INDEX

1. Sampling

A part of the population is known as sampling. The method consisting of the selecting for study, a portion of the 'universe' with a view to draw conclusions about the 'universe' or 'population' is known as sampling.

Sampling helps in time and cost saving. It also helps in checking their accuracy. But on the other hand, it demands exercise of great care & caution; otherwise, the results obtained may be incorrect or misleading.

2. Sampling and Sampling Methods

Researcher usually draw conclusions about large groups by taking a sample. A sample is a segment of the population selected to represent the whole population. The main benefits of sampling are as follows:

a) **Reduction in overall cost of research**

b) **Less time consuming, and in certain cases, this is desirable as well.**

c) **In case the population is consistent, this becomes even more desirable.**

Sampling and Sampling Methods

Ideally, the sample should be representative and should allow the researcher to make accurate estimates of the thoughts and behavior of the larger population. The population does not mean only human population; it can be factories, schools, etc. Population is denoted by N and sample as n.

The factors affecting interferences drawn from a sample are dependent upon the following:

a. **Sample Size: The larger the sample, the more is the accuracy.**

Sampling and Sampling Methods

b. Variation in population: The greater the variation in population, the greater will be the uncertainty of outcome. The higher the consistency in population, the more confident we are about the quality of outcome. The higher the variations in population, the larger should be the sample size.

3. Factors affecting Sample Survey

a. The size of the population
b. Amount of funds budgeted for the study
c. Facilities
d. Time

4. Sampling Procedure

i. Purpose of the survey
ii. Measurability
iii. Degree of precision
iv. Information about population
v. Nature of the population
vi. Geographical srea of study & size of the population
vii. Financial limits
viii. Time limitation
ix. Economy

5. Types of Samples

I. Probability or random sampling

Each person in the population has equal, independent, & known chances of being selected. In case, there are 100 elements in a population, every element has 1/100 chance of being selected in a sampling exercise. Here, independence means that selection of one element is neither being affected by the selection of other elements nor it will affect the other elements.

a. Simple random sampling

Every element or member of the population has a known and equal chance of being selected.

Suitability: This type of sampling is suited for a small homogeneous population.

Advantages: It is one of the easiest methods, all the elements in the population have an equal chance of being selected, simple to understand, does not require prior knowledge of the true composition of the population.

Types of Samples

Disadvantages: It is often impractical because of non-availability of population list or of difficulty in enumerating the population, does not ensure proportionate representation and it may be expensive in time and money. The amount of sampling error associated with any sample drawn can easily be computed. But it is greater than that in other probability samples of the same size, because it is less precise than other methods.

b. Stratified random sampling

In case the population is heterogeneous, the population can be divided into different strata. The population within a stratum is homogeneous with respect to the characteristic under study. Population is divided into mutually exclusive groups such as age groups, and random samples are drawn from each group. The population in a particular stratum may be in proportion to its population. Suppose there are 1,000 students in a college, 600 study humanities and 400 study commerce. In a sample of 100, 60 students will be from humanities and 40 from commerce, that is, in the same ratio as in the overall population.

Types of Samples

c. Cluster sampling

The simple & stratified sampling is adopted in situations when population size is small, and units are identifiable. But if the population is larger, the researcher can go for cluster sampling. The population is divided into mutually exclusive groups, and the researcher draws a sample of the group to interview.

For example, in a national level survey, at the first few levels, a few states may be selected. Within the states, a few districts may be selected; then, within each district, blocks may be selected and then villages. It is termed as **'multistage cluster sampling'**.

Cluster sampling

Suitability

The application of cluster sampling is extensive in farm management surveys, socio-economic surveys, rural credit surveys, demographic studies, ecological studies, public opinion polls, and large-scale surveys of political and social behavior, attitude surveys and so on.

Cluster sampling

Advantages: The advantages of this method is it is easier and more convenient, cost of this much less, promotes the convenience of field work as it could be done in compact places, it does not require more time, units of study can be readily substituted for other units, and it is more flexible.

Disadvantages: The cluster sizes may vary, and this variation could increase the bias of the resulting sample. The sampling error in this method of sampling is greater and the adjacent units of study tend to have more similar characteristics than do units distantly apart.

d. Area sampling

This is an important form of cluster sampling. In larger field surveys cluster consisting of specific geographical areas like districts, villages or blocks in a city are randomly drawn. As the geographical areas are selected as sampling units in such cases, their sampling is called area sampling. Some researchers feel that it is not a separate method of sampling, but forms part of cluster sampling.

