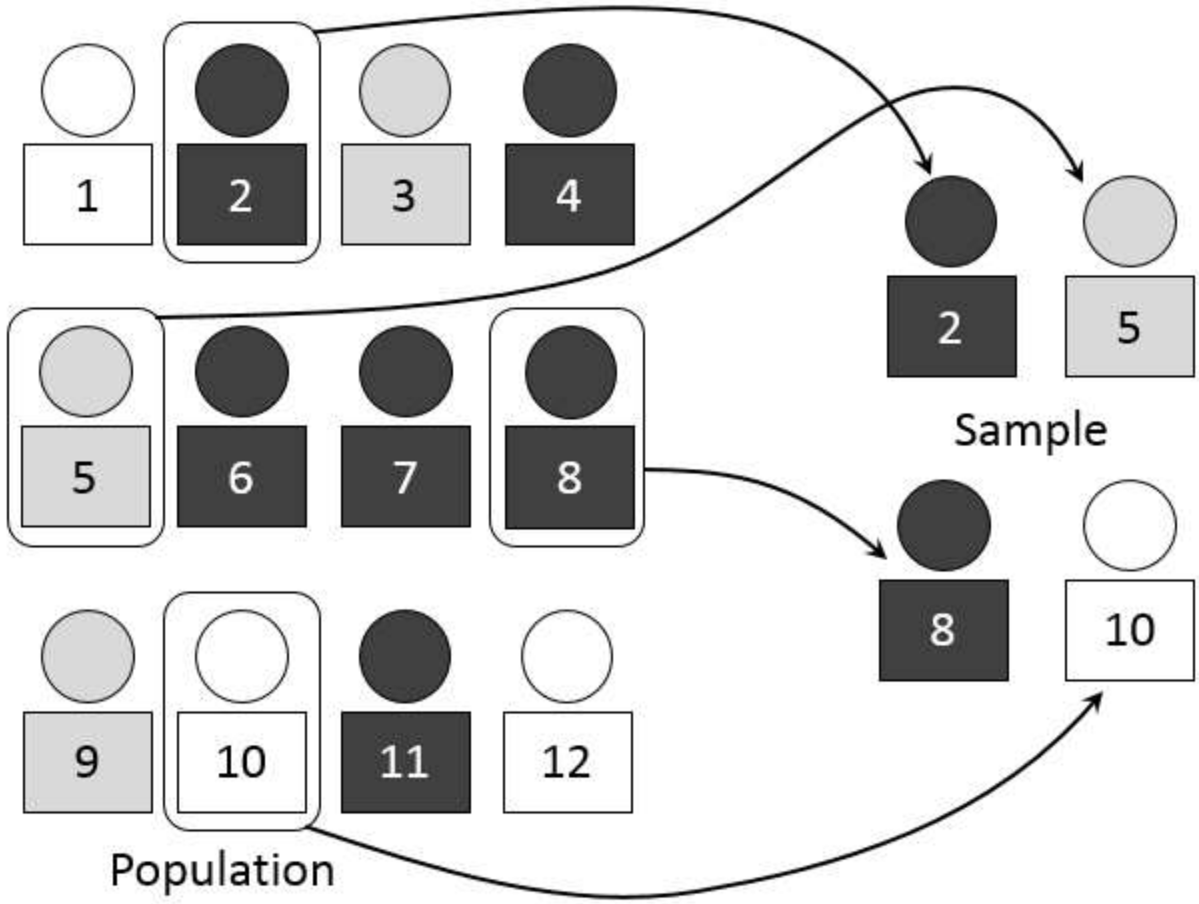
Sampling Techniques and Steps

In statistics and quantitative research methodology, a **sample** is a set of individuals or objects collected or selected from a statistical population by a defined procedure.[1] The elements of a sample are known as **sample points**, **sampling units** or observations.[citation needed] When conceived as a data set, a sample is often denoted by capital roman letters such {\displaystyle X} and {\displaystyle Y} , with its elements expressed in lower-case (e.g., {\displaystyle x_{3}}) and the sample size denoted by the letter {\displaystyle n} .[2][3]

Typically, the population is very large, making a census or a complete enumeration of all the individuals in the population either impractical or impossible. The sample usually represents a subset of manageable size. Samples are collected and statistics are calculated from the samples, so that one can make inferences or extrapolations from the sample to the population.

