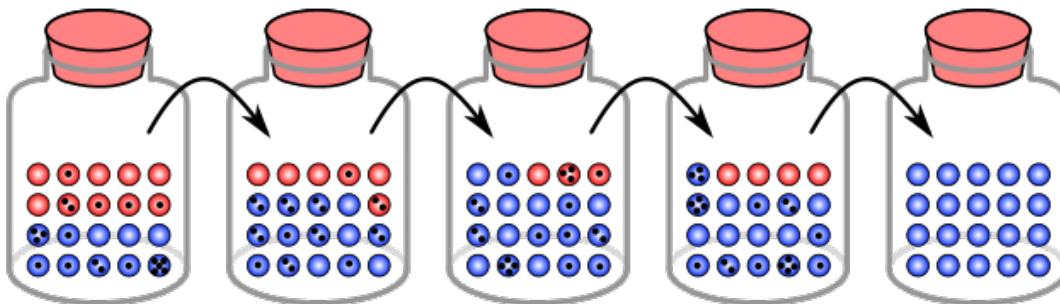


Statistics: Randomness and Survey

Samplings

[See online here](#)

This article discusses the randomness of a sample for selection in a population and survey sampling method and techniques. The types of sampling include simple random sample, sampling frame, stratified sampling, cluster sampling, multi-stage sampling, and systematic sampling. The article also explains common mistakes in sampling survey.



Randomness

Random sampling refers to an **unpredictable selection of samples** where each unit of value in a data set has an equal probability or a chance of selection. No set prediction can define which sample will be chosen exactly. Its selection has to be simple and fair, and every sample should be given equality of selection.

In a random sample, the outcomes of a few separate draws are **independent** of each other, and do not impact on the selection of other samples. Random sampling is a hard technique to follow unless performed through specific software, such as R, Excel or SAS.

Example: Each unit or member of a data set is assigned a specific number which helps in the selection of numbers randomly. Suppose 150 members are taken as a population for the lottery of a vacation ticket in a corporate office; 10 tickets are available and the lottery method has to be used. The lottery method is the way to choose any number out of the 150 employee names written on paper and accumulated in a big box, and now the box will be rotated to mix the papers well and randomly 10 papers will be chosen. In this method, each of the 150 employees has an equal chance of selection.

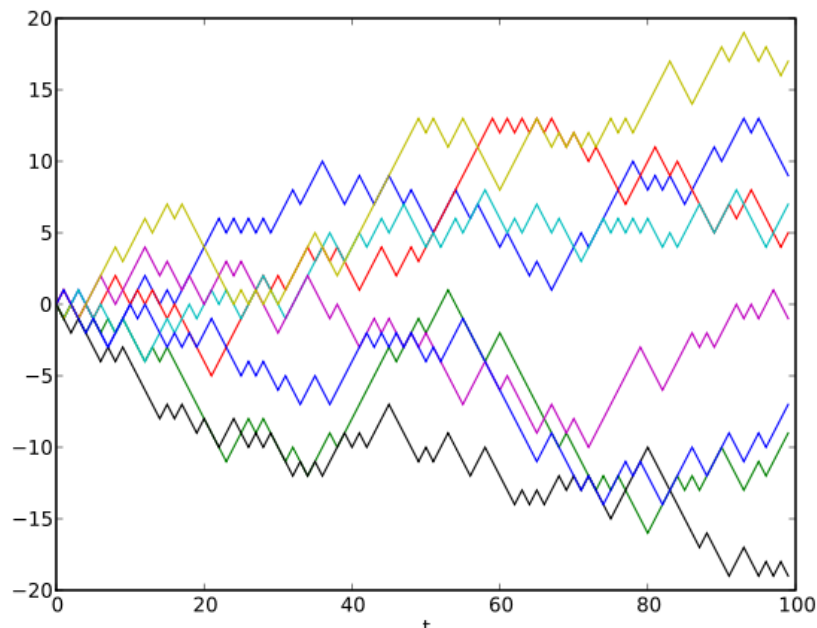


Image: "Eight different random walks." by Morn created this work entirely by him-/herself. License: [GFDL](#)

Sample Surveys

The sample surveys are conducted to get an insight into a small group of people on average and consider that it will serve as information about the whole population.