II. Non-probability or Non-Random Sampling

Some persons in the population have a greater or unknown chance of being selected. This type of design is used when the number of elements in a population are either unknown or can’t be individually identified.

Advantages: The only merits of this type of sampling are simplicity, convenience and low cost.

Disadvantages: The demerits are it does not ensure a selection chance to each population unit. The selection probability sample may not be a representative one. The selection probability is unknown. It suffers from sampling bias which will distort results.

a. Convenience sampling

The researcher selects the easiest population members from which to obtain information.
Suitability: Though this type of sampling has no status, it may be used for simple purposes such as testing ideas or gaining ideas or rough impression about a subject of interest.

Advantage: It is the cheapest and simplest, it does not require a list of population and it does not require any statistical expertise.

Disadvantage: The disadvantage is that it is highly biased because of researcher's subjectivity, it is the least reliable sampling method, and the findings cannot be generalized

b. Judgement or purposive sampling

The researcher uses his/her judgement to select population members who are likely to provide accurate information. This can be used for historical research or descriptive research.

Suitability: This is used when what is important is the typicality and specific relevance of the sampling units to the study and not their overall representativeness to the population.

Advantage: It is less costly and more convenient and guarantees inclusion of relevant elements in the sample.

Disadvantage: It is less efficient for generalizing, does not ensure the representativeness, requires more prior extensive information and does not lend itself for using inferential statistics.

c. Quota sampling

The researcher finds and interviews a prescribed number of people in each of several categories. Here, the main criterion used by the researcher is the ease to access sample population. The sample is selected from a location convenient to him or her; here, there are some possibilities to include people with some visible characteristics. However, the results may not be generalized to larger populations.

Suitability: It is used in studies like marketing surveys, opinion polls, and readership surveys; which do not aim at precision, but to get quickly some crude results.

c. Quota sampling

Advantage: It is less costly, takes less time, non need for a list of population, and field work can easily be organized.

Disadvantage: It is impossible to estimate sampling error, strict control if field work is difficult, and subject to a higher degree of classification.

d. Accidental sampling

It is akin to quota sampling, but mostly used in market research, where a researcher can come across any person, and they may not have any information.

In some books, convenience and accidental Sampling are mentioned as same.

e. Snowball sampling: In this kind of sampling, the information may be selected from few individuals, and they may identify other people for the purpose of gathering information. They may also become a part of the sample. It creates a network of sample elements.

Types of Samples

Suitability: It is very useful in studying social groups, informal groups in a formal organization, and diffusion of information among professional of various kinds.

Advantage: It is useful for smaller populations for which no frames are readily available.

Disadvantage: The disadvantage is that it does not allow the use of probability statistical methods. It is difficult to apply when the population is large. It does not ensure the inclusion of all the elements in the list.

6. Sampling Errors

Sampling errors are statistical errors that arise when a sample does not represent the whole population. They are the difference between the real values of the population and the values derived by using samples from the population. Sampling errors occur when numerical parameters of an entire population are derived from a sample of the entire population. Since the whole population is not included in the sample, the parameters derived from the sample differ from those of the actual population.

They may create distortions in the results, leading users to draw incorrect conclusions. When analysts do not select samples that represent the entire population, the sampling errors are significant.

