduces power. It should rarely be used in medical research anyway.

Other more complex forms of comparative studies may require different approaches:

Multiple Regression: Power analysis are not usually done for sample size estimation for multiple regression. A useful formula is N ≥ 50 + 8V, where N is the sample size and "V" is the number of independent variables in the model. For example, with 5 independent variables in the proposed model, the sample size will be 90.

Logistic Regression: Same as for multiple regression. However, because the outcome variable is binary rather than continuous, one may even opt for an even larger sample size.

Factor analysis: Like multiple regression, there are no power tables available for sample size estimation in factor analysis. A useful rule of thumb is that you must have a minimum of five subjects per variable, but you must make this up to at least 100, if what you calculate is less than 100.

Case Control Studies

Case control studies are studies where the group with the disease/condition (Case) is compared with the group without the disease/condition (Control) in regards to exposure to the risk factor under question.

The formula for sample size calculation for this design, like in the case of descriptive designs, depends on whether the outcome variable is categorical or continuous.

For categorical outcomes, the following formula may be used:

$$SampleSize = (\frac{r+1}{r})\frac{(\overline{p})(1-\overline{p})(Z_{\beta}+Z_{\alpha/})^2}{(p_1-p_2)^2} \quad (7)$$

Where:

R = Ratio of control to cases. Use 1 when case and control are equal.

$\overline{p}$ = Average proportion exposed= (Proportion of exposed cases + proportion of control)/2

z_{α} = 1.96 at 5% type I error and 2.58 at 1% type 1 error

Z_{β} = O.842 at 80% power, and 1.28for 90% power.

P_1-p_2 = Effect size = difference in proportion based on previous studies. P_1 is proportion in case group and P_2 proportion in control group.

Example

Supposing you want to do a study looking for a link between long-term smoking and delayed union of bone fractures. So, you will have two groups of patients: those with delayed union and those with normal union. Then you will give both groups a questionnaire on their smoking habits. Exposure (Yes or No) is a qualitative variable and equation 7 will be used for calculating the sample size. Let's assume that you want to fix the power at 80%, alpha at 0.05, and the expected proportions in case group and control group are 0.45 and 0.25 respectively, and you want to have equal number of cases and control. Then the sample size per group will be:

R = $\frac{1+2}{1}$ = 2

$\overline{p} = \frac{(0.45 + 0.25)}{2}$ = 0.35

Therefore, **Sample Size** = $2\frac{(0.35)(1-0.35)(0.84+1.96)^2}{(0.45-0.25)^2}$

= 44.59

So, you are going to have 45 subjects in each group

In case control studies with continuous outcomes, the following formula may be used:

$$SampleSize = (\frac{r+1}{r})\frac{SD^2(Z_\beta + Z_{\alpha/})^2}{d^2} \quad (8)$$

Where

SD: Standard deviation which is usually obtained from previous studies.

d = Expected mean difference between case and control. This may be from previous studies or just estimated.

Example

In the previous examples, the researcher is interested in mean duration of fracture union, which is a continuous variable. The expected mean difference is 4 days and the standard deviation was 7 days. All other parameters are the same as in the previous example.

$$SampleSize = 2\frac{7^2(0.84+1.96)^2}{4^2}$$

=48.02 in each group.

Yet Another Caveat!

Please note that all these formulas are for independent samples. The calculations will be different for paired or repeated samples designs.

Case control versus cohort studies?

Cohorts and case control studies are forms of observational studies. Randomized controlled clinical studies have always been the gold standard in Evidenced Based Medicine (Level I Evidence). However, RCTs are difficult to design for surgical investigations. Instead, well-designed cohort and case-control studies, recognized as level II or III evidence, can play an important role in deriving evidence for surgery specialties, and studies have shown that the results obtained from such studies, when well-designed are comparable to RCTs.

It is sometimes difficult for young researchers to differentiate between case control studies and cohort studies. Both study groups of patients either prospectively or retrospectively. However, here is the MAJOR difference; in cohort studies, a disease-free population is divided into groups based on their exposure to risk factor, and followed up in time until the disease or outcome of interest occurs.

However, in case control studies, the population is grouped on the outset into diseased and disease-free groups. Data about exposure to the risk factors are then collected retrospectively by interviews, records, or surveys.
So, in case-control studies, you have the answers (i.e. outcome) and you are working your ways towards the questions (i.e. causes), whereas, in cohort studies, you have your questions (prospective causes) and you are working your way towards the answers.
Expectedly, cohort studies are much more difficult to conduct than case-control studies and they tend to have much larger sample sizes than case control studies of similar power.

Attrition!