The sample may be drawn from a population *without replacement* (i.e. no element can be selected more than once in the same sample), in which case it is a subset of a population; or *with replacement* (i.e. an element may appear multiple times in the one sample), in which case it is a multisubset.[4]



A **complete sample** is a set of objects from a parent population that includes *all* such objects that satisfy a set of well-defined selection criteria.[5][*failed verification*] For example, a complete sample of Australian men taller than 2 m would consist of a list of *every* Australian male taller than 2 m. But it wouldn't include German males, or tall Australian females, or people shorter than 2 m. So to compile such a complete sample requires a complete list of the parent population, including data on height, gender, and nationality for each member of that parent population. In the case of human populations, such a complete list is unlikely to exist (the human population being in the billions). But such complete samples are often available in other disciplines, such as the set of players in a major sports league, the birth dates of the members of a parliament, or a complete magnitude-limited list of astronomical objects.

An **unbiased (representative) sample** is a set of objects chosen from a complete sample, using a selection process that does not depend on the properties of the objects.[6] For example, an unbiased sample of Australian men taller than 2 m might consist of a randomly sampled subset of 1% of Australian males taller than 2 m. But one chosen from the electoral register might not be unbiased since, for example, males aged under 18 will not be on the electoral register. In an astronomical context, an unbiased sample might consist of that fraction of a complete sample for which data are available, provided the data availability is not biased by individual source properties.

The best way to avoid a biased or unrepresentative sample is to select a random sample, also known as a probability sample. A random sample is defined as a sample where each individual member of the population has a known, non-zero chance of being selected as part of the sample.[7] Several types of random samples are simple random samples, systematic samples, stratified random samples, and cluster random samples.

A sample that is not random is called a non-random sample or a non-probability sampling.[8] Some examples of nonrandom samples are convenience samples, judgment samples, purposive samples, quota samples, snowball samples, and quadrature nodes in quasi-Monte Carlo methods.

REFERENCES

1. *Peck, Roxy; Olsen, Chris & Devore, Jay (2008), Introduction to Statistics and Data Analysis (3rd ed.), Belmont, Cal.: Thomson Brooks/Cole, p. 8, ISBN 978-0-495-11873-2, LCCN 2006933904, retrieved 2009-08-04*
2. ^ *"List of Probability and Statistics Symbols". Math Vault. 2020-04-26. Retrieved 2020-08-21.*
3. ^ *"What Is the Meaning of Sample Size?". Sciencing. Retrieved 2020-08-21.*
4. ^ *Borzyszkowski, Andrzej M.; Sokołowski, Stefan, eds. (1993), "A characterization of Sturmian morphisms" (PDF), Mathematical Foundations of Computer Science 1993. 18th International Symposium, MFCS'93 Gdańsk, Poland, August 30–September 3, 1993 Proceedings, Lecture Notes in Computer Science, **711**, pp. 281–290, CiteSeerX 10.1.1.361.7021, doi:10.1007/3-540-57182-5_20, ISBN 978-3-540-57182-7, Zbl 0925.11026*
5. ^ Pratt, J. W., Raiffa, H., and Schaifer, R. (1995). *Introduction to Statistical Decision Theory*. Cambridge, Mass.: MIT Press. ISBN 9780262161442. MR1326829
6. ^ Lomax, R. G. and Hahs-Vaughan, Debbie L. An introduction to statistical concepts (3rd ed).
7. ^ *Cochran, William G. (1977). Sampling techniques (Third ed.). Wiley. ISBN 978-0-471-16240-7.*
8. ^ *Johan Strydom (2005). Introduction to Marketing (Third ed.). Wiley. ISBN 978-0-471-16240-7.*

