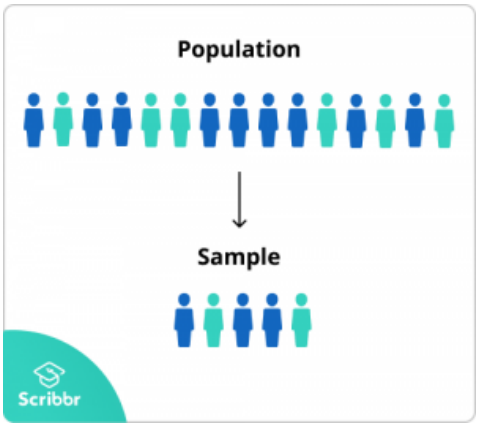
Sampling methods

- The **population** is the entire group that you want to draw conclusions about.
- The **sample** is the specific group of individuals that you will collect data from.

The population can be defined in terms of geographical location, age, income, and many other characteristics.



It can be very broad or quite narrow: maybe you want to make inferences about the whole adult population of your country; maybe your research focuses on customers of a certain company, patients with a specific health condition, or students in a single school.

If the population is very large, demographically mixed, and geographically dispersed, it might be difficult to gain access to a representative sample.

Sampling frame

The sampling frame is the actual list of individuals that the sample will be drawn from. Ideally, it should include the entire target population (and nobody who is not part of that population).

Example
You are doing research on working conditions at Company X. Your population is all 1000 employees of the company. Your sampling frame is the company's HR database which lists the names and contact details of every employee.

Sample size

The number of individuals in your sample depends on the size of the population, and on how precisely you want the results to represent the population as a whole.

You can use a sample size calculator to determine how big your sample should be. In general, the larger the sample size, the more accurately and confidently you can make inferences about the whole population.

What are sampling methods?

In a statistical study, sampling methods refer to how we select members from the population to be in the study.

If a sample isn't randomly selected, it will probably be biased in some way and the data may not be representative of the population.

There are many ways to select a sample—some good and some bad.

If a sample is to be used, by whatever method it is chosen, it is important that the individuals selected are representative of the whole population. This may involve specifically targeting hard to reach groups. For example, if the electoral roll for a town was used to identify participants, some people, such as the homeless, would not be registered and therefore excluded from the study by default.

There are several different sampling techniques available, and they can be subdivided into two groups: probability sampling and non-probability sampling. In probability (random) sampling, you start with a complete sampling frame of all eligible individuals from which you select your sample. In this way, all eligible individuals have a chance of being chosen for the sample, and you will be more able to generalise the results from your study. Probability sampling methods tend to be more time-consuming and expensive than non-probability sampling. In non-probability (non-random) sampling, you do not start with a complete sampling frame, so some individuals have no chance of being selected. Consequently, you cannot estimate the effect of sampling error and there is a significant risk of ending up with a non-representative sample which produces non-generalisable results. However, non-probability sampling methods tend to be cheaper and more convenient, and they are useful for exploratory research and hypothesis generation.

Probability Sampling Methods

1. Simple random sampling

In this case each individual is chosen entirely by chance and each member of the population has an equal chance, or probability, of being selected. One way of obtaining a random sample is to give each individual in a population a number, and then use a table of random numbers to decide which individuals to include.[1] For example, if you have a sampling frame of 1000 individuals, labelled 0 to 999, use groups of three digits from the random number table to pick your sample. So, if the first three numbers from the random number table were 094, select the individual labelled "94", and so on.

As with all probability sampling methods, simple random sampling allows the sampling error to be calculated and reduces selection bias. A specific advantage is that it is the most straightforward method of probability sampling. A disadvantage of simple random sampling is that you may not select enough individuals with your characteristic of interest, especially if that characteristic is uncommon. It may also be difficult to define a complete sampling frame and inconvenient to contact them, especially if different forms of contact are required (email, phone, post) and your sample units are scattered over a wide geographical area.

Every member and set of members has an equal chance of being included in the sample. Technology, random number generators, or some other sort of chance process is needed to get a simple random sample.

Example—A teachers puts students' names in a hat and chooses without looking to get a sample of students.

Why it's good: Random samples are usually fairly representative since they don't favor certain members.