Sampling takes on two forms in statistics: probability sampling and non-probability sampling:

- **Probability sampling** uses random sampling techniques to create a [sample](#).
- **Non-probability sampling techniques** use non-random processes like researcher judgment or [convenience sampling](#).

Probability sampling is based on the fact that **every member of a population has a known and equal chance of being selected**. For example, if you had a population of 100 people, each person would have odds of 1 out of 100 of being chosen. With non-probability sampling, those odds are not equal. For example, a person might have a better chance of being chosen if they live close to the researcher or have access to a computer.

Probability sampling gives you the best chance to create a sample that is truly representative of the population.

The goal of sampling is to examine a **part of the whole population** to acquire results and to learn about the whole population. The problem with sampling is that results are normally not very accurate, and it is difficult to infer that results of a small sample can be projected to the whole population precisely.

Types of Probability Sampling

Some of the common types of sampling include:

Simple random sample

Random sampling refers to an unpredictable selection of sample where each unit of value in a data set has an equal probability or a chance of selection. No set prediction can define which sample will be chosen exactly. [Simple random](#) sampling is a completely random method of selecting subjects. These can include assigning numbers to all subjects and then using a random number generator to choose random numbers.

Classic [ball and urn experiments](#) are other examples of this process (assuming the balls are sufficiently mixed). The members whose numbers are chosen are included in the sample.

Consider selecting a sample of size n . If this sample is drawn so that *every possible* sample of size n has the same chance of being selected, it is said to be a simple random sample (SRS).

Example: Suppose we have a piece of land and we want to estimate the volume of timber or the number of woodpecker nests on the piece of land. A census might be too costly. One simple way to take a sample might be to divide the area into equal-sized blocks.

The blocks should be small enough to survey reliably. Suppose the area is divided into 36 blocks and we've decided to survey a sample of 9 blocks. To select an SRS, label the blocks in any order. Go to the random number table and select a row at random" to generate the sample. For example, suppose we choose row 7:

Starting at row 7, select an SRS of 9 plots, and mark them on the drawn blocks.

In an SRS, every combination of 9 blocks has the same probability of being selected. Selecting an SRS does not guarantee that the particular sample selected is perfectly representative of the population.

It is not the sample you select which is unbiased; it's the procedure by which the sample is selected which is unbiased. If we were selecting an SRS from an alphabetical list of 36 people, we probably wouldn't worry that the names weren't evenly distributed through the list, since we have no reason to believe that the variable being measured (e.g.: their opinion on some issue) is associated with their position on the list.

Sampling frame

A sampling frame is the list of units, members or individuals from which a population has to be selected. The simplest way to select a sample to generate results for the whole population is through simple random sampling (SRS).

Samples which are selected randomly differ from the other samples. Each newly drawn sample has a different value of variables. The difference between samples is termed as **sample variability**. Sample variability is a normal part of the sampling technique; it does not infer that the sample does not represent the entire population.

Stratified sampling

When **samples are drawn from more than one group of the population**, it is known as stratified sampling. In case sample variability is high and more complicated, sampling is required to draw a sample which represents the whole population, so sampling is used. Each **homogeneous group from a population** is known as "**strata**." Samples from different strata are drawn and later combined to get the collective results about a whole

population.

Example: Suppose a team of researchers has researched the demographics of students from within a High School in London. They found the percentage of different subjects as follow: 16% major in accounting, 38% major in English, 14% major in science and 32% major in mathematics. The division of the population has shown four strata representing different groups.

The research team then found the proportion of each strata. It has been observed that the proportion of each stratum is not same. The research team then re-sampled the 1,000 students, finding out 160 students with accounting, 380 students in English, 140 students in science and 320 students in mathematics groups. The division of the population in different groups has made a better representation of students.

Cluster sampling

A cluster sampling technique is used in order to **take samples for a large number of the population divided in different clusters**. Researchers have to search and select a number of clusters to be included in a sample from the whole population. This technique is helpful in marketing researches. Different clusters are selected for the sampling of different groups in a population. A **“one stage” cluster design** refers to taking samples from different groups.

Example: The most common cluster used in research is a geographical cluster. For example, a researcher wants to survey academic performance of high school students in Spain.

- The whole population of the country can be divided in different clusters i.e. cities of Spain.
- Now, considering the requirements of his research, he can select the clusters (cities here) through a random or systematic sampling technique.
- With the help of the systematic or random sampling method, the researcher can count all high school students, or a number of subjects can be selected from each cluster (city) in order to acquire the desired results.
- The best thing about cluster sampling is that offers equal chances of selection for all clusters in a population.