Sampling Errors

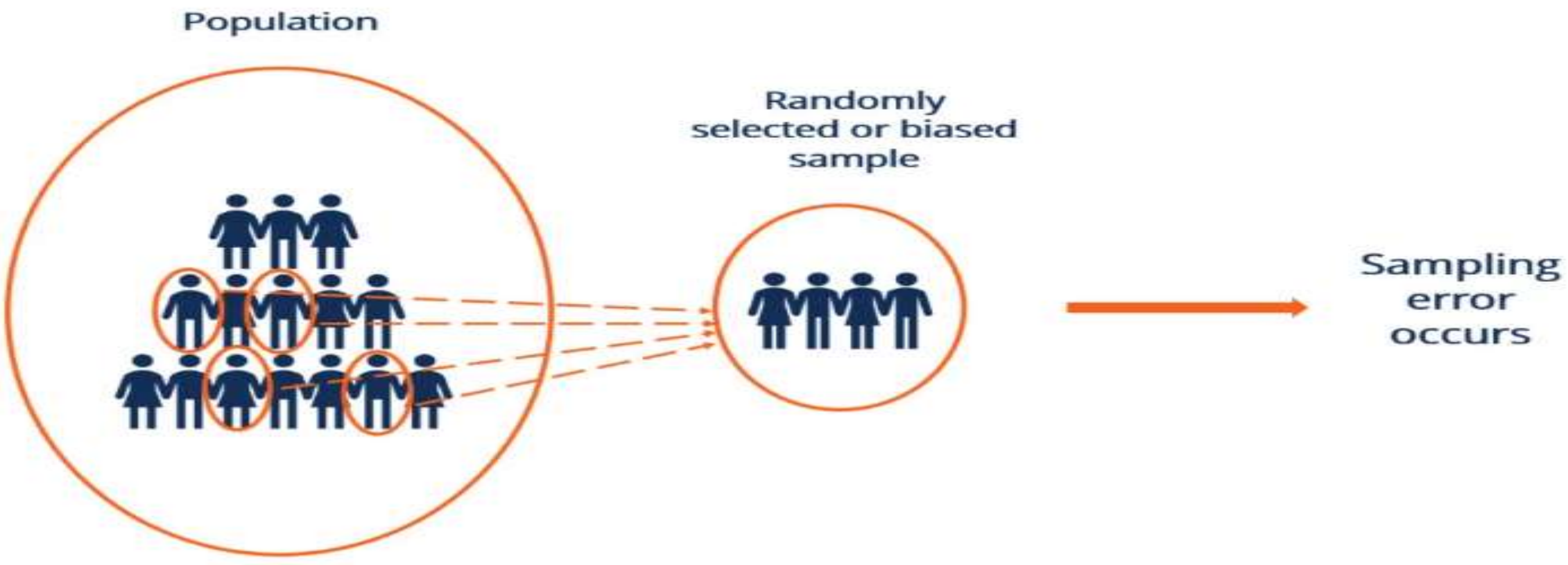
Sampling errors are deviations in the sampled values from the values of the true population emanating from the fact that a sample is not an actual representative of a population of data.

Since there is a fault in the data collection, the results obtained become invalid. Furthermore, when a sample is selected randomly, or the selection is based on bias, it fails to denote the whole population, and sampling errors will certainly occur.

These can be prevented if analysts select subsets or samples of data to represent the whole population effectively. Sampling errors are affected by factors such as the size and design of the sample, population variability, and sampling fraction. Increasing the size of samples can eliminate sampling errors. However, to reduce them by half, the sample size needs to be increased by four times. If the selected samples are small and do not adequately represent the whole data, the analysts can select a greater number of samples for satisfactory representation.

Sampling Errors

The population variability causes variations in the estimates derived from different samples, leading to larger errors. The effect of population variability can be reduced by increasing the size of the samples so that these can more effectively represent the population. Moreover, sampling errors must be considered when publishing survey results so that the accuracy of the estimates and the related interpretations can be established.



7. Categories of Sampling Errors

- **Population Specification Error** – Happens when the analysts do not understand who to survey. For example, for a survey of breakfast cereals, the population can be the mother, children, or the entire family.
- **Selection Error** – Occurs when the respondents' survey participation is self-selected, implying only those who are interested respond. Selection errors can be reduced by encouraging participation.
- **Sample Frame Error** – Occurs when a sample is selected from the wrong population data.
- **Non-Response Error** – Occurs when a useful response is not obtained from the surveys. It may happen due to the inability to contact potential respondents or their refusal to respond.

Non-Sampling Error

Non-sampling error refers to an error that arises from the result of data collection, which causes the data to differ from the true values. It is different from sampling error, which is any difference between the sample values and the universal values that may result from a limited sampling size.

Non-sampling errors can come in various forms, including non-response error, measurement error, interviewer error, adjustment error, and processing error.

8. Mechanics of Non-Sampling Error

Non-sampling error can arise when either a sample or an entire population (census) is taken. It falls under two categories

i. Random errors

Random errors are errors that cannot be accounted for and just happen. In statistical studies, it is believed that each random error offsets each other. So, they are of little to no concern.

ii. Systematic errors

Systematic errors affect the sample of the study and, as a result, will often create useless data. A systematic error is consistent and repeatable, so the study's creators must take great care to mitigate such an error. Non-sampling errors can occur from several aspects of a study. The most common non-sampling errors include errors in data entry, biased questions and decision-making, non-responses, false information, and inappropriate analysis.

