



sampling

1. The act, process, or technique of selecting an appropriate sample.

A small portion, piece, or segment selected as a sample.

1. (Statistics) the process of selecting a random sample

2. (Pop Music) the process of taking a short extract from (a record) and mixing it into a different backing track

3. (Telecommunications) a process in which a continuous electrical signal is approximately represented by a series of discrete values, usually regularly spaced

4. the act or process of selecting a sample for testing.

5. the sample so selected. [1630–40]

Noun 1. sampling - (statistics) the selection of a suitable sample for study

selection, choice, option, pick - the act of choosing or selecting; "your choice of colors was unfortunate"; "you can take your pick"

random sampling - the selection of a random sample; each element of the population has an equal chance of been selected

proportional sampling, representative sampling, stratified sampling - the population is divided into subpopulations (strata) and random samples are taken of each stratum

statistics - a branch of applied mathematics concerned with the collection and interpretation of quantitative data and the use of probability theory to estimate population parameters

2. sampling - items selected at random from a population and used to test hypotheses about the population

sample distribution, sample

acceptance sampling - a statistical procedure for accepting or rejecting a batch of merchandise or documents; involves determining the maximum number of defects discovered in a sample before the entire batch is rejected

distribution, statistical distribution - (statistics) an arrangement of values of a variable showing their observed or theoretical frequency of occurrence

statistics - a branch of applied mathematics concerned with the collection and



interpretation of quantitative data and the use of probability theory to estimate population parameters

random sample - a sample in which every element in the population has an equal chance of being selected

proportional sample, representative sample, stratified sample - the population is divided into strata and a random sample is taken from each stratum

3. sampling - measurement at regular intervals of the amplitude of a varying waveform (in order to convert it to digital form)

measurement, measuring, mensuration, measure - the act or process of assigning numbers to phenomena according to a rule; "the measurements were carefully done"; "his mental measurings proved remarkably accurate"

telecom, telecommunication - (often plural) systems used in transmitting messages over a distance electronically

Sampling

Sampling is the process of selecting units (e.g., people, organizations) from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen. Let's begin by covering some of the [key terms in sampling](#) like "population" and "sampling frame." Then, because some types of sampling rely upon quantitative models, we'll talk about some of the [statistical terms used in sampling](#). Finally, we'll discuss the major distinction between [probability](#) and [Nonprobability](#) sampling methods and work through the major types in each.

In applications:

[Probability Sampling](#): Simple Random Sampling, Stratified Random Sampling, Multi-Stage Sampling

- What is each and how is it done?
- How do we decide which to use?
- How do we analyze the results differently depending on the type of sampling?

[Non-probability Sampling](#): Why don't we use non-probability sampling schemes? Two reasons:

- We can't use the mathematics of probability to analyze the results.



- In general, we can't count on a non-probability sampling scheme to produce representative samples.

In mathematical statistics books (for courses that assume you have already taken a probability course):

- Described as [assumptions about random variables](#)
- [Sampling with replacement versus sampling without replacement](#)

What are the main types of sampling and how is each done?

Simple Random Sampling: A simple random sample (SRS) of size n is produced by a scheme which ensures that each subgroup of the population of size n has an equal probability of being chosen as the sample.

Stratified Random Sampling: Divide the population into "strata". There can be any number of these. Then choose a simple random sample from each stratum. Combine those into the overall sample. That is a stratified random sample. (Example: Church A has 600 women and 400 men as members. One way to get a stratified random sample of size 30 is to take a SRS of 18 women from the 600 women and another SRS of 12 men from the 400 men.)

Multi-Stage Sampling: Sometimes the population is too large and scattered for it to be practical to make a list of the entire population from which to draw a SRS. For instance, when the a polling organization samples US voters, they do not do a SRS. Since voter lists are compiled by counties, they might first do a sample of the counties and then sample within the selected counties. This illustrates two stages. In some instances, they might use even more stages. At each stage, they might do a stratified random sample on sex, race, income level, or any other useful variable on which they could get information before sampling.

How does one decide which type of sampling to use?

The formulas in almost all statistics books assume simple random sampling. Unless you are willing to learn the more complex techniques to analyze the data after it is collected, it is appropriate to use simple random sampling. To learn the appropriate formulas for the more complex sampling schemes, look for a book or course on sampling.