Steps In Sampling Techniques

According to Dalen (1981), there are several steps that researchers must consider in determining the sample, namely:

1. Determine the population,

2. Looking for accurate data on population units,

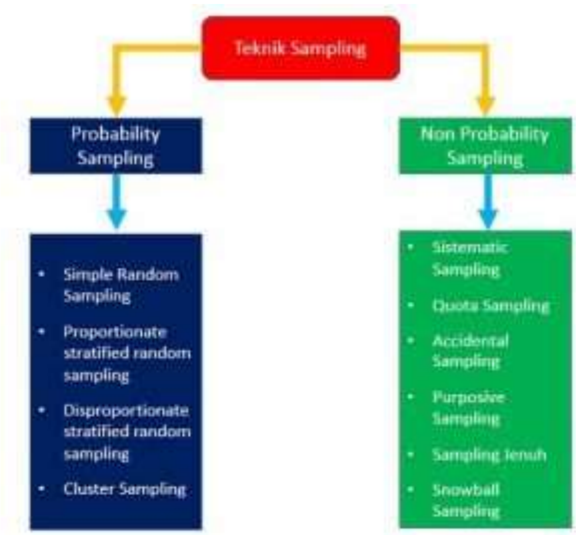
3. Choosing a representative sample,

4. Determine the adequate number of samples.

Types of Sampling Techniques

To determine the sample in research, there are various sampling techniques used. The sampling technique is based on randomization, namely taking subjects randomly from the collection, which can be grouped into 2, namely nonprobability sampling and probability sampling. These sampling techniques can be seen in the following scheme.

According to Sugiyono (2001), to determine the sample to be used in the study, various sampling techniques were used. It is schematically shown in the following diagram:



From the diagram above, it explains to us that the sampling technique can be grouped into two, namely: The first sampling technique is Probability Sampling and the second is Nonprobability Sampling.

Included in the probability sampling group include: simple random sampling, proportionate stratified random sampling, disproportionate stratified random sampling, and area (cluster) sampling (also called sampling by area). Meanwhile, nonprobability sampling includes: systematic sampling, quota sampling, accidental sampling, purposive sampling, saturated sampling, and snowball sampling.

Here's the explanation:

1. Probability Sampling

Probability sampling is a sampling technique that provides equal opportunities for each element (member) of the population to be selected as members of the sample. With probability sampling, the sampling is random or random from the existing population.

Probability sampling sample techniques include:

a. Simple Random Sampling

Simple Random Sampling is stated as simple (simple) because the sampling of members of the population is done randomly without paying attention to the strata in the population.

Simple random sampling is a technique for obtaining samples that are directly carried out in the sampling unit. Then each sampling unit as an element of a remote population has the same opportunity to become a sample or to represent its population. This method is done when members of the population are considered homogeneous.

This technique can be used if the number of sampling units in a population is not too large. The method of sampling with simple random sampling can be done by lottery, ordinal, or random number tables.

To determine the sample in this way is quite simple, but in practice it will take time. Especially if the number is large, the sample is large.

b. Proportionate Stratified Random Sampling

Proportionate Stratified Random Sampling is commonly used in populations that have stratified or layered arrangements. This technique is used when the population has members / elements that are not homogeneous and proportionally stratified. The downside of this method is that if there is no investigation of the subject list, it cannot be stratified.

c. Disproportionate Stratified Random Sampling

Disproportionate Stratified Random Sampling is used to determine the number of samples if the population is stratified but less proportional.

d. Cluster Sampling (Area Sampling)

Cluster Sampling (Area Sampling) is also cluster random sampling. This sampling technique is used when the population does not consist of individuals, but consists of groups of individuals or clusters. Area sampling technique is used to determine the sample if the object to be studied or the data source is very broad.