2. Systematic sampling

Individuals are selected at regular intervals from the sampling frame. The intervals are chosen to ensure an adequate sample size. If you need a sample size n from a population of size x, you should select every x/n^{th} individual for the sample. For example, if you wanted a sample size of 100 from a population of 1000, select every $1000/100 = 10^{th}$ member of the sampling frame.
Systematic sampling is often more convenient than simple random sampling, and it is easy to administer. However, it may also lead to bias, for example if there are underlying patterns in the order of the individuals in the sampling frame, such that the sampling technique coincides with the periodicity of the underlying pattern. As a hypothetical example, if a group of students were being sampled to gain their opinions on college facilities, but the Student Record Department's central list of all students was arranged such that the sex of students alternated between male and female, choosing an even interval (e.g. every 20^{th} student) would result in a sample of all males or all females. Whilst in this example the bias is obvious and should be easily corrected, this may not always be the case.

Members of the population are put in some order. A starting point is selected at random, and every n^{\text{th}}nthn, start superscript, start text, t, h, end text, end superscript member is selected to be in the sample.

Example—A principal takes an alphabetized list of student names and picks a random starting point. Every 20^{\text{th}}20th20, start superscript, start text, t, h, end text, end superscript student is selected to take a survey.

3. Stratified sampling

In this method, the population is first divided into subgroups (or strata) who all share a similar characteristic. It is used when we might reasonably expect the measurement of interest to vary between the different subgroups, and we want to ensure representation from all the subgroups. For example, in a study of stroke outcomes, we may stratify the population by sex, to ensure equal representation of men and women. The study sample is then obtained by taking equal sample sizes from each stratum. In stratified sampling, it may also be appropriate to choose non-equal sample sizes from each stratum. For example, in a study of the health outcomes of nursing staff in a county, if there are three hospitals each with different numbers of nursing staff (hospital A has 500 nurses, hospital B has 1000 and hospital C has 2000), then it would be appropriate to choose the sample numbers from each hospital*proportionally* (e.g. 10 from hospital A, 20 from hospital B and 40 from hospital C). This ensures a more realistic and accurate estimation of the health outcomes of nurses across the county, whereas simple random sampling would over-represent nurses from hospitals A and B. The fact that the sample was stratified should be taken into account at the analysis stage.
Stratified sampling improves the accuracy and representativeness of the results by reducing sampling bias. However, it requires knowledge of the appropriate characteristics of the sampling frame (the details of which are not always available), and it can be difficult to decide which characteristic(s) to stratify by.

The population is first split into groups. The overall sample consists of some members from every group. The members from each group are chosen randomly.

Example—A student council surveys 100100100 students by getting random samples of 252525 freshmen, 252525 sophomores, 252525 juniors, and 252525 seniors.

Why it's good: A stratified sample guarantees that members from each group will be represented in the sample, so this sampling method is good when we want some members from every group.

4. Clustered sampling

In a clustered sample, subgroups of the population are used as the sampling unit, rather than individuals. The population is divided into subgroups, known as clusters, which are randomly selected to be included in the study. Clusters are usually already defined, for example individual GP practices or towns could be identified as clusters. In single-stage cluster sampling, all members of the chosen clusters are then included in the study. In two-stage cluster sampling, a selection of individuals from each cluster is then randomly selected for inclusion. Clustering should be taken into account in the analysis. The General Household survey, which is undertaken annually in England, is a good example of a (one-stage) cluster sample. All members of the selected households (clusters) are included in the survey.[1]

Cluster sampling can be more efficient that simple random sampling, especially where a study takes place over a wide geographical region. For instance, it is easier to contact lots of individuals in a few GP practices than a few individuals in many different GP practices. Disadvantages include an increased risk of bias, if the chosen clusters are not representative of the population, resulting in an increased sampling error.

The population is first split into groups. The overall sample consists of every member from some of the groups. The groups are selected at random.

Example—An airline company wants to survey its customers one day, so they randomly select 555 flights that day and survey every passenger on those flights.

Why it's good: A cluster sample gets every member from some of the groups, so it's good when each group reflects the population as a whole.

Non-Probability Sampling Methods

1. Convenience sampling

Convenience sampling is perhaps the easiest method of sampling, because participants are selected based on availability and willingness to take part. Useful results can be obtained, but the results are prone to significant bias, because those who volunteer to take part may be different from those who choose not to (volunteer bias), and the sample may not be representative of other characteristics, such as age or sex. Note: volunteer bias is a risk of all non-probability sampling methods.

2. Quota sampling

This method of sampling is often used by market researchers. Interviewers are given a quota of subjects of a specified type to attempt to recruit. For example, an interviewer might be told to go out and select 20 adult men, 20 adult women, 10 teenage girls and 10 teenage boys so that they could interview them about their television viewing. Ideally the quotas chosen would proportionally represent the characteristics of the underlying population.