The sample size you estimate determines the number of complete cases which are needed for analysis. But it is rare for all subjects who started a study to complete it. Some may drop out, others may relocate and some may have incomplete data, especially on the key outcome variables. To deal with this, you should decide on an "attrition rate" and increase the sample size by this.

For example, if you expect to lose about 12% of the sample to attrition, then the sample size should be increased by a factor of 1 / (1 - 0.1) or 1.11. So, you will enroll 11% more subjects that the sample size calculation called for. In many instances, 10% is routinely taken as the attrition rate, but this may be changed if theory or experience so dictate.

Strategies for Minimizing Sample Size and Maximizing Power

Sometimes, matters of economy, time and ethics may make the calculated sample size impractical. What can we do in these circumstances?

i. You should check your calculations again!
ii. You should review the elements of your calculations. Is the effect size unreasonably small or the variability unreasonably large? Can you modify the alpha or beta or both without any major problem?
iii. If possible, use continuous variables instead of dichotomous variables. Continuous variables inherently have more information and so for any given sample size you get greater power (or a smaller sample size for a given power).
iv. Use paired measurements if this is an option as these tend to reduce the between-subject part of the variability of the outcome variable.
v. Use more precise measurements.
vi. Use unequal group sizes, if it is easier to recruit in one group than another (e.g. case-control).

Computer Programs for Sample Size Estimation

There are many software packages for calculating sample size. They range from free, open source to shareware and commercial. They can be basically divided into three categories:

i. Specialized Sample Size software programs
ii. General statistics software programs with capacity for Sample Size calculation.
iii. Web -based sample size calculators

No single program exists that allows researchers to determine sample size across all statistical

Procedures, therefore the decision about which program to use for sample size estimation

should be based on the following considerations: (a) cost; (b) ease of use, (c) availability of documentation and technical support and; (d) analytical and graphical capabilities. Users seeking to use these programs should be familiar with the statistical procedure to use.

The programs

This article will not attempt to describe all the available software programs for sample size estimation. We will categorize the software into:

i. General standalone Programs for sample size calculation
ii. Single procedure Programs for sample size calculations
iii. General purpose statistical software with capacity for sample size calculation
iv. Web-based sample size calculators.

Programs with multiple sample size calculation procedures

These programs can be used for sample size determinations for more than one procedures. Examples are free programs such as:

i. G*Power (For a user-friendly tutorial, please see: https://stats.idre.ucla.edu/other/gpower/)
ii. PS (http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize) and
iii. UnifyPow (Download from http://www.bio.ri.ccf.org/power.html). This is really a module that run within SAS.

The commercial programs include:

i. PASS,
ii. nQuery,
iii. Sample Power from SPSS,
iv. StatMate from Graphpad,
v. Power on X,
vi. Power and
vii. Precision and Statistical Power Analysis.

Single procedure Programs for sample size calculations

These programs are designed to estimate the sample size for just one statistical procedure. These are rather complex procedures like structural equation modelling and multilevel designs. They are generally free

General purpose statistical software with capacity for sample size calculation

i. R (Please see http://www.statmethods.net/stats/power.html for a tutorial on using R for sample size estimation. R is free
ii. Epi info (https://wwwn.cdc.gov/epiinfo/user-guide/StatCalc/How-to-Population-Survey-or-Description-Study.html) Epi info is free.
iii. SAS
iv. Stata
v. Minitab

Web-based sample size calculators.

Online statistic programs are increasing every day. As a rule, they tend to be user-friendly and so should be easy to use by nonstatisticians. However, like most things which are available online,

researchers should exercise caution in using online calculators. In a damning 2014 study, Meysamie et. al assessed 60 online sample size calculators and found that only five of them performed all four most commonly used sample size calculations (estimating proportion or mean and comparing proportions or means). The authors also compared the results from the online calculators with manual calculations by the authors using recommended formulae. They found that just 26% of the sites provide at least one accurate result. In fact, only one site had accurate results for all four calculations and just one other (Open epi) had three accurate results. The authors advised that sites which provide documentations on the formulae they used may be more reliable as users can at least test the accuracy of their computations.

Further Readings

1. Kasiulevicius V, Sapoka V, Filipaviciute R. Sample size calculation in epidemiological studies. Gerontology. 2006;7:225–31.
2. Cai J, Zeng D. Sample size/power calculation for case-cohort studies. Biometrics. 2004;60:1015–24
3. Meysamie A, Taee F, Mohammadi-Vajari M-A, Yoosefi-Khanghah S, Emamzadeh-Fard S and Abbassi M. Sample size calculation on web, can we rely on the results? J Med Stat Inform. 2014;
4. Patrick Dattalo. A Review of Software for Sample Size Determination. Evaluation & the Health Professions Volume 32 Number 3 September 2009 229-248

Next Topic: Formulating your questionnaire

See you next week.