9. Types of Non-Sampling Errors

i. Non-response error

A non-response error is caused by the differences between the people who choose to participate compared to the people who do not participate in a given survey. In other words, it exists **when people are given the option to participate but choose not to; therefore, their survey results are not incorporated into the data.**

ii. Measurement error

A measurement error refers to all errors relating to the measurement of each sampling unit, as opposed to errors relating to how they were selected. The error often arises when there are confusing questions, low-quality data due to sampling fatigue (i.e., someone is tired of taking a survey), and low-quality measurement tools.

Types of Non-Sampling Errors

iii. Interviewer error

Interviewer error occurs when the interviewer (or administrator) makes an error when recording a response. In qualitative research, an interviewer may lead a respondent to answer a certain way. In quantitative research, an interviewer may ask the question differently, which leads to a different result.

iv. Adjustment error

An adjustment error describes a situation where the analysis of the data adjusts it so that it is not entirely accurate. Forms of adjustment error include errors with weighting the data, data cleaning, and imputation.

v. Processing error

A processing error arises when there is a problem with processing the data that causes an error of some kind. An example will be if the data were entered incorrectly or if the data file is corrupt.

10. Sampling Error vs. Non-Sampling Error

Sampling error and non-sampling error are used in similar contexts, but there are some crucial differences between both concepts. They include:

a. **Sampling error can arise even when no apparent mistake's been made, as opposed to non-sampling error, which arises when a mistake occurs.**
b. **Sampling error occurs when the sample is not representative of the universal truth, whereas non-sampling error is specific to a certain study design.**
c. **Sampling error can be reduced greatly as sampling size increases, but non-sampling error requires more methodical processes to reduce.**
d. **Sampling error is often caused by internal factors, whereas non-sampling error is caused by external factors not entirely related to a survey, study, or census.**

