**Chapter-6**

**Sampling**

**Introduction**

Sampling strategy for data collection is an important decision at the time of designing research. At this stage it is important to decide as to *who* and *how many* would be your subjects from whom the information (data) is to be gathered. In empirical studies the information (data) may be gathered either by *sample* *survey* or *census* depending upon the requirement of study. Census covers all units of population and is a complete enumeration of the population[[1]](#footnote-1) while sample survey covers only part of the total population for information on the topic under study. These data are then extended to make inferences about the whole population.[[2]](#footnote-2) Since it is not possible to gather information from the entire population of interest, it is preferred to create a *representative sample*. For example, if effective implementation of the Protection of Women from Domestic Violence Act, 2005 is to be studied in the state of Haryana; or if the impact of the Sexual Harassment of Women at Workplace (Prevention, Prohibition and Redressal) Act, 2013 is to be studied in private sector in a big city, it will not be feasible to gather relevant information from each and every person who is affected by the said laws, given the time and cost constrains.

Hence the researcher has to select the ‘sample’ for collecting information and make estimates about the population on that basis. Sample if selected properly, serves as the replica of the whole for identifying the characteristic features to understand problems of the whole population under study. A sample is a part of a larger population, taken to be the representative of that population. Sample comprises of subjects chosen as a cross-section of the larger group.[[3]](#footnote-3)

Any socio-legal research is undertaken to address legal issues having social implications and social issues having legal implications. For carrying out such a research it would be relevant to gather information from the general public, police personnel, judges, advocates, accused, suspects, prisoners and other stake holders affected (in a given context) by law or those who are part of the legal system. However, in doctrinal research the relevant information (data) is collected from the legal materials. Thus in doctrinal legal research *information sources* have to be sampled for the purpose of data collection. For example, in a study on capital punishment, the researcher has to decide the number of cases to be studied on capital punishment decided by the court within a specific duration, let us say, between 2000 to 2024. Another example could be to study the pattern of judicial view on rape cases from 1990’s till date. Depending upon the requirement of study, the sample will comprise of all the relevant cases decided by the Hon’ble Supreme Court or the trial courts within this duration for the purpose of analysis and interpretation.

**Meaning of Sample**

In scientific terminology the subset of the whole is known as *sample*. Sample is a small set of cases a researcher selects from a large pool and generalises to the large population.In other wordsthe sample has been defined “as the set of observations taken from the population for the purpose of obtaining information about the population.”[[4]](#footnote-4) A **sample** is the number of individual cases (all elements) that is ultimately drawn and from which/whom data is generated.[[5]](#footnote-5) Thus a “sample” is a portion, piece, or segment that is representative of a whole.

**Characteristics of a good sample as under:**

1. **True-representativeness-** A sample may be comprehensive in traits but may not be a good representative of the population. It is important to have representative sample for credible results.
2. **Generalisability**- Sample should be such that the results of the sample study can be applied, in general, for the universe with a reasonable level of confidence.
3. **Practicability**- A good sample has the practicability for research situation.
4. **Equal Chance to be chosen**-Every individual in the chosen population should have an equal chance to be included in the sample.

**Meaning and Purpose of Sampling**

The process of selecting a manageable number from a delineated and identifiable group (population of interest) by employing scientific measures is known as ‘sampling’. It is a crucial step in designing research. There are two major concerns in sampling – *sampling strategies* and *sampling size*. Sampling addresses the questions of ‘who’ and ‘how many or how much’ data or information is required for any research. Like *who* will be the subjects, respondents, participants, institutions, organisations, individuals, groups; and *how much data* would be enough for making a claim with confidence. To cover the whole study population is not practically feasible. Usually data is collected from a scientifically (statistically in quantitative studies) selected ‘sample’ of relatively small size (as compared to the total population) in order to extrapolate population attributes and make estimates about the whole population on that basis. A good sample makes the research work more accurate and credible.

The quality of data in terms of accuracy and representativeness is dependent upon sampling method. Sampling is integral to the whole research process as it has substantial implications on the quality of research findings in terms of the ability to draw accurate inferences and meaningful research conclusions.[[6]](#footnote-6) However, sampling has been considered to be ‘the messiest part of inferential empirical work’.[[7]](#footnote-7) Depending on the requirement of study the data thus collected may be quantitative or qualitative. “Sampling involves determining and selecting the circle of those to be observed.” [[8]](#footnote-8)

|  |
| --- |
| Sampling is a statistical procedure of drawing a small number of elements from a population and drawing conclusions regarding the population. |

**Sampling Difference in quantitative and qualitative study**

The choice of sampling is not dependent solely on the whims, choice or convenience of the researcher. Rather it is guided primarily by the requirement of study undertaken and is need-driven. There is difference in *sampling strategies* and the *purpose* of sampling in respect of quantitative and qualitative studies.

*Qualitative sample* is not intended to replicate the information from ‘small set of cases’ to the entire population. Rather qualitative sampling is restricted to a limited number of cases from the relevant categories for deeper insight into the larger processes, complex situations, events, or relationships. The qualitative sample brings out distinct features of people, and social settings, by highlighting the key dimensions in the complex social life.[[9]](#footnote-9) . This enables the researcher to make the sample more diverse and rich by including wide range of informants.[[10]](#footnote-10)

*For example*, to study the problem of child marriage in a village, data may be collected from a small number of families (ten to fifteen) who have given their children in marriage or who are likely to marry off their children before legally permissible age. This will give understanding about the reasons for such a practice, the prevailing trend, level of awareness of the people about the law prohibiting child marriage, and the challenges in the implementation of law. Information may also be gathered from the concerned authorities responsible for preventing child marriages to gauge the efficacy of law.

