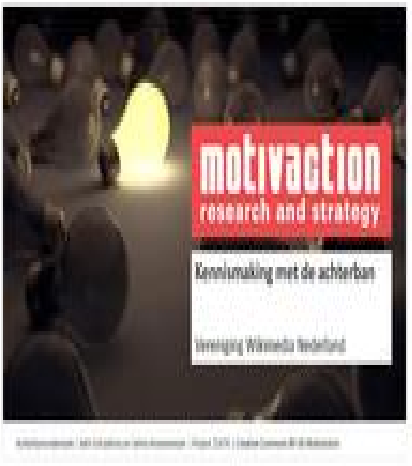
Sample (statistics)

From Wikipedia, the free encyclopedia

In statistics and quantitative research methodology, a **data sample** is a set of data collected and/or selected from a statistical population by a defined procedure.

Contents



Acting on instructions from the Vereniging Wikimedia Nederland (nl), Motivaction International B.V. conducted a survey of the Wikimedia Nederland supporters and members and the users of the various Wikimedia initiatives including, in particular, the users of Wikipedia.

Kinds of samples

In other words, the population is very large, making a census or a complete enumeration of all the values in the population impractical or impossible. The sample represents a subset of manageable size. Samples are collected and statistics are calculated from the samples so that one can make inferences or extrapolations from the sample to the population. This process of collecting information from a sample is referred to as sampling. The data sample may be drawn from a population without replacement, in which case it is a subset of a population; or with replacement, in which case it is a multisubset.

A **complete sample** is a set of objects from a parent population that includes ALL such objects that satisfy a set of well-defined selection criteria. For example, a complete sample of Australian men taller than 2m would consist of a list of **every** Australian male taller than 2m. But it wouldn't include German males, or tall Australian females, or people shorter than 2m. So to compile such a complete sample requires a complete list of the parent population, including data on height, gender, and nationality for each member of that parent population. In the case of human populations, such a complete list is unlikely to exist, but such complete samples are often available in other disciplines, such as complete magnitude-limited samples of astronomical objects.

An **unbiased (representative) sample** is a set of objects chosen from a complete sample using a selection process that does not depend on the properties of the objects. For example, an unbiased sample of Australian men taller than 2m might consist of a randomly sampled subset of 1% of Australian males taller than 2m. But one chosen from the electoral register might not be unbiased since, for example, males aged under 18 will not be on the electoral register. In an astronomical context, an unbiased sample might consist of that fraction of a complete sample for which data are available, provided the data availability is not biased by individual source properties.

The best way to avoid a biased or unrepresentative sample is to select a random sample, also known as a probability sample. A random sample is defined as a sample where each individual member of the population has a known, non-zero chance of being selected as part of the sample. Several types of random samples are simple random samples, systematic samples, stratified random samples, and cluster random samples.

A sample that is not random is called a non-random sample or a non-probability sampling. Some examples of nonrandom samples are convenience samples, judgment samples, purposive samples, quota samples, snowball samples, and quadrature nodes in quasi-Monte Carlo methods.

Statistic samples have multiple uses. They can be used in many situations.

Mathematical description of random sample

In mathematical terms, given a random variable *X* with distribution *F*, a random sample of length *n* (where *n* may be any of 1,2,3,...) is a set of *n* independent, identically distributed (iid) random variables with distribution *F*.[1]

A sample concretely represents *n* experiments in which the same quantity is measured. For example, if *X* represents the height of an individual and *n* individuals are measured, X_i will be the height of the *i*-th individual. Note that a sample of random variables (i.e. a set of measurable functions) must not be confused with the realizations of these variables (which are the values that these random variables take, formally called random variates). In other words, X_i is a function representing the measurement at the i-th experiment and $x_i = X_i(\omega)$ is the value actually obtained when making the measurement.

The concept of a sample thus includes the *process* of how the data are obtained (that is, the random variables). This is necessary so that mathematical statements can be made about the sample and statistics computed from it, such as the sample mean and covariance.

See also

- Estimation theory
- Replication (statistics)
- Sample size determination
- Sampling (statistics)
- Survey sampling

Notes

1. **^** Samuel S. Wilks, *Mathematical Statistics*, John Wiley, 1962, Section 8.1

External links

- Statistical Terms Made Simple (http://www.socialresearchmethods.net/kb/sampstat.php)

Retrieved from "http://en.wikipedia.org/w/index.php?title=Sample_(statistics)&oldid=599545484"
Categories: Sampling (statistics)