The weakness of this sampling technique can be seen from the sampling error rate. If it is more compared to sampling based on strata because it is very difficult to find clusters that are exactly the same level of heterogeneity with other clusters in the population.

2. Nonprobability sampling

Nonprobability sampling is a sampling technique that does not provide equal opportunities / opportunities for every element or member of the population to be selected as samples. The types of sampling techniques include:

a. Systematic Sampling

Systematic sampling is a sampling technique based on the sequence of members of the population who have been given serial numbers.

b. Quota Sampling or Quota Sampling

Quota sampling is a technique for determining a sample of a population that has certain characteristics to the desired quota. This technique the population number is not taken into account but is classified into several groups. Samples were taken by giving a certain quorum to the group. Data collection was carried out directly on the sampling unit. After the quota was met, data collection was stopped.

This technique is usually used and designed for research where a small sample is required where each case is studied in depth. And the danger, if the sample is too small, it will not be able to represent the population.

c. Accidental Sampling or Accidental Sampling

Accidental sampling is a sampling technique based on chance, in which anyone who happens to meet a researcher can be used as a sample, if it is deemed that the person who happened to be met is suitable as a data source.

In accidental sampling techniques, sampling is not determined in advance. The researcher immediately collected data from the sampling unit encountered.

d. Purposive Sampling

Purposive sampling is a sampling technique with certain considerations. The selection of a group of subjects in purposive sampling, is based on certain characteristics which are considered to have a close relationship with the characteristics of the population that have been previously known. So in other words, the sample unit contacted is adjusted according to certain criteria that are applied based on the research objectives or research problems.

e. Saturated Sampling

Saturated sampling is a sampling technique when all members of the population are used as samples. This is often done when the population is relatively small, less than 30 people. Saturated sample is also called a census, where all members of the population are sampled.

f. Snowball Sampling

Snowball sampling is a sampling technique that is initially small in number, then the sample is asked to select its friends to be the sample. And so on, so that the number of samples is getting bigger and bigger. It's like a snowball rolling, getting bigger and bigger. In qualitative research, a lot of using purposive and snowball samples.

Selection of the Type of Sampling Technique

Sorting out the types of probability and nonprobability sampling techniques is based on the existence of randomization or randomness, which is taking subjects randomly from the collection. In the case that randomization applies, each research subject has the same opportunity to be a member of the sample in line with the assumption that basically the probability distribution of events is present in all sections.

The Purpose of Sampling Technique according to Sugiarto in Martono (2010: 75). If it is not possible to observe all members of the population, this can happen if the members of the population are very large. Observation of an entire population can be destructive. Save costs, time and energy used. Able to provide accurate, more comprehensive and in-depth (comprehensive) information. (Martono, 2011: 75).

Selection of sampling techniques must be based on two important things, namely, reliability and efficiency. A reliable sample is a sample that has high reliability. This means that the smaller the sampling error, the lower the reliability of the sampling. If it is related to the variance of the statistical value, the criteria apply that the lower the variant, the higher the reliability of the sample obtained.

HOW TO CALCULATE SAMPLE

How to calculate the formula for the sample size of a study is very much determined by the research design used and the data taken. This type of observational research using a cross-sectional design will be different from the case-control study and the celebrity, likewise if the data collected is a proportion it will be different from if the data used is continuous data. In public health research, most use a cross-sectional or cross-sectional design or approach, although some use case controls or cohorts.

There are many formulas for calculating the minimum sample size of a study, but in this article, we will present some formulas that are most frequently used by researchers.

Cross-sectional Research Sample Formula

For survey research, usually the formula that can be used uses binominal proportions. If the population size (N) is known, it is searched using the following formula:

$$n = \frac{Z^2_{1-\alpha/2}\, p\, (1-p)\, N}{d^2(N-1) + Z^2_{1-\alpha/2}\, p\, (1-p)}$$

With a known population (N), the researcher can do random sampling).