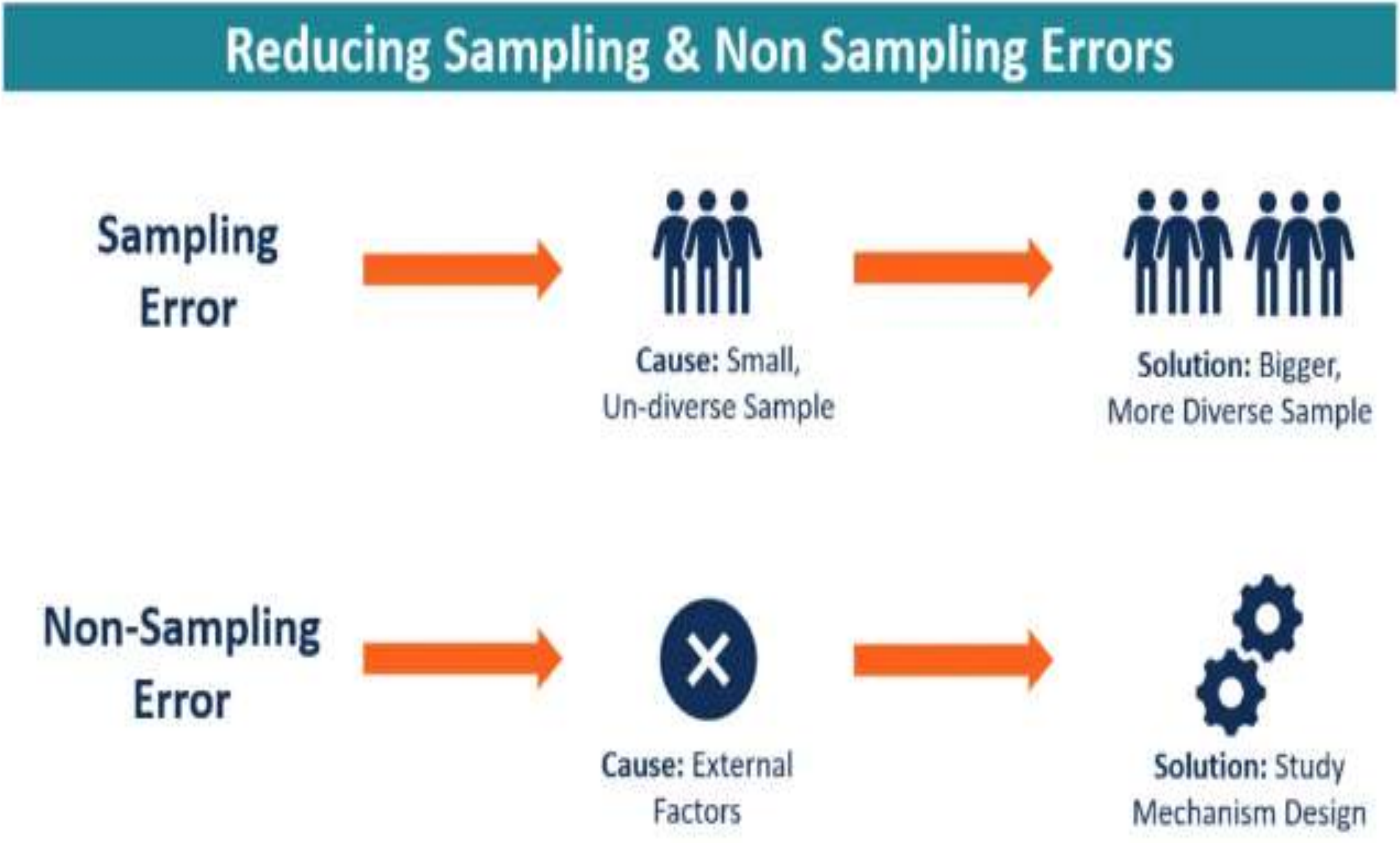
11. How to Reduce Errors

Reducing non-sampling error is not as easily achieved as reducing sampling error. With sampling error, you can reduce the risk of error by simply increasing the sample size. It will not work for non-sampling error, which is often very difficult to detect and eliminate (unless very methodical consideration is given to the source of the error).

To effectively reduce non-sampling error, careful consideration must be taken by those designing the study to ensure the validity of the results. As such, a researcher may design a mechanism into the study to reduce the error while subsequently not introducing another error.

For example, a researcher may pay the individual a bonus depending on the accuracy of their data entry, or they may film all interviews to ensure that the interviewer stays on topic and on script.

How to Reduce Errors



12. Sample Size

Scientific studies often rely on surveys distributed among a sample of some total population. Your sample will need to include a certain number of people, however, if you want it to accurately reflect the conditions of the overall population it's meant to represent. To calculate your necessary sample size, you'll need to determine several set values and plug them into an appropriate formula.

I. Part One: Determining Key Values

(a) **Know your population size.** Population size refers to the total number of people within your demographic.

(b) **Determine your margin of error.** Margin of error, also referred to as "confidence interval," refers to the amount of error you wish to allow in your results.

Sample Size

(c) Set your confidence level. Confidence level is closely related to confidence interval (margin of error). This value measures your degree of certainty regarding how well a sample represents the overall population within your chosen margin of error.

(d) Specify your standard of deviation. The standard of deviation indicates how much variation you expect among your responses.

(e) Find your Z-score. The Z-score is a constant value automatically set based on your confidence level. It indicates the "standard normal score," or the number of standard deviations between any selected value and the average/mean of the population.

Sample Size

- You can calculate z-scores by hand, look for an online calculator, or find your z-score on a z-score table. Each of these methods can be complex. However, Since confidence levels are standardized, most researchers simply memorize the necessary z-score for the most common confidence levels

80% confidence => 1.28 z-score

85% confidence => 1.44 z-score

90% confidence => 1.65 z-score

95% confidence => 1.96 z-score

99% confidence => 2.58 z-score

Sample Size

II. Part Two: Using the Standard Formula

(a)Look at the equation. If you have a small to moderate population and know all of the key values, you should use the standard formula. The standard formula for sample size is

Sample Size =

[z2 * p(1-p)] / e2 / 1 + [z2 * p(1-p)] / e2 * N]

N = population size

z = z-score

e = margin of error

p= standard of deviation

Sample Size

(b) Plug in your values. Replace the variable placeholders with the numerical values that actually apply to your specific survey.

Example: Determine the ideal survey size for a population size of 425 people. Use a 99% confidence level, a 50% standard of deviation, and a 5% margin of error.