Stratified random sampling gives more precise information than simple random sampling for a given sample size. So, if information on all members of the population is available that divides them into strata that seem relevant, stratified sampling will usually be used.



If the population is large and enough resources are available, usually one will use multi-stage sampling. In such situations, usually stratified sampling will be done at some stages.

How do we analyze the results differently depending on the different type of sampling?

The main difference is in the computation of the estimates of the variance (or standard deviation). An excellent book for self-study is *A Sampler on Sampling*, by Williams, Wiley. In this, you see a rather small population and then a complete derivation and description of the sampling distribution of the sample mean for a particular small sample size. I believe that is accessible for any student who has had an upper-division mathematical statistics course and for some strong students who have had a freshman introductory statistics course. A very simple statement of the conclusion is that the variance of the estimator is smaller if it came from a stratified random sample than from simple random sample of the same size. Since small variance means more precise information from the sample, we see that this is consistent with stratified random sampling giving better estimators for a given sample size.

Non-probability sampling schemes

These include voluntary response sampling, judgement sampling, convenience sampling, and maybe others.

In the early part of the 20th century, many important samples were done that weren't based on probability sampling schemes. They led to some memorable mistakes. Look in an introductory statistics text at the discussion of sampling for some interesting examples. The introductory statistics books I usually teach from are *Basic Practice of Statistics* by David Moore, Freeman, and *Introduction to the Practice of Statistics* by Moore and McCabe, also from Freeman. A particularly good book for a discussion of the problems of non-probability sampling is *Statistics* by Freedman, Pisani, and Purves. The detail is fascinating. Or, ask a statistics teacher to lunch and have them tell you the stories they tell in class. Most of us like to talk about these! Someday when I have time, maybe I'll write some of them here.

Mathematically, the important thing to recognize is that the discipline of statistics is based on the mathematics of probability. That's about random variables. All of our formulas in statistics are based on probabilities in sampling distributions of estimators. To create a sampling distribution of an estimator for a sample size of 30, we must be able to consider all possible samples of size 30 and base our analysis on how likely each individual result is.

In mathematical statistics books (for courses that assume you have already taken a probability course) the part of the problem relating to the sampling is described as assumptions about random variables.



Mathematical statistics texts almost always says to consider the X's (or Y's) to be independent with a common distribution. How does this correspond to some description of how to sample from a population? **Answer:** simple random sampling [with replacement](#).

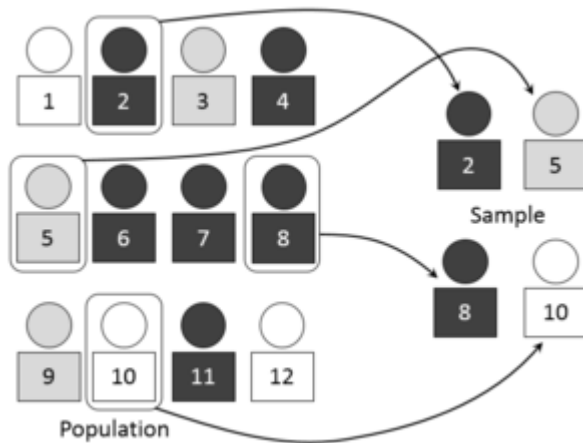
Sampling methods

Within any of the types of frame identified above, a variety of sampling methods can be employed, individually or in combination. Factors commonly influencing the choice between these designs include:

- Nature and quality of the frame
- Availability of auxiliary information about units on the frame
- Accuracy requirements, and the need to measure accuracy
- Whether detailed analysis of the sample is expected
- Cost/operational concerns

Simple random sampling

Main article: [Simple random sampling](#)



A visual representation of selecting a simple random sample

In a simple random sample (SRS) of a given size, all such subsets of the frame are given an equal probability. Furthermore, any given *pair* of elements has the same chance of selection as any other such pair (and similarly for triples, and so on). This minimises bias and simplifies analysis of results. In particular, the variance between individual results within the sample is a good



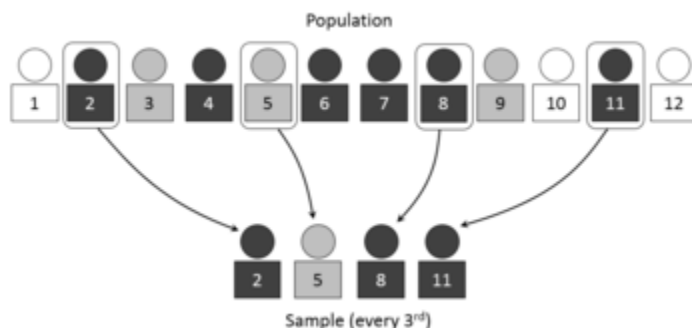
indicator of variance in the overall population, which makes it relatively easy to estimate the accuracy of results.