On the other hand in *quantitative sampling*, cases/units are selected as sample which closely represents features of interest in a much larger collection of cases called population.[[11]](#footnote-11) The *quantitative studies* aim at creating a *representative sample* that represents approximate aggregate characteristics of the population, which may be limited to the characteristic/s required for a specific research, *e.g.* sex, age, educational qualification *etc*. [[12]](#footnote-12) Examining and analysing the sample data enables extrapolation of sample results to the entire population with accuracy to a great extent. This is possible only with *properly selected sample* as the representative of the larger population. *For example*, a researcher wants to study the incidence of eve teasing among the female students in different age brackets on the university campus. If the majority of female students are in the age bracket of 22 to 25 years, the sample comprising mainly of girls between 19 to 21 years cannot be considered to be representative of the population. The researcher has to be careful in selecting the sample. Another example can be taken of a study that requires information from men who have suffered because of misuse of Section 85[[13]](#footnote-13) of *Bharatiya Nyaya Sanhita*, 2023 (section 498 A, IPC). For this study, sample comprising of majorly unmarried men will not bring about reliable information. Valid inferences can be drawn about the study population only by studying the representative sample. It is only then the research findings can be replicated to the whole population.

**Difference in sampling between Quantitative and Qualitative Studies**

|  |  |  |
| --- | --- | --- |
| No. | **Quantitative** **study** | **Qualitative** **study** |
| 1 | Focus on maximising statistical *representativeness of population* based on the chosen sample. Quantitative sample represents the characteristics that reflects a targeted population. | Comprehensive/in-depth *representation of a phenomenon of interest.* Qualitative sample is not representative of the larger population. |
| 2 | Generalisations are made for larger group (population) on the basis of inferences drawn from sample data. | The study findings can be applied to other contexts through extrapolation to an extent. |
| 3 | Probability based approach of sampling. | Non-probability based approach of sampling. |
| 4 | * Usually favours large sample size for accuracy * Pre-determined (statistically)   representative sample | * Usually favours smaller sample size * Sample size is dependent on search of saturation point aimed at theoretical completeness rather than maximising the quality of statistical inference. |

**According to Patricia Leavy the Steps in Sampling are as follows:**[[14]](#footnote-14)

**First** step is to determine broadly the *elements/units* in your study, like- person, group, or organisation, in which you are interested.

**Second** step is to identify the *population* (a group of elements/units about which claims will be made later).

**Third** step is to determine the study population, *i.e*. *the sampling frame*.

For example, to study whether the student bodies in all the city colleges comply with Lyngdoh committee guidelines in conducting the student elections or not. The *first* thing to determine is the *element/unit* to be studied. The unit to be studied will be *individual* *college student* studying in the city colleges. *Secondly*, the *population* (taken from the universe comprising of all the college students in the city colleges) to be studied will comprise of all the college students involved in politics (student leaders, office bearers and other executive members) from the *selected* major city colleges. After determining the *study population* (list of students to be procured from the respective colleges)*,* prepare the *sample frame* out of that study population.

**Fundamental Terms in Sampling-**

It is important to understand the important terms in sampling.

|  |  |
| --- | --- |
| **Sample** | A “sample” is the subset of the whole, *a small part of a bigger whole on the basis of which one may know about the characteristic features of the whole*. |
| **Sampling Unit/Element** | *Basic unit about which information is collected and is the basis of analysis*, *eg*, family, head of the family, couple, group (class). |
| **Universe** | *Total of the items or units in any field of enquiry*. |
| **Population** | *Total of items about which information is desired*. The aggregate or the collection of units (individuals, households, prisons, hospitals etc) to be surveyed focussing on a specific characteristic. |
| **Sampling Frame** | The list of units (individuals, households, government institutions). *A complete list or collection from which the sample participants are to be drawn*. It is a researcher’s list to specify the population of interest. |
| **Sample size** | *The number of participants in a sample*. |

**Sampling Unit/Element**

The elements serve as the basis of the analysis and are the basic units about which information is collected.[[15]](#footnote-15)A sample is drawn from a large collection of cases/units. Each case/unit is the sampling element. It is the unit of analysis or a case in a population. It could be a person, a family, offenders, victims, business entities, civil servants, courts, internet service providers, a nation, an organization, a state or a social action (e.g., an arrest, a divorce, or an adoption).[[16]](#footnote-16)

It refers to the level of aggregation of data collected during the subsequent data analysis stage. Some of the examples of different sizes of units are as follows:

**Single person as unit**- For example if the study problem focuses on the legal awareness of employees belonging to an organisation, regarding some service covered under the Consumer Protection law, the unit of analysis would be an individual.

**Two person group- Dyad**- If the study requires ‘two-person interaction’, then the unit will be several two-person groups (known as Dyads) for the purpose of analysis. Let us say a study about gay couples or other live-in partners or married couples etc.

**A group of individuals**- If the study concerns an issue relating to a group, unit of analysis will be at group level. Here it is important to note that the relevant data is to be gathered from all individuals in a group, but the individual data would be aggregated into the group data to see the differences among all the groups. For example, to study any problem at the university level various departments of the University will be taken as separate groups. Data analysis will be done at the department level. Each department is taken as a separate identifiable group as one unit comprising of aggregate of individuals in the department. For the purpose of analysis comparisons will be made treating each department as a separate unit.