For 99% confidence, you would have a z-score of 2.58.

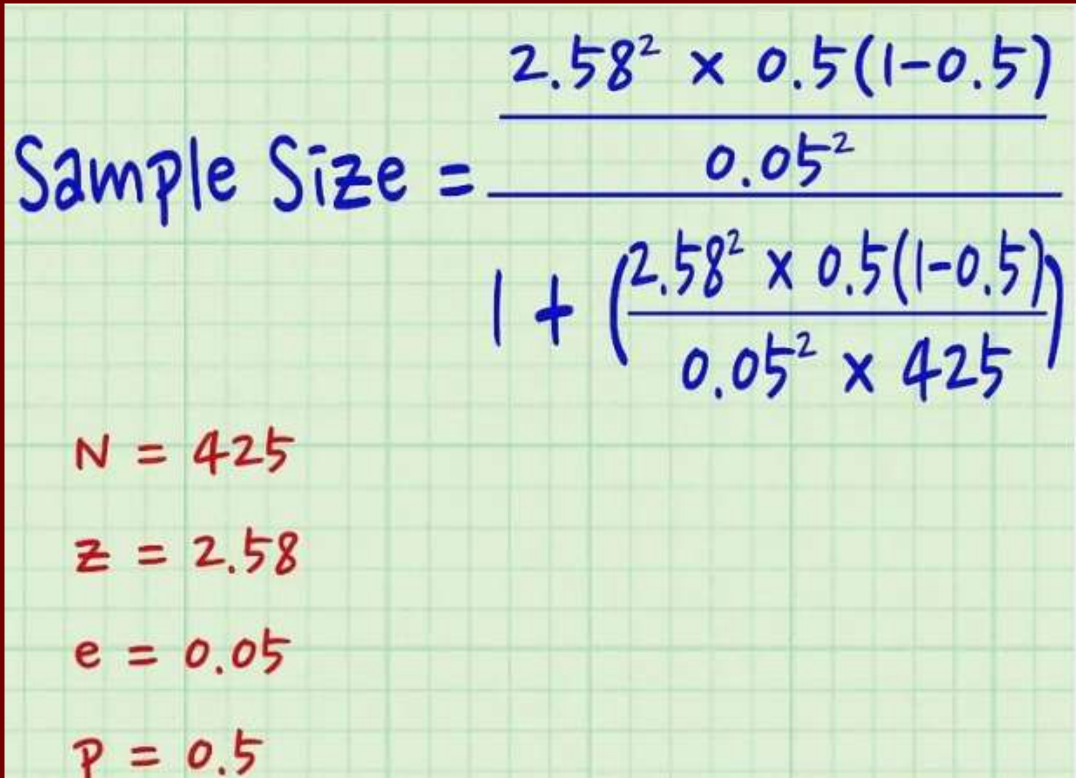
This means that

N = 425

z = 2.58

e = 0.05

p = 0.5



Sample Size

(c) Do the math. Solve the equation using the newly inserted numerical values. The solution represents your necessary sample size.

Example: Sample Size = [z2 * p(1-p)] / e2 / 1 + [z2 * p(1-p)] / e2 * N]

$$\text{Sample size} = \frac{\frac{2.58^2 \times 0.5(1-0.5)}{0.052}}{1 + \left(\frac{2.58^2 \times 0.5(1-0.5)}{0.052 \times 425}\right)}$$

$$= \frac{\frac{6.6564 \times 0.25}{0.0025}}{1 + \left(\frac{6.6564 \times 0.25}{1.0625}\right)}$$

$$= \frac{665.64}{2.566} = 259.39$$

Sample Size

III. Part Three: Creating a Formula for Unknown or Very Large Populations

(a) Examine the formula. If you have a very large population or an unknown one, you'll need to use a secondary formula. If you still have values for the remainder of the variables, use the equation:

Sample Size = [z2 * p(1-p)] / e2

z = z-score

e = margin of error

p = standard of deviation

Sample Size

(b) Plug your values into the equation. Replace each variable placeholder with the numerical values chosen for your survey.

Example: Determine the necessary survey size for an unknown population with a 90% confidence level, 50% standard of deviation, a 3% margin of error.

For 90% confidence, use the z-score would be 1.65.