Multi-stage sampling

Multi-stage sampling is a **type of cluster sampling** in which the population is first divided into different groups for selection samples. In this kind of sampling, all clusters are not taken to choose samples. Samples from randomly taken clusters are chosen for analysis of data to generate results from the whole population.

Systematic sampling

Systematic sampling is a type of probability sampling method where a large population is filtered for a selection of sample members. **The members are selected from a random starting point at a fixed periodic interval**. The fixed periodic interval is known as a sampling interval. One sample interval can be calculated by dividing the total population from the number of clusters or groups.

Advantages and Disadvantages

Each probability sampling method has its own unique advantages and disadvantages.

Advantages

- **Cluster sampling:** convenience and ease of use.
- **Simple random sampling:** creates samples that are highly representative of the population.
- **Stratified random sampling:** creates strata or layers that are highly representative of strata or layers in the population.
- **Systematic sampling:** creates samples that are highly representative of the population, without the need for a [random number generator](#).

Disadvantages

- **Cluster sampling:** might not work well if unit members are not **homogeneous** (i.e. if they are different from each other).
- **Simple random sampling:** tedious and time-consuming, especially when creating larger samples.
- **Stratified random sampling:** tedious and time-consuming, especially when creating larger samples.
- **Systematic sampling:** not as random as simple random sampling,

Valid Surveys

In order to achieve **reliability** and **validity** in a survey, careful measures and proper consideration is required. If a researcher wants to conduct a research through surveys, high-quality data is required to generate the required results.

The high-quality data requirements include respondent effort requested, data collection method, order, structure and format of data collection questionnaires and forms and accuracy of elicited information, along with many other things.

Common Mistakes in Survey Sampling

Population specification error occurs when the researcher does not understand who she should survey.

Sample frame error happens when the selection is done in a wrong or inappropriate sub-population.

Selection error happens when respondents choose their own selection during a research study. It means respondents who are not willing to be a party of study or survey try to avoid it. It gives rise to selection error and can be mitigated by taking follow-ups and chasing respondents for the desired response.

Non-response error happens in case the non-respondents are other persons than respondents. The reason for the occurrence of this error is due to the unwillingness of potential respondents to be part of the study, or potential respondents are not contacted.

Sampling errors: the variation in the number of samples is the reason for sampling error. It can be mitigated by:

- Careful sample designs

- Large samples
- Multiple contacts to assure a representative response

Bias

A biased sampling method represents some kind of **avored or discriminated outcomes**. Sampling bias is a systematic bias, or known as ascertainment bias.

Example: Due to the emerging requirements of marketing in the corporate sector, telephone sampling is very common these days. A simple random sample can be selected from a long sampling list or frame consisting of telephone numbers of various prospective customers in a city, district or specific area. It is considered that all members of a specific city or community have an equal chance of selection in telephone sampling which is not correct here.

Those who don't have phones are excluded here. It also misses members or customers who don't have mobile phones. It also excludes those who are not intended to be a part of research survey, or don't respond to telephone calls or respond through answering machines. This way, there are several respondents missing from a survey which is assumed to be taken or conducted by providing an equal chance or selection to the whole customer population in an area.

Non-response bias

Sometimes, in survey sampling, individuals chosen for the sample are **unwilling or unable to participate** in the survey. A non-response bias is a bias that results when respondents differ in meaningful ways from non-respondents. Non-response is often the problem with mail surveys, where the response rate can be very low.

Response bias

The tendency or inclination of a person towards **answering survey questions in a misleading or untruthful way** is known as response bias. Suppose a person is very influenced by his own perception or attitude from his past experience or stereotypical behavior of society. It creates a problem in the effective research purpose as the researcher may not be aware that the respondent did not answer in an unbiased way.

References

[Cluster Sampling](#). via explorable.com

[Biased Sampling and Extrapolation](#). via ma.utexas.edu

[Validity and Reliability in Surveys](#). via relevantinsights.com

[Statistics and Probability Dictionary](#). via stattrek.com

[Response Bias: Definition and Examples](#). via statisticshowto.com

Legal Note: Unless otherwise stated, all rights reserved by Lecturio GmbH. For further legal regulations see our [legal information page](#).