Decision about choosing unit of analysis depends upon the questions to be addressed by the study which is further decisive in the selection of the following -

* The data collection methods
* Sample size
* Variables

**Universe and Population**

Even though universe and population are usually used interchangeably but there is a difference in both, as *universe* is wider than the *population*. From a statistical point of view, the term ‘*universe’* refers to the total of the items or units in any field of enquiry, whereas the term ‘*popu­lation’* refers to the total of items about which information is desired. Thus, all units in any field of enquiry constitute universe and all ele­mentary units (on the basis of one characteristic or more) constitute population.[[17]](#footnote-17) Let us take an *example*, if a researcher wants to study mentally ill undertrial prisoners in the jails of Sate of Haryana, the aggregate of all the prisoners in these jails will constitute the universe. While the population of the study as defined by the research question would be limited to the aggregate of all the mentally ill undertrial prisoners in these jails. In statistics, information is often inferred about a population by studying a finite number of individuals from that population, *i.e.,* the population is sampled, and it is assumed that characteristics of the sample are representative of the overall population.[[18]](#footnote-18)

In simple words all the individuals of interest to the researcher are referred as the population or the universe of study.[[19]](#footnote-19) Population is the abstract idea of a large group of many cases from which a researcher draws a sample and to which results from a sample are generalised. The population may be finite or infinite. An example of infinite population could be the number of drug addicts in a city, because there is no way to identify this number. While the number of drug addicts in a city taking treatment in government hospitals or rehabilitation centres would be an example of finite population.

To define the population, it is required to specify the elements and identify its “geographical” and “temporal boundaries” as well as any other relevant boundaries.[[20]](#footnote-20) Population includes *persons or things* that fit in the previously defined selection criteria, about which the conclusions are to be made. [[21]](#footnote-21) Due to affordability and time considerations the data analysis of the sample is used to make estimate or prediction about the whole population out of which the sample has been drawn.

**Sampling Frame**

A sampling frame is a list of observational units to be sampled. A good sampling frame is crucial for accurate sampling. The sampling frame refers to the list of units (eg, persons, households, businesses, etc) in the survey population. The sampling frame is important because the selection of sample is drawn from this list. It reflects the coverage of the target population and is also decisive of the choice of the data collection method.[[22]](#footnote-22) Only those out of the sampled population who have consented for the study will be included in the sampling frame. Therefore, *the sampling frame comprises of a list of the entire sampled population who are alive and available and have given their consent*. Sampling frame is also known as source list. The sampling frame should be comprehensive and should contain authentic information about the elements or units to be studied. The frame should contain contact points (relevant information like phone number, address) for each of the units to help in accessing the population.[[23]](#footnote-23)

To construct a Sampling Frame requires-

* Enumeration of every individual in the sampled population
* Attaching an *identifier* to each individual

(usually, the position on the list serves as this identifier)

**Sample size**

Technically, the size of the sample depends upon the precision the researcher desires in estimating the population parameter at a particular confidence level. There is no rigid rule for the determination of sample size. However, a larger sample is much more likely to be representative of the population, and also more accurate and precise. *Sample size is defined as the number of individuals included in research study to represent a population.* Sample size comprises of the total number of respondents included in a study, and is further broken down by demographics such as age, gender, location so that the sample represents the entire population. Determining the appropriate sample size is one of the most important factors in statistical analysis.[[24]](#footnote-24)

Sample size refers to number of participants or other units needed to address research questions.[[25]](#footnote-25) Following factors may be taken into consideration while determining the appropriate sample size[[26]](#footnote-26):

* Number of cases required to answer your research questions or hypotheses.
* The availability of resources (monetary and time).
* Use of research method or methods.
* Heterogeneity or homogeneity of the population of interest.

In case of heterogeneous population with multiple variables, a large sample may be required.[[27]](#footnote-27)

* Evaluation of high variation in outcomes of the underlying population, requires a larger sample.

Even though the procedure of selecting a sample differs according to the type of the sample selected, certain fundamental rules remain the same. These include:

1. The universe and population must be defined precisely;
2. The size of the sample to be selected should be pre-determined;
3. The unit of the sample should be defined.
4. The appropriate source list/sampling frame should be prepared before-hand in case it does not already exist.

**Types of Sampling Techniques**

There are a number of alternative techniques for drawing a sample. These techniques can be broadly grouped into two categories- *probability sampling* and *non-probability* or *purposeful sampling*. These general categories of sampling bear different strengths to suit specific needs of any research project based on its objectives.

The table below illustrates the main differences between probability and non-probability sampling methods. [[28]](#footnote-28)

|  |  |
| --- | --- |
| **Probability Sampling** | **Non-probability Sampling** |
| Random or chance sampling | Purposive, judgement, deliberate sampling |
| Equal chance for each member of population to get selected | Not everyone has equal chance to get selected |
| Useful for collecting data from diverse population | Useful in specific environment with sampling group members sharing similar characteristics |
| Useful for inferential generalisations about the population | No generalisations based on observations about the sample |
| Complex but more accurate in terms of representative sample | Simpler but less accurate alternative for getting representative sample |
| Types   * Simple random sampling * Systematic sampling * Stratified sampling * Cluster sampling | Types   * Quota sampling * Cluster sampling * Convenience sampling * Purposive/judgemental sampling |

Notwithstanding these key differences, non-probability (purposeful) and probability sampling techniques can be combined usefully. Depending on the research objectives, the method to be used for sampling can differ.

**Probability Sampling**

The essence of probability sampling technique lies in *randomization* or *chance.* Random selection is likely to yield a truly representative sample of the population. Probability sampling is regarded as the “gold standard” for creating a representative sample.[[29]](#footnote-29) Each member of the population stands an equal chance of being selected for the study. Sample is drawn from a large

collection of cases/units.