However, if the population size (N) is unknown or (N-n) / (N-1) = 1, the sample size is calculated using the following formula:

$$n = \frac{Z^2_{\alpha}\, p\, q}{d^2} = \frac{Z^2\ p\,(1-p)}{d^2}$$

(Snedecor GW & Cochran WG, 1967)
(Lemeshowb dkk, 1997)

Information :

n = minimum number of samples required

= degree of confidence

p = proportion of children who are exclusively breastfed

q = 1-p (proportion of children not exclusively breastfed

d = limit of error or absolute precision

If set = 0.05 or Z1- / 2 = 1.96 or Z2

1- / 2 = 1,962 or rounded to 4, then the formula for the given magnitude of N is sometimes changed to:

$$n = \frac{4\ p\ q}{d^2}$$

Sample Research Sample Size Formulas

For example, suppose we want to find a minimal sample size for a study looking for the determinants of exclusive breastfeeding. To get the p value, we have to look at existing research or literature. From the results of research by Suyatno (2001) in the Demak-Central Java region, the proportion of infants (p) who were given exclusive breastfeeding was around 17.2%. This means that the p value = 0.172 and the q value = 1 - p. With the limit of error (d) set at 0.05 and an Alfa value = 0.05, the number of samples needed is:

Example of a Cross Sectional Sample Formula =

$$n = \frac{1{,}96^2 \,.\, 0{,}172 \,.\, 0{,}828}{0{,}05^{\ 2}}$$

= 219 people (minimum number)

If no p value is found from research or other literature, then a maximal estimation can be done with p = 0.5. If you want to be careful, the d value is around 2.5% (0.025) or smaller. Simplification of the formula above is widely known as the Slovin Formula.

Case Control and Cohort Research Sample Formulas

The formula used to find the sample size for both case control and cohort is the same, especially when using a measure of proportion. It's just that for famous research, there are also those that use a continuous data measure (mean value).

The sample size for the case control study aims to find the minimum sample for each case group and the control group. Sometimes researchers make a comparison between the number of samples of the case and control groups not necessarily 1: 1, but also 1: 2 or 1: 3 in order to get better results.

Minimum Sample Formula for Case Control Research Sample Size

The formula that is widely used to find a minimum sample of case-control research is as follows:

$$n = \frac{(p0.q0 + p1.q1)(Z_{1-\alpha/2} + Z_{1-\beta})^2}{(p1 - p0)^2}$$

Keterangan :

n = jumlah sampel minimal kelompok kasus dan kontrol
$Z_{1-\alpha/2}$ = nilai pada distribusi normal standar yang sama dengan tingkat kemaknaan α (untuk α = 0,05 adalah 1,96)
$Z_{1-\beta}$ = nilai pada distribusi normal standar yang sama dengan kuasa (*power*) sebesar diinginkan (untuk ß=0,10 adalah 1,28)
p0 = proporsi paparan pada kelompok kontrol atau tidak sakit
p1 = proporsi paparan pada kelompok kasus (sakit)
qo = 1 – p0 dan q1 = 1 – p1

Minimum Sample Formulas for Cohort Research Sample Size

In a typical study, what is sought is the minimum amount for the exposure and non-exposure groups or the exposed and unexposed groups. If a proportion data is used, then for the cohort study the p0 value in the above formula is the proportion of sickness in the unexposed population and p1 is the proportion of sickness in the exposed population or the value p1 = p0 x RR (Relative Risk).