VS can be vulnerable to sampling error because the randomness of the selection may result in a sample that doesn't reflect the makeup of the population. For instance, a simple random sample of ten people from a given country will *on average* produce five men and five women, but any given trial is likely to overrepresent one sex and underrepresent the other. Systematic and stratified techniques attempt to overcome this problem by "using information about the population" to choose a more "representative" sample.

SRS may also be cumbersome and tedious when sampling from an unusually large target population. In some cases, investigators are interested in "research questions specific" to subgroups of the population. For example, researchers might be interested in examining whether cognitive ability as a predictor of job performance is equally applicable across racial groups. SRS cannot accommodate the needs of researchers in this situation because it does not provide subsamples of the population. "Stratified sampling" addresses this weakness of SRS.

Systematic sampling

Main article: [Systematic sampling](#)



A visual representation of selecting a random sample using the systematic sampling technique

Systematic sampling (also known as interval sampling) relies on arranging the study population according to some ordering scheme and then selecting elements at regular intervals through that ordered list. Systematic sampling involves a random start and then proceeds with the selection of every k th element from then onwards. In this case, $k = (\text{population size} / \text{sample size})$. It is important that the starting point is not automatically the first in the list, but is instead randomly chosen from within the first to the k th element in the list. A simple example would be to select every 10th name from the telephone directory (an 'every 10th' sample, also referred to as 'sampling with a skip of 10').



As long as the starting point is [randomized](#), systematic sampling is a type of [probability sampling](#). It is easy to implement and the [stratification](#) induced can make it efficient, *if* the variable by which the list is ordered is correlated with the variable of interest. 'Every 10th' sampling is especially useful for efficient sampling from [databases](#).

For example, suppose we wish to sample people from a long street that starts in a poor area (house No. 1) and ends in an expensive district (house No. 1000). A simple random selection of addresses from this street could easily end up with too many from the high end and too few from the low end (or vice versa), leading to an unrepresentative sample. Selecting (e.g.) every 10th street number along the street ensures that the sample is spread evenly along the length of the street, representing all of these districts. (Note that if we always start at house #1 and end at #991, the sample is slightly biased towards the low end; by randomly selecting the start between #1 and #10, this bias is eliminated.)

However, systematic sampling is especially vulnerable to periodicities in the list. If periodicity is present and the period is a multiple or factor of the interval used, the sample is especially likely to be *un*representative of the overall population, making the scheme less accurate than simple random sampling.

For example, consider a street where the odd-numbered houses are all on the north (expensive) side of the road, and the even-numbered houses are all on the south (cheap) side. Under the sampling scheme given above, it is impossible to get a representative sample; either the houses sampled will *all* be from the odd-numbered, expensive side, or they will *all* be from the even-numbered, cheap side, unless the researcher has previous knowledge of this bias and avoids it by using a skip which ensures jumping between the two sides (any odd-numbered skip).

Another drawback of systematic sampling is that even in scenarios where it is more accurate than SRS, its theoretical properties make it difficult to *quantify* that accuracy. (In the two examples of systematic sampling that are given above, much of the potential sampling error is due to variation between neighbouring houses – but because this method never selects two neighbouring houses, the sample will not give us any information on that variation.)

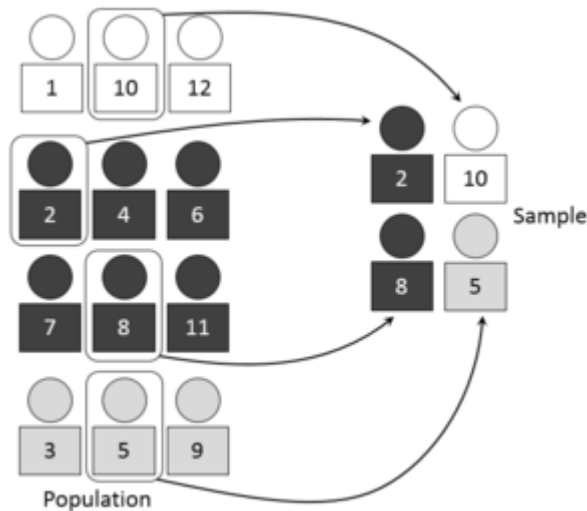
As described above, systematic sampling is an EPS method, because all elements have the same probability of selection (in the example given, one in ten). It is *not* 'simple random sampling' because different subsets of the same size have different selection probabilities – e.g. the set {4,14,24,...,994} has a one-in-ten probability of selection, but the set {4,13,24,34,...} has zero probability of selection.