This sampling approach is complex and tedious and has several subtypes.

*Methods of Drawing Random Samples*

1. Fishbowl Draw Method/lottery
2. Using Table of Random Numbers
3. Method of computer-determined Randomness

*Fishbowl Draw Method/lottery*

It is the simplest method used to select elements from population in a random manner. The entire elements of the population are numbered on the paper slips, folded and put in a bowl or a container. All the slips are same in all respects and are mixed up thoroughly. The researcher thereafter randomly picks up one number at a time till the number reaches the desired sample size.

*Using Table of Random Numbers*

Equal chance of being included in the study can be achieved by assigning a number to each sample unit and selecting members of the sample by using a random number table.[[30]](#footnote-30) The quantitative researchers employ these techniques for sampling. The table of random numbers consists of a continuous row-column sequence of numbers which do not appear in any particular sequence, nor does any number appear more frequently than the other one. After specifying the number of elements in the population, the researcher numbers them from one to N (total number of elements in the population). Let us say the population size is 300 and the researcher intends to select 30 cases randomly from this population. Then enter the table at any point and move systematically to any direction- right, left, up, down, diagonally. Continue moving on systematically through the table until 30 elements have been systematically selected.

*Computer-determined Randomness*

This method is preferred in case of large population size. The data is fed in the computer to obtain a random number of elements corresponding to the elements in population.

***Characteristics of Probability Sampling-***

**One**, probability sample may be representative of the population.

**Two**, the observations (data) obtained by employing probability sample are used to draw inferences for making generalisations about the population.

**Three**, every member of the population has an *equal* and *non-zero* chance of being selected for the study sample.

Simply put, *“the chance that each element in the population will be included in the sample can be statistically determined, and the chance of inclusion, no matter how small, will be a number above zero. Each element has some chance of inclusion.”*[[31]](#footnote-31) In other words they cannot have “no chance” of being sampled.

**Steps to be followed in probability sampling:**

|  |
| --- |
| For example, to study the **population** of all law students on a campus-  At first the researcher must decide the **parameters** to define the type of students to be included in the study, like undergraduate and postgraduate students; local and international students; hostellers and day scholar students *etc*.  After deciding on the actual population under study or the **sampling frame**, next task is to prepare a list of students who fulfil the criteria.  **Final step** is to identify a method/technique for the selection of students to be invited to participate in the study. |

Adapted from Wing Hong Chui, *Quantitative Legal Research,* Research Methods for Law 55 (Mike McConville & Wing Hong Chui ed. 2007).

**Types or Techniques of Probability Sampling:** There are a number of techniques for taking probability sample. Four important techniques have been discussed as follows:

1. Simple random sampling.
2. Systematic sampling.
3. Stratified sampling.
4. Cluster sampling.
5. *Simple Random Sampling (SRS)*

SRS is the basic and most common random sampling strategy in which every element in the study population has an equal chance of being selected. It is considered to be the best selection procedure to avoid sample bias. The method requires developing an accurate sampling frame, selection of elements from the sampling frame according to mathematical random procedure, and then locate the exact element that was selected for inclusion in the sample.

|  |
| --- |
| After having defined students’ type, let us assume that there are 1000 law students in the sampling frame.  The researcher decides to choose a sample of 200 out of the sampling frame of 1000 students.  Each student on the list of 1000 students may be assigned a unique number and 200 numbers can be randomly generated either by a computer, or the researcher may use a list of random numbers[[32]](#footnote-32) to select the elements. |

Example: Wing Hong Chui, *Quantitative Legal Research,* Research Methods for Law 55 (Mike McConville & Wing Hong Chui ed. 2007).

SRS serves as a foundation upon which all other types of random sampling are based because this method of sampling can be readily applied in conjunction with all other probability methods. Random sampling cannot guarantee perfect and fool-proof representativeness of the population but brings it close to the population. Random sampling is not ideal but it relies on statistical probability to ensure the representativeness of the sample. This means that to ensure representativeness a relatively large sample would be required.[[33]](#footnote-33)

One of the major drawbacks of the SRS is that it does not ensure inclusion of elements in the given sample that exist in small numbers in the population. Suppose in a population of 1000 people only 10 persons possess trait ‘X’. Considering the researcher draws a sample of 50 persons out of 1000 people, the chances are very grim that the persons with trait ‘X’ are included in the sample even if the researcher wanted.

This method may take two forms-*Sampling with replacement* and *sampling without sampling*. *Sampling with replacement* -For example to select a sample of 20 students from LL.B. first year consisting of 60 students in a class, write the roll numbers/names of all the students on a slip and put them in a box. Pick up one slip at a time, note down the name and again put it back and mix the numbers thoroughly, so on and so forth till the required number of elements has been selected. If the same name is repeated, it may be ignored, instead another slip is taken out. In such a case the probability will remain the same for each element to be selected even till the end, i.e., 1/60 as the total number of slips remains the same out of which selection is to be made. In case of sampling with replacement , chance of the same case being selected more than once is increased.

*Sampling without replacement*- Once a slip is selected it is not to be put back again. This makes the probability of the second slip 1/59.

1. *Systematic Random Sampling*

Systematic sampling involves “choosing samples in a systematic pattern” by taking every *n*th element in the sampling frame until the total is reached. According to this sampling strategy the first element in the study population is selected randomly and then every *n*th element, after the first element, is selected.