If the p-value is continuous data (for example, average weight, height, BMI and so on) or not in the form of proportions, the sample size for the group is determined based on the following formula:

$$n = \frac{2(Z_{1-\alpha/2} + Z_{1-\beta})^2 \sigma^2}{(U1 - U2)^2}$$

Keterangan :

n = jumlah sampel tiap kelompok
$Z_{1-\alpha/2}$ = nilai pada distribusi normal standar yang sama dengan tingkat kemaknaan α (untuk α= 0,05 adalah 1,96)
$Z_{1-\beta}$ = nilai pada distribusi normal standar yang sama dengan kuasa (*power*) sebesar diinginkan (untuk ß=0,10 adalah 1,28)
σ = standar deviasi kesudahan (*outcome*)
U1 = *mean outcome* kelompok tidak terpapar
U2 = *mean outcome* kelompok terpapar

Sample Case Research Sample Size Formula

For example, let's say we want to find a minimal sample in a study on the effect of exclusive breastfeeding on baby weight. By using a significance level of 95% or Alfa = 0.05, and a power level of 90% or ß = 0.10, and the outcome observed was that the baby's body weight was determined to have an assumed value of SD = 0.94 kg, and the estimated difference between the mean outcome (outcome) weight of the unexposed and exposed groups during the first 4 months of infant life (U0 - U1) is 0.6 kg (referring to the results of the study by Piwoz, et al. 1994). the samples needed for each observation group, either exposed or unexposed are:

$$n = \frac{2(1,96 + 1,28)^2 \ (0,94)^2}{(0,6)^2}$$

= 51.5 people or rounded off: 52 people / group

In a typical study, the number of lost to follow or after-care individuals should be increased during observation, usually assuming 15%. In the example above, the minimum sample required is n = 52 (1 + 0.15) = 59.8 babies or rounded up to 60 babies for each group, either exposed or unexposed groups or a total of 120 babies for both groups. .

Experimental Research

According to Supranto J (2000) for experimental research with a completely randomized design, randomized group or factorial, it can simply be formulated:

(t-1) (r-1)> 15

where: t = number of treatment groups

j = number of replications

Sample Case Large Formula Research Sample Experiment

For example: If there are 4 treatments, then the number of repetitions for each treatment can be calculated:

(4 -1) (r-1)> 15

(r-1)> 15/3

r> 6

To anticipate the loss of experimental units, a correction is made with 1 / (1-f) where f is the proportion of experimental units that are lost or have withdrawn or dropped out.

Purposive Sampling Formulas

Basically, sampling is saturated then simple random sampling is the best sampling technique. However, we cannot close our eyes to the existence of certain criteria that can lead to bias in the results of the study. Therefore, the purposive technique needs to be considered for its use. Talking about the formula for determining the number of samples based on a purposive basis, will be a dilemma. Because even though we already know the list of populations that we will examine, there are times when the number is insufficient if we are going to apply the simple random sampling formula because of limitations or criteria. Then all of that is returned to the researcher, emphasizing the sufficient number or the tight limitations of the sample.

Purposive Sampling Steps

The steps in applying this technique are as follows:

1. Determine whether the objective of the study requires certain criteria in the sample to avoid bias.

2. Determine the criteria.

3. Determine the population based on a thorough preliminary study.

4. Determine the minimum number of samples that will be used as research subjects and meet the criteria.

Purposive Sampling Requirements

The requirements for using this technique include:

1. The criteria or limits are carefully defined.

2. Samples taken as research subjects are samples that meet predetermined criteria.

The advantages and disadvantages of purposive sampling

Advantages:

1. The selected sample is a sample in accordance with the research objectives.

2. This technique is an easy way to implement.

3. The selected sample is usually an individual or person who is easy to find or approach by the researcher.

Diadvantages

1. There is no guarantee that the number of samples used is representative in terms of numbers. 2. Where is not as good as sample random sampling.

3. Not included in the random sampling method.

4. Cannot be used as a generalization to draw statistical conclusions.

Purposive technique is an alternative that needs to be considered but also needs to be careful in using it. So that readers must be really careful before actually using this purposive technique. And don't forget, you should also describe these techniques in your research methodology chapter.

Thus the above has been described and discussed together regarding various sampling techniques in detail. Henceforth, to be more observant or careful in determining the sampling technique. So that the samples taken will be in accordance with your research problem.