Systematic sampling can also be adapted to a non-EPS approach; for an example, see discussion of PPS samples below.



Stratified sampling

Main article: [Stratified sampling](#)



A visual representation of selecting a random sample using the stratified sampling technique



It has been suggested that portions of this section be [moved](#) into [Stratified sampling](#).
([Discuss](#))

Where the population embraces a number of distinct categories, the frame can be organized by these categories into separate "strata." Each stratum is then sampled as an independent sub-population, out of which individual elements can be randomly selected.^[1] There are several potential benefits to stratified sampling.

First, dividing the population into distinct, independent strata can enable researchers to draw inferences about specific subgroups that may be lost in a more generalized random sample.

Second, utilizing a stratified sampling method can lead to more efficient statistical estimates (provided that strata are selected based upon relevance to the criterion in question, instead of availability of the samples). Even if a stratified sampling approach does not lead to increased statistical efficiency, such a tactic will not result in less efficiency than would simple random sampling, provided that each stratum is proportional to the group's size in the population.

Third, it is sometimes the case that data are more readily available for individual, pre-existing strata within a population than for the overall population; in such cases, using a stratified



sampling approach may be more convenient than aggregating data across groups (though this may potentially be at odds with the previously noted importance of utilizing criterion-relevant strata).

Finally, since each stratum is treated as an independent population, different sampling approaches can be applied to different strata, potentially enabling researchers to use the approach best suited (or most cost-effective) for each identified subgroup within the population.

There are, however, some potential drawbacks to using stratified sampling. First, identifying strata and implementing such an approach can increase the cost and complexity of sample selection, as well as leading to increased complexity of population estimates. Second, when examining multiple criteria, stratifying variables may be related to some, but not to others, further complicating the design, and potentially reducing the utility of the strata. Finally, in some cases (such as designs with a large number of strata, or those with a specified minimum sample size per group), stratified sampling can potentially require a larger sample than would other methods (although in most cases, the required sample size would be no larger than would be required for simple random sampling).

A stratified sampling approach is most effective when three conditions are met

1. Variability within strata are minimized
2. Variability between strata are maximized
3. The variables upon which the population is stratified are strongly correlated with the desired dependent variable.

Advantages over other sampling methods

1. Focuses on important subpopulations and ignores irrelevant ones.
2. Allows use of different sampling techniques for different subpopulations.
3. Improves the accuracy/efficiency of estimation.
4. Permits greater balancing of statistical power of tests of differences between strata by sampling equal numbers from strata varying widely in size.

Disadvantages

1. Requires selection of relevant stratification variables which can be difficult.
2. Is not useful when there are no homogeneous subgroups.
3. Can be expensive to implement.

Poststratification



Stratification is sometimes introduced after the sampling phase in a process called "poststratification".^[1] This approach is typically implemented due to a lack of prior knowledge of an appropriate stratifying variable or when the experimenter lacks the necessary information to create a stratifying variable during the sampling phase. Although the method is susceptible to the pitfalls of post hoc approaches, it can provide several benefits in the right situation. Implementation usually follows a simple random sample. In addition to allowing for stratification on an ancillary variable, poststratification can be used to implement weighting, which can improve the precision of a sample's estimates.^[1]

Oversampling

Choice-based sampling is one of the stratified sampling strategies. In choice-based sampling,^[5] the data are stratified on the target and a sample is taken from each stratum so that the rare target class will be more represented in the sample. The model is then built on this [biased sample](#). The effects of the input variables on the target are often estimated with more precision with the choice-based sample even when a smaller overall sample size is taken, compared to a random sample. The results usually must be adjusted to correct for the oversampling.

Probability-proportional-to-size sampling

In some cases the sample designer has access to an "auxiliary variable" or "size measure", believed to be correlated to the variable of interest, for each element in the population. These data can be used to improve accuracy in sample design. One option is to use the auxiliary variable as a basis for stratification, as discussed above.