**Steps to be followed in Systematic Random Sampling**

* The very first step is to number each element in the sampling frame.
* Then the sampling interval is to be calculated, that serves as the basis of randomisation/random selection method to select the sampling interval.

Unlike simple random sampling, the sampling interval is systematically calculated.

|  |
| --- |
| If in a sample frame of 1000 students, sample of 200 students is to be selected, this will be done by dividing 1000 by 200, the sampling interval comes out to be five. With five as the sampling interval; the researcher may begin with any random number as a starting point, for example starting from # 26 on the list. This is to be followed by 31, 36, 41, 46 and so on till the number so selected reaches 200. |

Example: Wing Hong Chui, *Quantitative Legal Research,* Research Methods for Law 55 (Mike McConville & Wing Hong Chui ed. 2007).

In the above examples, the process involved in random and systematic sampling is random, leading to a possibility, though unlikely, that the sample drawn will contain males only. Also, it may be possible that smaller sub-groups such as international students are under-represented or excluded.

*Stratified Random Sampling*

In case of heterogeneous population the stratified random sampling would prove better in producing representative sample. It involves the division of the sampling frame into groups in order to ensure the sample to be representative.[[34]](#footnote-34) In Stratified Random Sampling all the elements in the study population are divided into two or more groups based on a shared characteristic. This division into groups may be based on single criterion such as gender, yielding two strata- male and female, or may be upon a combination of two or more criteria such as gender and graduation, yielding four strata, namely, male undergraduates, male graduates, female undergraduates, female graduates. Stratification divides the sampling frame into subgroups, subpopulations or *strata*, which are non-overlapping, “mutually exclusive and also exhaustive”[[35]](#footnote-35) and together constitute whole population.[[36]](#footnote-36) After dividing the whole population into two or more strata (internally homogeneous) a simple random sample of the desired number is drawn from each stratum.

For example, if the researcher identifies *gender and class* as important variables, the population would be divided into working-class males, working-class females, middle-class males, middle-class females, upper-class males and upper-class females. The sample would then be selected at random from each of these groups. It is to be further ensured that the proportions of the sample in each category were the same as the proportions in the population as a whole. If 20 percent of the population were found to be working-class females, 20 percent of the sample would be working-class females. [[37]](#footnote-37)

**Steps to be followed in Stratified Random Sampling**

|  |
| --- |
| First step is to identify the variables/characteristics that are likely to affect the results.  Second step is to stratify the population to make separate homogeneous groups based on the shared characteristics- called *strata*.  After this initial exercise of forming the *strata*, conduct any of the simple random, systematic, or cluster sampling on each *strata*. |

One advantage of this sampling is that it ensures better coverage of the population than random and systematic sampling. However, this procedure requires greater effort in defining strata and identifying the characteristics of each stratum.[[38]](#footnote-38)

Example- suppose the investigator divides the population of 10,000 individuals into 6,000 males and 4,000 females. If she intends to draw a sample of 1000 individuals from the set of 10,000 and randomly draws both the males and females in equal numbers, say 500 each; it will constitute the example of disproportionate stratified sampling. However, if 600 males and 400 females are drawn in the sample, it will be an example of proportionate stratified sampling.

*Cluster Sampling*

Cluster sampling is another method of probability sampling. This is a multistage sampling strategy. Firstly, pre-existing clusters are randomly selected from a population. Next, elements in each cluster are sampled (in some cases, all elements in each cluster are included in the sample). For example, if your population is all students from the Panjab University Law institutions/colleges (affiliated to the university) which participate in legal aid clinics, you might get a list of all the institutions/colleges in the region that run legal aid clinics. Then by following random selection several such institutions/colleges are selected—each serving as a cluster. The students involved in legal aid clinics at those institutions/colleges would comprise your sample.[[39]](#footnote-39)

This sampling method has utility where large geographical area is to be covered because the researcher need not have the list of individuals in a given area. The clusters of elements are sampled and not the individuals, thus the respondents can be readily substituted for other respondents within the same random section. Cluster sampling possesses the trait of flexibility.

**Sampling error** and **Sampling bias**

There are three concepts relevant to probability samples - sampling error, random sampling, and sampling bias[[40]](#footnote-40) and should be taken care.

Sampling error occurs in all probability samples and is unavoidable because no sample can ever totally represent the population. The sampling error is bound to occur due to almost inevitable gap between a sample’s representativeness and the population’s known or unknown characteristics. *The size of the sampling error generally decreases as the size of the sample increases.* Sampling error may result from faulty sample frame, non-response (change of address/mobile number, death), lack of population specification.

Sampling bias may occur because of flaw in *sample selection process*. It is systematic, and increasing the size of the sample just increases the effect of the bias. Sampling bias occurs when the sample is not representative of the population. A systematic bias occurs if some members of the population are systematically more likely to be selected in a sample than others, or when the research is conducted on a group of participants who do not actually represent the population, or when members of intended population are selected incorrectly – either because they stand a higher or lower chance of being selected. Some of the examples of sampling bias are- In a study on the violations of Motor Vehicle Law, students of 10th grade are included as participants, who are legally not competent to drive and have never driven .

**Non-probability Sampling Techniques**

Non-probability sampling techniques are also known as *purposeful, purposive or judgement* sampling. Each element in the population does not have any probability of being chosen as sample subject. The sample group members are selected non-randomly; therefore, only certain members of the population have a chance to participate in the study.[[41]](#footnote-41) These techniques are a less accurate substitute when representative sample is required. This is a strategic approach to sampling in which “information-rich cases” are sought out in order to best address the research purpose and questions.[[42]](#footnote-42) Purposeful sampling strategies are typically used in qualitative studies that require in-depth understanding from a small sample.