Whilst this has the advantage of being relatively straightforward and potentially representative, the chosen sample may not be representative of other characteristics that weren't considered (a consequence of the non-random nature of sampling). [2]

3. Judgement (or Purposive) Sampling

Also known as selective, or subjective, sampling, this technique relies on the judgement of the researcher when choosing who to ask to participate. Researchers may implicitly thus choose a "representative" sample to suit their needs, or specifically approach individuals with certain characteristics. This approach is often used by the media when canvassing the public for opinions and in qualitative research.

Judgement sampling has the advantage of being time-and cost-effective to perform whilst resulting in a range of responses (particularly useful in qualitative research). However, in addition to volunteer bias, it

is also prone to errors of judgement by the researcher and the findings, whilst being potentially broad, will not necessarily be representative.

4. Snowball sampling

This method is commonly used in social sciences when investigating hard-to-reach groups. Existing subjects are asked to nominate further subjects known to them, so the sample increases in size like a rolling snowball. For example, when carrying out a survey of risk behaviours amongst intravenous drug users, participants may be asked to nominate other users to be interviewed.

Snowball sampling can be effective when a sampling frame is difficult to identify. However, by selecting friends and acquaintances of subjects already investigated, there is a significant risk of selection bias (choosing a large number of people with similar characteristics or views to the initial individual identified).

TABULATION OF DATA

Tabulation of data :

Tabulation may be defined as systematic presentation of data with the help of a statistical table having a number of rows and columns and complete with reference number, title, description of rows as well as columns and foot notes, if any.

Or

The process of placing classified data into tabular form is known as tabulation. A table is a symmetric arrangement of statistical data in rows and columns. Rows are horizontal arrangements whereas columns are vertical arrangements. It may be simple, double or complex depending upon the type of classification.

5 Major Objectives Of Tabulation:

(1) To Simplify the Complex Data

- It reduces the bulk of information i.e. raw data in a simplified and meaningful form so that it could be easily by a common man in less time.

(2) To Bring Out Essential Features of the Data

- It brings out the chief/main characteristics of data.
- It presents facts clearly and precisely without textual explanation.

(3) To Facilitate Comparison

- Presentation of data in row & column is helpful in simultaneous detailed comparison on the basis of several parameters.

(4) To Facilitate Statistical Analysis

- Tables serve as the best source of organized data for further statistical analysis.
- The task of computing average, dispersion, correlation, etc. becomes much easier if data is presented in the form of a table.

(5) Saving of Space

- A table presents facts in a better way than the textual form.
- It saves space without sacrificing the quality and quantity of data.

Types of Tabulation

(1) Simple Tabulation or One-way Tabulation

When the data are tabulated to one characteristic, it is said to be a simple tabulation or one-way tabulation.

For example: Tabulation of data on the population of the world classified by one characteristic like religion is an example of a simple tabulation.

(2) Double Tabulation or Two-way Tabulation

When the data are tabulated according to two characteristics at a time, it is said to be a double tabulation or two-way tabulation.

For example: Tabulation of data on the population of the world classified by two characteristics like religion and sex is an example of a double tabulation.

(3) Complex Tabulation

When the data are tabulated according to many characteristics, it is said to be a complex tabulation.

For example: Tabulation of data on the population of the world classified by three or more characteristics like religion, sex and literacy, etc. is an example of a complex tabulation.

Components of Data Tables

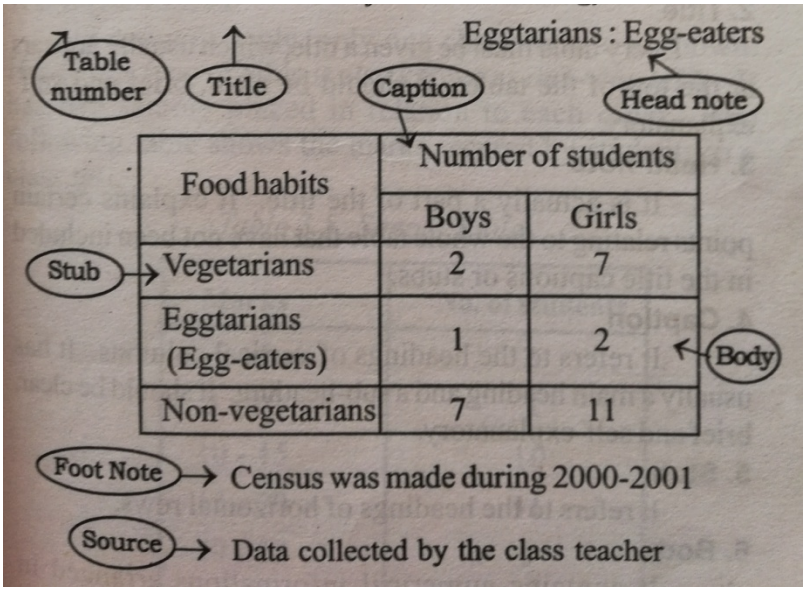
- **Table Number**: Each table should have a specific table number for ease of access and locating. This number can be readily mentioned anywhere which serves as a reference and leads us directly to the data