Another option is probability proportional to size ('PPS') sampling, in which the selection probability for each element is set to be proportional to its size measure, up to a maximum of 1. In a simple PPS design, these selection probabilities can then be used as the basis for [Poisson sampling](#). However, this has the drawback of variable sample size, and different portions of the population may still be over- or under-represented due to chance variation in selections.

Systematic sampling theory can be used to create a probability proportionate to size sample. This is done by treating each count within the size variable as a single sampling unit. Samples are then identified by selecting at even intervals among these counts within the size variable. This method is sometimes called PPS-sequential or monetary unit sampling in the case of audits or forensic sampling.

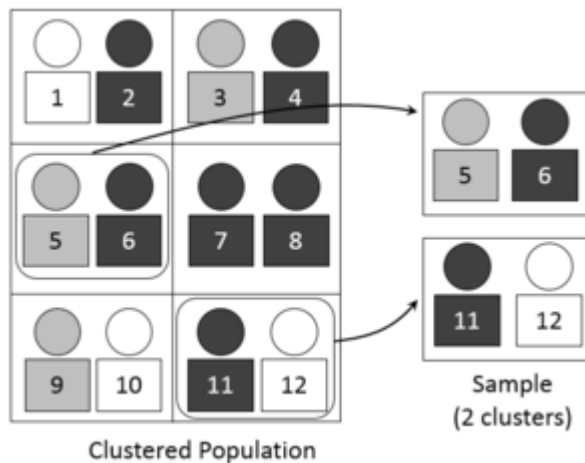
Example: Suppose we have six schools with populations of 150, 180, 200, 220, 260, and 490 students respectively (total 1500 students), and we want to use student population as the basis for a PPS sample of size three. To do this, we could allocate the first school numbers 1 to 150, the second school 151 to 330 (= 150 + 180), the third school 331 to 530, and so on to the last



school (1011 to 1500). We then generate a random start between 1 and 500 (equal to $1500/3$) and count through the school populations by multiples of 500. If our random start was 137, we would select the schools which have been allocated numbers 137, 637, and 1137, i.e. the first, fourth, and sixth schools.

The PPS approach can improve accuracy for a given sample size by concentrating sample on large elements that have the greatest impact on population estimates. PPS sampling is commonly used for surveys of businesses, where element size varies greatly and auxiliary information is often available—for instance, a survey attempting to measure the number of guest-nights spent in hotels might use each hotel's number of rooms as an auxiliary variable. In some cases, an older measurement of the variable of interest can be used as an auxiliary variable when attempting to produce more current estimates.^[6]

Cluster sampling



A visual representation of selecting a random sample using the cluster sampling technique

Sometimes it is more cost-effective to select respondents in groups ('clusters'). Sampling is often clustered by geography, or by time periods. (Nearly all samples are in some sense 'clustered' in time – although this is rarely taken into account in the analysis.) For instance, if surveying households within a city, we might choose to select 100 city blocks and then interview every household within the selected blocks.

Clustering can reduce travel and administrative costs. In the example above, an interviewer can make a single trip to visit several households in one block, rather than having to drive to a different block for each household.



It also means that one does not need a [sampling frame](#) listing all elements in the target population. Instead, clusters can be chosen from a cluster-level frame, with an element-level frame created only for the selected clusters. In the example above, the sample only requires a block-level city map for initial selections, and then a household-level map of the 100 selected blocks, rather than a household-level map of the whole city.

Cluster sampling (also known as clustered sampling) generally increases the variability of sample estimates above that of simple random sampling, depending on how the clusters differ between themselves, as compared with the within-cluster variation. For this reason, cluster sampling requires a larger sample than SRS to achieve the same level of accuracy – but cost savings from clustering might still make this a cheaper option.

[Cluster sampling](#) is commonly implemented as [multistage sampling](#). This is a complex form of cluster sampling in which two or more levels of units are embedded one in the other. The first stage consists of constructing the clusters that will be used to sample from. In the second stage, a sample of primary units is randomly selected from each cluster (rather than using all units contained in all selected clusters). In following stages, in each of those selected clusters, additional samples of units are selected, and so on. All ultimate units (individuals, for instance) selected at the last step of this procedure are then surveyed. This technique, thus, is essentially the process of taking random subsamples of preceding random samples.