This means

$z = 1.65$

$e = 0.03$

$p = 0.5$

$$\text{Sample size} = \frac{1.65^2 \times 0.5(1-0.5)}{0.03^2}$$

Sample Size

(c) Do the math. After plugging your numbers into the formula, solve the equation. Your answer will indicate your necessary sample size.

Example: Sample Size = [z2 * p(1-p)] / e2

Sample size

$$= \frac{1.65^2 \times 0.5(1-0.5)}{0.03^2}$$

$$= \frac{2.7225 \times 0.25}{0.0009}$$

$$= \frac{0.6806}{0.0009}$$

$$= 756.22$$

Sample Size

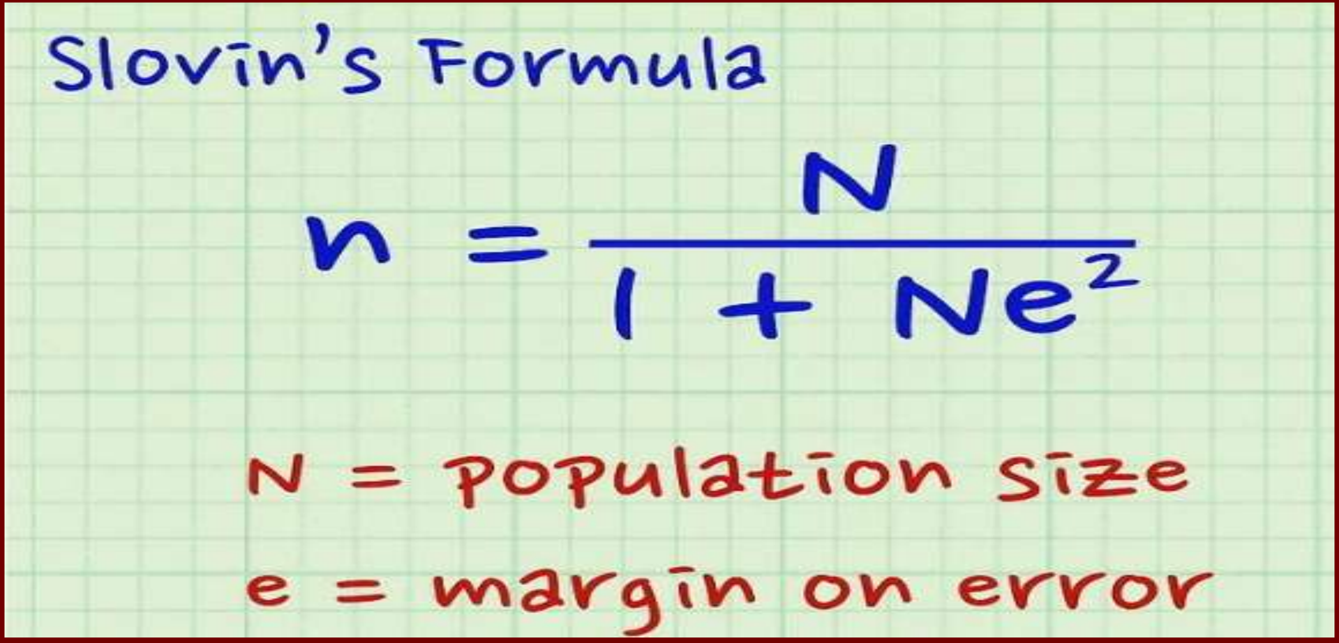
IV. Part Four: Using Slovin's Formula

(a) Look at the formula. Slovin's formula is a very general equation used when you can estimate the population but have no idea about how a certain population behaves. The formula is described as

Sample Size = N / (1 + N*e2)

N = population size

e = margin of error



Sample Size

(b) Plug in the numbers. Replace each variable placeholder with the numerical values that apply specifically to your survey.

Example: Calculate the necessary survey size for a population of 240, allowing for a 4% margin of error.

This means

$$N = 240$$

$$e = 0.04$$

$$n = \frac{240}{1 + 240\,(0.04^2)}$$

Sample Size

(c) **Do the math. Solve the equation using your survey-specific numbers. The answer you arrive at should be your necessary survey size.**

Example: Sample Size = N / (1 + N*e2)

$$n = \frac{240}{1 + 240(0.04^2)}$$

$$= \frac{240}{1 + 240(0.0016)}$$

$$= \frac{240}{1 + 0.384}$$

$$= \frac{240}{1.384} = 173.41$$