mentioned in that particular table. There are different practices with regard to the place where this number is to be given. The number may be given either in the centre at the top above the title or inside of the title at the top or in the bottom of the table on the left hand side. However, if space permits the table number should be given in the centre as is shown in the specimen table give on page. When there so that easy reference to it is possible.

- **Title:** A table must contain a title that clearly tells the readers about the data it contains, time period of study, place of study and the nature of classification of data. every table must be given suitable title. The title is a description of the contents of the table. A complete title has to answer the question what, where and when in that sequence. In other words:

  a)What precisely are the data in the table (i.e.) what categories of statistical data are shown? b) Where the data occurred 9i.e. the precise geographical, political or physical area covered)? C) When the data occurred (i.e. the specific time or period covered by the statistical materials in the table)?

- **Headnotes:** A headnote further aids in the purpose of a title and displays more information about the table. Generally, headnotes present the units of data in brackets at the end of a table title. It is a brief explanatory statement applying to all or a major part of the material in the table, and is placed below the point centered and enclosed in brackets. It is used to explain certain points relating to the whole table that have not been included in the title nor in the cations or studs. For example, the unit of measurement is frequently written as a head note, such as "in thousands" or "in million tonne3s or "in crores", etc.
- **Stubs:** These are titles of the rows in a table. Thus a stub display information about the data contained in a particular row. As distinguished from caption, stubs are the designations of the rows or row heading. They are at the extreme lift and perform the same function for the horizontal rows of numbers in the table as the column headings do for the vertical columns of numbers. The stubs are usually wide than column headings but should be kept as narrow as possible without sacrificing precision and clarity of statements.



- **Caption:** A caption is the title of a column in the data table. In fact, it is a counterpart if a stub and indicates the information contained in a column. It explains what the column represents it may consist of one or more column headings. Under a column heading there may be sub-heads. The caption should be clearly defined and placed at the middle of the column, if the different columns are expressed in different units. The units should be mentioned with the captions. As compared with the main part of the table the caption should be shown in smaller letters. This helps in saving space
- **Body or field:** The body of a table is the content of a table in its entirety. Each item in a body is known as a 'cell'. The body of the table contains the numerical information. This is the most vital art of the. Data presented in the body arranged according to description are classifications
- **Footnotes:** Footnotes are rarely used. In effect, they supplement the title of a table if required. Anything in a table which the reader may find difficult to understand from the title, captions and studs should be explained in footnotes. If footnotes are needed they are placed directly below the body of the table. Footnotes are used for the following main purposes:

  a)To points out any exceptions as to the basis of arriving at the data.
  b) Any special circumstances affecting the data, for example, strike, lock-out fire, etc.
  c)To clarify anything in the table.
  d) To give the source in case of secondary data
- **Source:** When using data obtained from a secondary source, this source has to be mentioned below the footnote.

Construction of Data Tables

There are many ways for construction of a good table. However, some basic ideas are:

- **The title should be in accordance with the objective of study:** The title of a table should provide a quick insight into the table.
- **Comparison:** If there might arise a need to compare any two rows or columns then these might be kept close to each other.
- **Alternative location of stubs:** If the rows in a data table are lengthy, then the stubs can be placed on the right-hand side of the table.
- **Headings:** Headings should be written in a singular form. For example, 'good' must be used instead of 'goods'.
- **Footnote:** A footnote should be given only if needed.
- **Size of columns:** Size of columns must be uniform and symmetrical.
- **Use of abbreviations:** Headings and sub-headings should be free of abbreviations.
- **Units:** There should be a clear specification of units above the columns.

The Advantages of Tabular Presentation

- **Ease of representation:** A large amount of data can be easily confined in a data table. Evidently, it is the simplest form of data presentation.
- **Ease of analysis:** Data tables are frequently used for statistical analysis like calculation of central tendency, dispersion etc.
- **Helps in comparison:** In a data table, the rows and columns which are required to be compared can be placed next to each other. To point out, this facilitates comparison as it becomes easy to compare each value.
- **Economical:** Construction of a data table is fairly easy and presents the data in a manner which is really easy on the eyes of a reader. Moreover, it saves time as well as space.