Multistage sampling can substantially reduce sampling costs, where the complete population list would need to be constructed (before other sampling methods could be applied). By eliminating the work involved in describing clusters that are not selected, multistage sampling can reduce the large costs associated with traditional cluster sampling.^[6] However, each sample may not be a full representative of the whole population.

Quota sampling

In **quota sampling**, the population is first segmented into [mutually exclusive](#) sub-groups, just as in [stratified sampling](#). Then judgement is used to select the subjects or units from each segment based on a specified proportion. For example, an interviewer may be told to sample 200 females and 300 males between the age of 45 and 60.

It is this second step which makes the technique one of non-probability sampling. In quota sampling the selection of the sample is non-[random](#). For example interviewers might be tempted to interview those who look most helpful. The problem is that these samples may be biased because not everyone gets a chance of selection. This random element is its greatest weakness and quota versus probability has been a matter of controversy for several years.



Minimax sampling

In imbalanced datasets, where the sampling ratio does not follow the population statistics, one can resample the dataset in a conservative manner called [minimax sampling](#).^[7] The minimax sampling has its origin in [Anderson](#) minimax ratio whose value is proved to be 0.5: in a binary classification, the class-sample sizes should be chosen equally.^[8] This ratio can be proved to be minimax ratio only under the assumption of [LDA](#) classifier with Gaussian distributions.^[8] The notion of minimax sampling is recently developed for a general class of classification rules, called class-wise smart classifiers. In this case, the sampling ratio of classes is selected so that the worst case classifier error over all the possible population statistics for class prior probabilities, would be the best.^[7]

Accidental sampling

[Accidental sampling](#) (sometimes known as **grab, convenience or opportunity sampling**) is a type of nonprobability sampling which involves the sample being drawn from that part of the population which is close to hand. That is, a population is selected because it is readily available and convenient. It may be through meeting the person or including a person in the sample when one meets them or chosen by finding them through technological means such as the internet or through phone. The researcher using such a sample cannot scientifically make generalizations about the total population from this sample because it would not be representative enough. For example, if the interviewer were to conduct such a survey at a shopping center early in the morning on a given day, the people that he/she could interview would be limited to those given there at that given time, which would not represent the views of other members of society in such an area, if the survey were to be conducted at different times of day and several times per week. This type of sampling is most useful for pilot testing. Several important considerations for researchers using convenience samples include:

1. Are there controls within the research design or experiment which can serve to lessen the impact of a non-random convenience sample, thereby ensuring the results will be more representative of the population?
2. Is there good reason to believe that a particular convenience sample would or should respond or behave differently than a random sample from the same population?
3. Is the question being asked by the research one that can adequately be answered using a convenience sample?

In social science research, [snowball sampling](#) is a similar technique, where existing study subjects are used to recruit more subjects into the sample. Some variants of snowball sampling, such as respondent driven sampling, allow calculation of selection probabilities and are probability sampling methods under certain conditions.



Line-intercept sampling

Line-intercept sampling is a method of sampling elements in a region whereby an element is sampled if a chosen line segment, called a "transect", intersects the element.

Panel sampling

Panel sampling is the method of first selecting a group of participants through a random sampling method and then asking that group for (potentially the same) information several times over a period of time. Therefore, each participant is interviewed at two or more time points; each period of data collection is called a "wave". The method was developed by sociologist [Paul Lazarsfeld](#) in 1938 as a means of studying [political campaigns](#).^[9] This [longitudinal](#) sampling-method allows estimates of changes in the population, for example with regard to chronic illness to job stress to weekly food expenditures. Panel sampling can also be used to inform researchers about within-person health changes due to age or to help explain changes in continuous dependent variables such as spousal interaction.^[10] There have been several proposed methods of analyzing [panel data](#), including [MANOVA](#), [growth curves](#), and [structural equation modeling](#) with lagged effects.

Snowball sampling

Snowball sampling involves finding a small group of initial respondents and using them to recruit more respondents. It is particularly useful in cases where the population is hidden or difficult to enumerate.

Theoretical sampling

Theoretical sampling^[11] occurs when samples are selected on the basis of the results of the data collected so far with a goal of developing a deeper understanding of the area or develop theory.

Replacement of selected units

Sampling schemes may be *without replacement* ('WOR'—no element can be selected more than once in the same sample) or *with replacement* ('WR'—an element may appear multiple times in the one sample). For example, if we catch fish, measure them, and immediately return them to the water before continuing with the sample, this is a WR design, because we might end up catching and measuring the same fish more than once. However, if we do not return the fish to the water (e.g., if we eat the fish), this becomes a WOR design.



