PARAMETER AND STATISTIC

In order to avoid verbal confusion with the statistical constants of the population, viz., mean, variance, etc., of the population which are usually referred to as parameters, statistical measures computed from the sample observations alone, e.g., mean, variance, etc., of the sample have been termed as statistic.

In practice parameter values are not known and their estimates based on the sample values are generally used. Thus statistic which may be regarded as an estimate of the parameter, obtained from the sample, is a function of the sample values only. It may be pointed out that a statistic, as it is based on sample values and as there are multiple choices of the samples that can be drawn from a population, varies from sample to sample. The determination or the characterization of the variation (in the values of the statistic obtains from different samples) that may be attributed to chance or fluctuations of sampling is one of the fundamental problems of the sampling theory.

(Unbiased Estimate)

A statistic t = t (x_1 , x_2 , ..., x_n), a function of the sample values x_1 , x_2................,x_n is an unbiased estimate of population parameterθ if E(t) = θ, i.e., if E(Statistics) = Parameter, then statistic is said to be an unbiased estimate of the parameter.

S**ampling Distribution**. The number of possible samples of size n that can be drawn from a finite population of size N is NCn. (If N is large or infinite, then we can draw a large number of such samples.) For each of these samples we can compute a statistic, say't' ....e.g., mean, variance, etc., which will obviously vary from sample to sample. The aggregate of the various values of the statistic under consideration so obtained (one from each sample), may be grouped into a frequency distribution which is known as the sampling distribution of the statistic. Thus, we can have the sampling distribution of the sample mean $\bar{x}$, the sample variance, etc.

Standard Error.

The standard deviation of the sampling distribution of a statistic is known as its Standard Error. The standard errors (S.E.) of some of the well-known statistics are given in Table, where n is the sample size, σ^2 the population variance, P the population proportion and Q = 1- P.

Utility of Standard Error.

S.E. plays a very important role in the large sample theory and forms the basis of the testing of hypothesis. If t is any statistic, then for large samples

Z = $\frac{t-E(t)}{\sqrt{V(t)}} \sim N(0,1)$

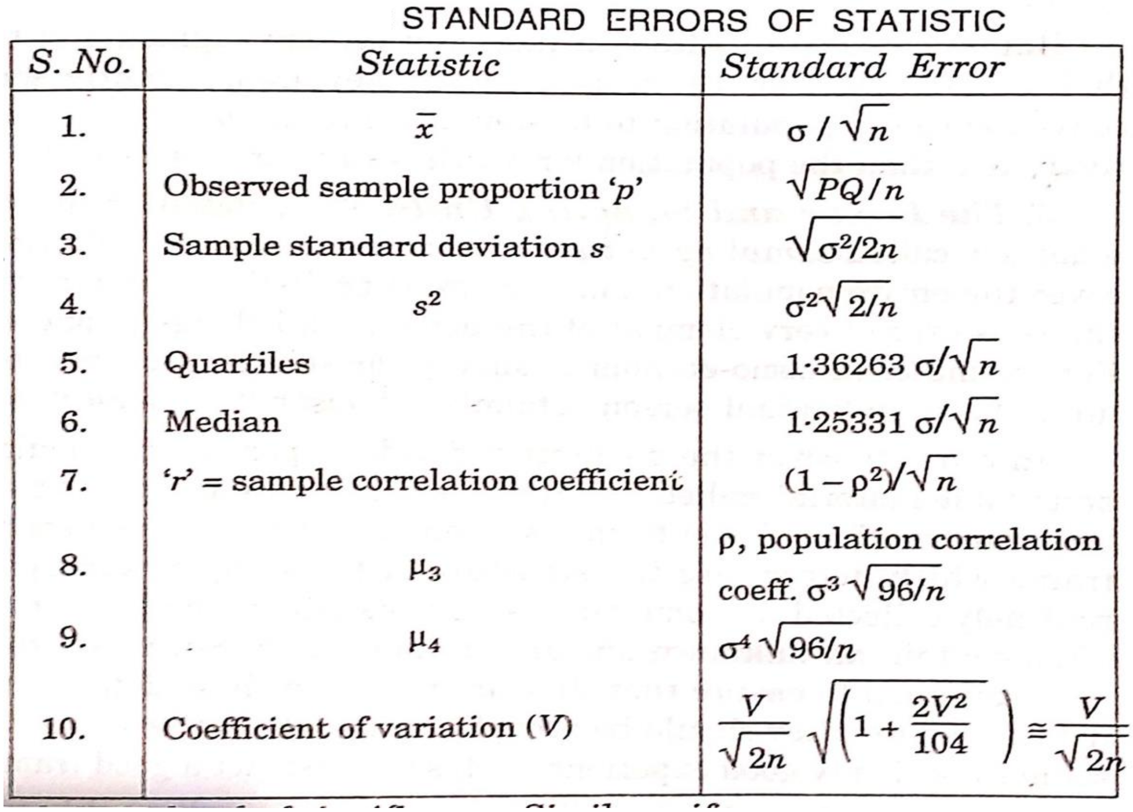
$$\Rightarrow \quad Z = \frac{t - E(t)}{S.E.(t)} \sim N(0,1)$$