Characteristics of Non-probability Sampling: The following are the main characteristics of non- probability sample:

1. There is no idea of population in non-probability sampling.
2. There is no probability of selecting any individual.
3. The observations of non-probability sample are not used for generalization purpose.

There are a number of non-probability sampling techniques namely, convenience sampling, quota sampling, snowball sampling, purposive sampling,

*Convenience Sampling*

Convenience samples are most convenient and economical but at the same time least reliable are not representative of the population. The main criteria for identifying the participants is the easy *accessibility*, *availability* or *convenience* of the participants. Street-corner interviews and phone-in polls are the examples of convenience sampling. Convenience sampling is also called *accidental, availability, or haphazard* sampling. A convenience sample is a potential source of participants that is easily accessible to the researcher. [[43]](#footnote-43) This approach is often used when the researcher has access to subjects within a particular institution, organization, business, group, and so forth. [[44]](#footnote-44)

This sample type may be suited for a few exploratory preliminary studies where some basic information is quickly required. It may also be helpful in some qualitative research studies when the purpose is something other than creating a representative sample. For example interviewing people visiting the mall about any cyber frauds with them, or interviewing female students about eve teasing outside the school after the school gets over for data collection. Another example could be that if you are interested in studying how female students in the university experience sexual harassment and, you may begin your sampling process in your own department and a few nearby departments, because of the convenience and accessibility of the respondents, it will not reflect the actual position of the problem at the university level. Accidental samples are not apt for making generalisations for the entire population with reasonable confidence. There remains a scope for the researcher’s bias in the selection of individuals.

*Quota sampling*

Quota sampling allows the researcher to control variables without having a sampling frame. [[45]](#footnote-45) Quota sampling is one of the most important types of non-probability sampling and has some apparent similarity with the stratified random sampling. It is a sampling procedure that gives choice and liberty to the researcher to identify certain characteristics of a population of interest and take decision about their overall presence in the population. This is followed by selection of cases as the representative of each of the relevant characteristics proportionate to their representation in the population.[[46]](#footnote-46)

|  |
| --- |
| **Quota sampling**  “A non-random sample in which the researcher first identifies general categories into which cases or people will be placed and then selects cases to reach a predetermined number in each category.” [[47]](#footnote-47) |

This is done by at first identifying relevant categories among the population being sampled to capture diversity among units. For example, male and female; or below 18 years of age, above 50 years of age educated and uneducated *etc*). Next is to determine the “quota” for each category by deciding the number of cases to be included in each category. Thus, the number of cases in various categories is fixed at the very beginning.[[48]](#footnote-48) Quota sampling requires a pre-set number of cases in each of several predetermined categories so as to reflect the diversity of the population by using haphazard methods. For example, for a sample of 50 persons the researcher decides to select 10 males and 10 females below 18 years, 10 males and 10 females above 50 years and 5 educated males and 5 uneducated females. Once quota has been fixed, the researcher may use convenience sampling thereafter. Another example can be taken, where the researcher is required to administer questionnaire to ten married females and ten married males aged between 20 and 35, five unmarried men and women of the same age group, and five divorced men and divorced women of the same age group. Once the pre-decided quota has been filled, thereafter no more responses from those categories will be collected.

There are three weaknesses of quota samples:

*First*, they are limited in scope and capture only a few aspects (e.g., gender and age) of all population diversity and ignore others (e.g., race-ethnicity, area of residence in the city, income level).

*Second*, the fixed number of cases in each category may not accurately reflect the proportion of cases in the total population for the category. Perhaps 20 percent of city residents are over 60 years old but are 10 percent of a quota.

*Lastly*, convenience sampling is used for the selection of each quota category. Due to convenience sampling there is the likelihood that the first twenty-five males under age 30 whom the researcher encounters are included, are highly educated; or may be the people who are friendly are more likely to be selected in the quota.[[49]](#footnote-49)

*Snowball Sampling*

Snowball sampling is a very specialised type of sampling and are usually used when other methods do not seem to be practical. [[50]](#footnote-50) This type of sampling has been regarded as most suitable where there is no sampling frame, where people to be studied are rare or widely spread, and where the people of interest are likely to know each other, expected to facilitate the researcher find more contacts. [[51]](#footnote-51) Snowball sampling is also called *chain sampling or chain referral, network, reputational sampling.* In snowball sampling one case leads to another by way of referrals provided by the participants that helps in getting more participants who may give additional information important for the study.Subjects with desired traits or characteristics give names of further appropriate subjects.[[52]](#footnote-52) It is a form of sequential sampling. It begins with a few participants, and the number goes on increasing with more and more persons being referred by the earlier participants and thus the chain of referrals continues. Sometimes the researcher seeks to study the ‘latent’ population which is not easily identifiable. For example- identifying drug addicts, prostitutes, hardened criminals *etc* is usually very difficult. This sampling design has been found to be very useful in such cases. For example, in a study on the violation of human rights of LGBTQ, it would be difficult to identify and locate such people. If the researcher locates 2-3 such people and each one of them gives 3 to 4 referrals, it will make the work of researcher easier in identifying more participants. When some names are repeatedly given, or no new names are emerging, that would be an indication to stop looking for more participants.

|  |
| --- |
| **Snowball sampling**  A non-random sample in which the researcher begins with one case and then, based on information about interrelationships from that case, identifies other cases and repeats the process again and again. [[53]](#footnote-53) |