Sample size

Main article: [Sample size](#)

Formulas, tables, and power function charts are well known approaches to determine sample size.

Steps for using sample size tables

Postulate the effect size of interest, α , and β .

1. Check sample size table^[12]
 1. Select the table corresponding to the selected α
 2. Locate the row corresponding to the desired power
 3. Locate the column corresponding to the estimated effect size.
 4. The intersection of the column and row is the minimum sample size required.

Sampling and data collection

Good data collection involves:

- Following the defined sampling process
- Keeping the data in time order
- Noting comments and other contextual events
- Recording non-responses

Errors in sample surveys

Survey results are typically subject to some error. Total errors can be classified into sampling errors and non-sampling errors. The term "error" here includes systematic biases as well as random errors.

Sampling errors and biases

Sampling errors and biases are induced by the sample design. They include:

1. **Selection bias**: When the true selection probabilities differ from those assumed in calculating the results.
2. **Random sampling error**: Random variation in the results due to the elements in the sample being selected at random.



Non-sampling error

Non-sampling errors are other errors which can impact the final survey estimates, caused by problems in data collection, processing, or sample design. They include:

1. **Over-coverage:** Inclusion of data from outside of the population.
2. **Under-coverage:** Sampling frame does not include elements in the population.
3. **Measurement error:** e.g. when respondents misunderstand a question, or find it difficult to answer.
4. **Processing error:** Mistakes in data coding.
5. **Non-response:** Failure to obtain complete data from all selected individuals.

After sampling, a review should be held of the exact process followed in sampling, rather than that intended, in order to study any effects that any divergences might have on subsequent analysis. A particular problem is that of *non-response*.

Two major types of non-response exist: unit nonresponse (referring to lack of completion of any part of the survey) and item non-response (submission or participation in survey but failing to complete one or more components/questions of the survey).^{[13][14]} In [survey sampling](#), many of the individuals identified as part of the sample may be unwilling to participate, not have the time to participate (opportunity cost),^[15] or survey administrators may not have been able to contact them. In this case, there is a risk of differences, between respondents and nonrespondents, leading to biased estimates of population parameters. This is often addressed by improving survey design, offering incentives, and conducting follow-up studies which make a repeated attempt to contact the unresponsive and to characterize their similarities and differences with the rest of the frame.^[16] The effects can also be mitigated by weighting the data when population benchmarks are available or by imputing data based on answers to other questions.

Survey weights

In many situations the sample fraction may be varied by stratum and data will have to be weighted to correctly represent the population. Thus for example, a simple random sample of individuals in the United Kingdom might include some in remote Scottish islands who would be inordinately expensive to sample. A cheaper method would be to use a stratified sample with urban and rural strata. The rural sample could be under-represented in the sample, but weighted up appropriately in the analysis to compensate.



Methods of producing random samples

Random number table

- Mathematical algorithms for [pseudo-random number generators](#)
- Physical randomization devices such as coins, playing cards or sophisticated devices such as [ERNIE](#)

History

Random sampling by using lots is an old idea, mentioned several times in the Bible. In 1786 Pierre Simon [Laplace](#) estimated the population of France by using a sample, along with [ratio estimator](#). He also computed probabilistic estimates of the error. These were not expressed as modern [confidence intervals](#) but as the sample size that would be needed to achieve a particular upper bound on the sampling error with probability 1000/1001. His estimates used [Bayes' theorem](#) with a uniform [prior probability](#) and assumed that his sample was random. [Alexander Ivanovich Chuprov](#) introduced sample surveys to [Imperial Russia](#) in the 1870s. [citation needed]

YARN SAMPLING:

1. Use of random numbers:

Table of random sampling number are normally used a small number of yarn bobbins are to be selected from comparatively small bulk size.

Total 10 package are to be selected at random from the consignment.

If consignment contains more than five cases, they are selected at random from it.& then two packages are selected at random from each case.

If no. of cases <5, then 10 packages are selected at random approximately, equal from each package.

FABRIC SAMPLING:



Fabric samples from warp and weft are taken separately.



Warp direction should be marked before it is cut out.



No two specimens should contain same warp or weft threads.



Samples should not be from within 50mm of selvedge.