: Thus, if the discrepancy between the observed and the expected (hypothetical) values of the statistic is greater then 1.96 times the S.E the hypothesis is, rejected at 5% level of significance. Similary,

if t - E(t) ≤ 1.96 x S.E. (t),

the deviation is not regarded significant at 5% level of significance. In other words the deviation' t - E(t), could have arisen due to fluctuations of sampling and the data do not provide us any evidence against the null hypothesis which may, therefore , be accepted at 5% level of Significance. Similarly we can discuss the significance of the difference at 1% level of significance

The .magnitude of the standard error gives an index of the precision of the estimate of the parameter. The reciprocal of the standard error is taken as the measure of reliability or precision of the sample.



STANDARD ERRORS OF STATISTIC

S. No.	Statistic	Standard Error
1.	$\bar{x}$	$\sigma / \sqrt{n}$
2.	Observed sample proportion 'p'	$\sqrt{PQ/n}$
3.	Sample standard deviation s	$\sqrt{\sigma^2/2n}$
4.	s^2	$\sigma^2\sqrt{2/n}$
5.	Quartiles	$1{\cdot}36263\ \sigma/\sqrt{n}$
6.	Median	$1{\cdot}25331\ \sigma/\sqrt{n}$
7.	'r' = sample correlation coefficient	$(1-\rho^2)/\sqrt{n}$
8.	μ_3	ρ, population correlation coeff. $\sigma^3\sqrt{96/n}$
9.	μ_4	$\sigma^4\sqrt{96/n}$
10.	Coefficient of variation (V)	$\frac{V}{\sqrt{2n}}\sqrt{\left(1+\frac{2V^2}{104}\right)} \cong \frac{V}{\sqrt{2n}}$

SAMPLING AND NON-SAMPLING ERRORS

The errors involved in the collection, processing and analysis of a data may be broadly classified under the following two heads:

(**i) Sampling Errors, and (ii) Non-sampling Errors**.

(i) Sampling Errors.

Sampling errors have their origin in sampling and arise due to the fact that only a part of the population (i.e., sample) has been used to estimate population parameters and draw inferences about the population. As such the sampling errors are absent in a complete enumeration survey. Sampling biases are primarily due to the following reasons:

1. Faulty selection of the sample. Some of the bias is introduced by the use of defective sampling technique for the selection of a sample, e.g., purposive or judgment sampling in which the investigator deliberately selects a representative sample to obtain certain results. This bias can select a representative sample to obtain certain results. This bias can be overcome by strictly adhering to a simple random sample or by selecting a sample at random subject to restrictions which while improving the accuracy are of such nature that they do not introduce bias in the results.

2. Substitution.

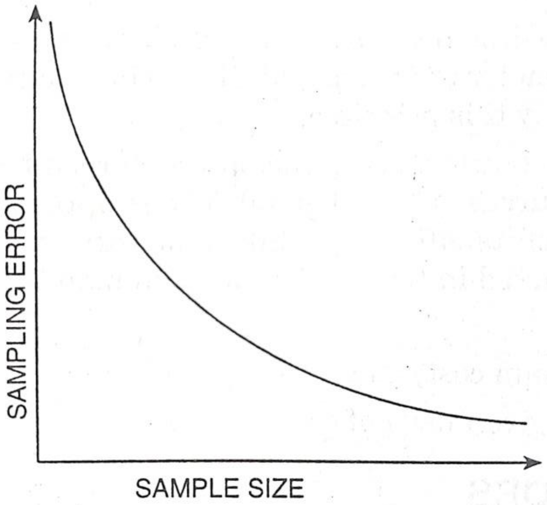
If difficulties arise in enumerating a particular sampling unit included in the random sample, the investigators usually substitute a convenient member of the population. This obviously leads to some bias since the characteristics processed by the substituted unit will usually be different from those possessed by the unit originally included in the sample.