*Purposive or Judgmental Sampling*

Purposive sampling (also known as judgmental sampling) is a non-random sampling method. It is used in exploratory research or field research. The method for selection of cases is either guided by the judgment of an expert, or is driven by a specific purpose in mind. The sample selected by this method would rarely be representative of the entire population. This method is appropriate to select unique cases that are informative, including people who are difficult-to-reach. [[54]](#footnote-54)

|  |
| --- |
| **Purposive sampling**  A non-random sample in which the researcher uses a wide range of methods to locate all possible cases of a highly specific and difficult-to reach population. [[55]](#footnote-55) |

The main focus of purposive sampling is to concentrate on people with particular characteristics who will better be able to assist with the relevant research. In this method “information rich cases” are selected which are best suited to address the research objectives and questions.[[56]](#footnote-56) This involves identification and selection of individuals on the basis of expertise, selecting people who are proficient and well-informed with a phenomenon of interest; or selection of people on the basis of specific purpose in mind. [[57]](#footnote-57) Besides knowledge and experience, the availability and willingness to participate are important factors in the selection of sample. Usually the researcher lays down the criteria for the subjects to be included in the sample. Whoever meets the criteria could be selected in the sample.

In order to carry out study about the banking frauds, bank officials would be in a position to give valuable input. According to this sampling strategy the selection of respondents is deliberate, best suited to provide the desired information about the phenomena of interest under study. Sample selection is dependent upon their ability, knowledge, experience, and willingness to explain the specific phenomena or issue.

In a study on prostitutes the researcher has to rely on the local knowledge to locate prostitutes (*e.g*., locations where prostitutes solicit, social groups with whom prostitutes associate) and local experts (*e.g*., police who work on other prostitutes) to locate possible prostitutes for inclusion in the research project. A researcher will use many different methods to identify the cases because the goal is to locate as many cases as possible. [[58]](#footnote-58)

**Conclusion:**

Qualitative data is best suited for most of the socio-legal studies that relate to various aspects of ‘society-law’ interaction. Such studies are important for deeper understanding of the changing social reality keeping up with the dynamism of law and its implications. In order to know and understand about the social/economic implications of law or to assess the need for a new law, or for amending the law; interface between law and society has to be studied. This requires collecting relevant information from various stakeholders.

For any scientific inquiry about a phenomenon/research interest, relevant information is to be gathered by employing different strategies. The nature and objectives of study are decisive in selecting the method for data collection. Quantitative studies are driven by representativeness of the sample that is achieved through randomisation. While qualitative studies are judgement and convenience driven and follow a non-random process to include diversity of information. With this fundamental difference, the sampling methods can be broadly categorised into probability and non-probability sampling methods suitable to cater the specific needs of a study. The use of probability sampling methods used in quantitative studies enable the researcher to make estimates about the population. The greater sample size may result in greater accuracy in drawing inferences about the population. On the other hand non-probability sampling strategies focus on the diversity of the information, which is not intended to be representative of the population. Hence the size of the sample does not affect the outcome of the study.

|  |
| --- |
| **Takeaways:**   1. ‘Sampling’ is a crucial step in designing research. 2. There are two major aspects of sampling – *sampling strategies* and *sampling size*. Sampling addresses the questions of ‘who’ and ‘how many or how much’ data or information is required for any research. 3. Major steps in sampling include-  * Identify *elements/units* in your study, like- person, group, or organisation, in which you are interested. * Identify the *population* (a group of elements/units about which claims will be made later). * Preparing *the sampling frame* from the population.  1. Sampling techniques are broadly categorised as- *Probability sampling* and *Non-probability sampling.* 2. In probability sampling -  * the sample is representative of the population of interest. * the observations (data) obtained by employing probability sample are used to draw inferences for making generalisations about the population. * every member of the population has an *equal* and *non-zero* chance of being selected for the study sample.  1. In non-probability sampling-  * There is no idea of population in non-probability sampling. * There is no probability of selecting any individual. * The observations of non-probability sample are not used for generalization purpose.  1. *Types of probability sampling*  * Simple random sampling. * Systematic sampling. * Stratified sampling. * Cluster sampling.  1. *Types of non-probability sampling*    * Quota sampling    * Cluster sampling    * Convenience sampling    * Purposive/judgemental sampling |

**Suggested Questions:**

1. Discuss which sampling methods should be employed in quantitative and qualitative studies respectively and why?
2. What do you mean by sampling? What are the factors that influence the decision to sample?
3. Explain the importance of sampling in research. What are the requisites of a good sample? What is the difference between probability and non-probability sampling?
4. Discuss the advantages and disadvantages of probability sampling methods.
5. Discuss the advantages and disadvantages of non-probability sampling methods.
6. What are the requisites of a good sample? Explain the advantages of randomising in scientific research.
7. Explain the difference between stratified random sampling and simple random sampling.

1. Australian Bureau of Statistics, https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Basic+Survey+Design+-+Samples+and+Censuses (last visited Aug. 07, 2022). [↑](#footnote-ref-1)
2. *Id.* [↑](#footnote-ref-2)
3. HARALAMBOS & HOLBORN, SOCIOLOGY, THEMES AND PERSPECTIVES, 911 (8th ed. 2020). [↑](#footnote-ref-3)
4. Australian Bureau of Statistics, *supra* note 1. [↑](#footnote-ref-4)
5. Patricia Leavy, Research Design 76, (2017)*.* [↑](#footnote-ref-5)
6. Summary, https://oxfordre.com/business/view/10.1093/acrefore/9780190224851.001.0001/acrefore-9780190224851-e-216#acrefore-9780190224851-e-216-div1-1 [↑](#footnote-ref-6)
7. Frans L. Leeuw With Hans Schmeets, Empirical Legal Research, A Guidance Book For Lawyers, Legislators And Regulators 130 (2016). [↑](#footnote-ref-7)
8. Imre Boncz, Introduction to Research Methodology 25 (2015). [↑](#footnote-ref-8)
9. W. Lawrence Neuman, Social Research Methods: Qualitative and Quantitative Approaches 247 (7th ed., 2014). [↑](#footnote-ref-9)
10. Editorial, *If you could just provide me with a sample: examining sampling in qualitative and quantitative research papers*, 2:3 EBN 70 (July 1999), https://ebn.bmj.com/content/ebnurs/2/3/68.full.pdf (last visited July 15, 2022). [↑](#footnote-ref-10)
11. Neuman, *supra* note 9. [↑](#footnote-ref-11)
12. Boncz, *supra* note 8. [↑](#footnote-ref-12)
13. Section 85, Bharatiya Nyaya Sanhita, 2023 makes it a penal offence against husband or his relatives for subjecting a woman to cruelty. [↑](#footnote-ref-13)
14. Leavy, *supra* note 5. [↑](#footnote-ref-14)
15. Boncz, *supra* note 8. [↑](#footnote-ref-15)
16. Neuman, *supra* note 9, at 250. [↑](#footnote-ref-16)
17. https://www.yourarticlelibrary.com/agriculture/7-fundamental-concepts-pertaining-to-sampling/44386 [↑](#footnote-ref-17)
18. https://www.calculator.net/sample-size-calculator.html (accessed on 13 July 2022) [↑](#footnote-ref-18)
19. Geoffrey Marczyk et al., Essentials of Research Design and Methodology 18 (2005). [↑](#footnote-ref-19)
20. # Neuman, *supra* note 9, at 250-51.

    [↑](#footnote-ref-20)
21. Boncz, *supra* note 8. [↑](#footnote-ref-21)
22. Australian Bureau of Statistics, *supra* note 1. [↑](#footnote-ref-22)
23. *Id.* [↑](#footnote-ref-23)
24. Frankline Kibuacha, *How to Determine Sample Size for a Research Study*, https://www.geopoll.com/blog/sample-size-research/ (last visited Jul. 13, 2022). [↑](#footnote-ref-24)
25. # Timothy C. Guetterman, Qualitative, Quantitative, and Mixed Methods Research Sampling Strategies, *Oxford Bibliographies*, last modified Feb. 26, 2020, https://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0241.xml.

    [↑](#footnote-ref-25)
26. Leavy, *supra* note 5, at 77-78. [↑](#footnote-ref-26)
27. Boncz, *supra* note 8, at 27. [↑](#footnote-ref-27)
28. *Business Research Methodology*, https://research-methodology.net/sampling-in-primary-data-collection/probability-sampling/ (last visited Aug. 28, 2024). [↑](#footnote-ref-28)
29. Neuman, *supra* note 9, at 250. [↑](#footnote-ref-29)
30. Haralambos, *supra* note 3. [↑](#footnote-ref-30)
31. Leavy, *supra* note 5, at 78. [↑](#footnote-ref-31)
32. Random numbers can be taken from random number table, a table of numbers chosen in a mathematically random way. [↑](#footnote-ref-32)
33. HARALAMBOS, *supra* note 3. [↑](#footnote-ref-33)
34. HARALAMBOS, *supra* note 3, at 912. [↑](#footnote-ref-34)
35. Wing Hong Chui, *Quantitative Legal Research,* Research Methods for Law 55 (Mike McConville & Wing Hong Chui ed. 2007). [↑](#footnote-ref-35)
36. A.K.Singh, Tests, Measurements and Research Methods in Behavioural Sciences 347 (2017). [↑](#footnote-ref-36)
37. HARALAMBOS, *supra* note 3, at 912. [↑](#footnote-ref-37)
38. Chui, *supra* note 35. [↑](#footnote-ref-38)
39. Leavy, *supra* note 5, at 79. [↑](#footnote-ref-39)
40. Editorial, *supra* note 10, at 68-69. [↑](#footnote-ref-40)
41. https://research-methodology.net/sampling-in-primary-data-collection/probability-sampling/ [↑](#footnote-ref-41)
42. Leavy, *supra* note 5, at 80. [↑](#footnote-ref-42)
43. Marczyk et al., *supra* note 19 at 83. [↑](#footnote-ref-43)
44. Leavy, *supra* note 5, at 110. [↑](#footnote-ref-44)
45. Haralambos, *supra* note 3, at 912. [↑](#footnote-ref-45)
46. Leavy, *supra* note 5, at 149. [↑](#footnote-ref-46)
47. Neuman, *supra* note 9, at 249. [↑](#footnote-ref-47)
48. *Id*. [↑](#footnote-ref-48)
49. # *Id*.

    [↑](#footnote-ref-49)
50. Haralambos, *supra* note 3, at 913. [↑](#footnote-ref-50)
51. *Id*. [↑](#footnote-ref-51)
52. Chui, *supra* note 35, at 56. [↑](#footnote-ref-52)
53. Neuman, *supra* note 9, at 275. [↑](#footnote-ref-53)
54. *Id.* at 273-74. [↑](#footnote-ref-54)
55. *Id.* [↑](#footnote-ref-55)
56. Leavy, *supra* note 5, at 79. [↑](#footnote-ref-56)
57. Neuman, *supra* note 9, at 274. [↑](#footnote-ref-57)
58. *Id.* [↑](#footnote-ref-58)