3. Faulty demarcation of sampling units.

Bias due to defective demarcation of sampling units is particularly significant in area surveys such as agricultural experiments in the field or crop cutting survey, etc. In such surveys, while dealing with border line cases, it depends more or less on the discretion of the investigator whether to include them in the sample or not,

4. Constant error due to improper choice of the statistics for estimating the population Parameters.

For example, if x_1, x_2, ..., x_n is a sample of independent observations, then the sample variance $s^2 = \sum_{i=1}^{n} {(x_i - \overline{x})^2}/{n}$ as an estimate of the population variance σ^2 is biased whereas the statistic $\frac{1}{n}\sum_{i=1}^{n}(x - \bar{x})^2$, is an unbiased estimate of σ^2.



Remark: Increase in the sample size (i.e., the number of units in the sample) usually results in the decrease in sampling error. In fact, in many situations this decrease in sampling error is inversely proportional to the square root of the sample size as illustrated in Figure

. **(ii) Non-sampling Errors**. As distinct from sampling errors which are due to the inductive process of inferring about the population on the basis of a sample, the non-sampling errors primarily arise at the stages of observation, ascertainment and processing of the data and are thus present in both the complete enumeration survey and the sample survey. Thus, *the data obtained in a complete census, although free from sampling errors, would still be subject to non-sampling errors whereas data obtained in a sample survey should be subject to both sampling and non-sampling errors.*

Non-sampling errors can occur at every stage of the planning or execution of census or sample survey. The preparation of an exhaustive list of all the sources of non-sampling errors is a very difficult task. However, a careful examination of the major phases of a survey (complete or sample) indicates that some of the more important non-sampling errors arise from the following factors :

1. ***Faulty Planning or Definitions***
   The planning of a survey consists in explicitly stating the objectives of the survey. These objectives are then translated into (i) a set of definitions of the characteristics for which data are to be collected, and (ii) into a set of specifications for collecting, processing and publishing. Here the non-sampling errors can be due to:

   (a) Data specification being inadequate and inconsistent with respect to the objectives of the survey.

   (b) Error due to location of the units and actual measurement of the characteristics, errors in recording the measurements, errors due to ill-designed questionnaire, etc.

(c) Lack of trained and qualified investigators and lack of adequate supervisory staff.

2. Response Errors. These errors are introduced as a result of the responses furnished by the respondents and may be due to any of the following reasons:

(i) Response errors may be accidental. For example, the respondent may misunderstand a particular question and accordingly furnish improper information un-intentionally.

(ii) Prestige bias**.** An appeal to the pride or prestige of person interviewed may introduce yet another kind of bias, called prestige bias by virtue of which he may upgrade his education, intelligence quotient, occupation, income, etc., or downgrade his age, thus resulting in wrong answers.

(iii) Self-interest

. Quite often, in order to safeguard one's self-interest, one may give incorrect information, e.g., a person may give an underestimate of his salary or production and an over-statement of his expenses or requirements, etc.

(iv) Bias due to interviewer.

Sometimes the interviewer may affect the accuracy of the response by the way he asks questions or records them. The information obtained on suggestions from the interviewer is very likely to be influenced by interviewer's beliefs and prejudices.

(v) Failure of respondent's memory.

One source of error which is common to most of the methods of collecting information is that of 'recall'. Many of the questions in surveys refer to happenings or conditions in the past and there is a problem both of remembering the event and associating it with the correct time period.

3. Non-response Biases.

Non-response biases occur if full information is not obtained on all the sampling units. In house-to-house survey, non-response usually results if the respondent is not found at home even after repeated calls, or if he/she is unable to furnish the information on all the questions or if he/she refuses to answer certain questions. Therefore, some bias is introduced as a consequence of the exclusion of a section of the population with certain peculiar characteristics, due to non-response.

4. Errors in Coverage.

If the objectives of the survey are not precisely stated in clear cut terms, this may result in

(i) the inclusion in the survey of certain units which are not to be included, or

(ii) The exclusion of certain units which were to be included in the survey under the objectives. For example, in a census to determine the number of individuals in the age group, say, 20 years to 50 years, more or less serious errors may occur in deciding whom to enumerate unless particular community or area is not specified and also the time at which the age is to be specified.

5. Compiling Errors.

Various operations of data processing such as editing and coding of the responses, tabulation and summarizing the original observations made in the survey are a potential source of error. Compilation errors are subject to control though verification, consistency check, etc.

6. Publication Errors.

Publication errors, i.e., the errors committed during presentation and printings of tabulated results are basically due to two sources. The first refers to the mechanics of publication—the proofing error and the like. The other, which is of more serious nature, lies in the failure of the survey organization to point out the limitations of the statistics.