1B SAMPLING THEORY

1. Population or Universe

Population in statistics means the whole of the information which comes under the purview of statistical investigation. It is the totality of all the observations of a statistical experiment or enquiry.

A population may be finite or infinite according as the number of observations or items in it are finite or infinite. The population of weights of students of class XII in a government school is an example of a finite population. The population of pressure at different points in the atmosphere is an example of an infinite population.

Types of Population:

a) Finite Population: When the items in the population are fixed and limited.
Example : No. of students in the class

b) Infinite Population: If a population consist of infinite no. of items its an infinite population. If a sample is known to have been drawn from a continuous probability distribution, then the population is infinite. Example : Population of all real numbers lying between 5 and 20.

c) Real Population: A Population consisting of the items which are all present physically is termed as real population.

d) Hypothetical Population: The Population consists of the results of the repeated trails is named as hypothetical population The tossing of a coin repeatedly results into a hypothetical population of heads and tails.

2. Sample

A part of the population selected for study is called a sample. In other words, the selection of a group of individuals or items from a population in such a way that this group represents the population, is called a sample.

1. Sampling is a process whereby we judge the characteristics or draw inference about the totality or Universe (known as population) on the basis of judging the characteristics of a selected portion taken from that totality (known as sample).
2. Sample: Sample is the part of population selected on some basis it is a finite subset

of the population.

3. Sample Units : Units forming the samples are called Sample Units.
4. Sample Frame : A complete list of sampling units is called Sample Frame
5. Sample Faction : $\frac{n}{N}$ is called Sampling Fraction where n = Sample Size and N = Population Size.
6. Complete enumeration or census : In case of enumeration, information is collected for each and every unit. The aggregate of all the units under consideration is called the 'population' or the 'universe'. The results are more accurate and reliable but it involves lot of time, money and man power

3. Parameter and Statistic

There are various statistical measures in statistics such as mean, median, mode, standard deviation, coefficient of variation etc. These statistical measures can be computed both from population (or universe) data and sample data.

Parameter : Any statistical measure computed from population data is known as parameter.

Statistics : Any statistical measure computed from sample data is known as statistic. Thus a parameter is a statistical measure which relates to the population and is based on population data, whereas a statistic is a statistical measure which relates to the sample and is based on sample data. Thus a population mean, population median, population variance, population coefficient of variation etc., are all parameters. Statistic computed from a Sample such as sample mean, sample variance etc.

	Notations	
Statistical Measure	**Population**	**Sample**
Mean	μ	x
Standard deviation	σ	s
Proportion	P	p
Size	N	n

Related MCQ's:

1. The aggregate or totality of statistical data forming a subject of investigation is known as :

a) Sample
b) Population
c) Both a) and b) above
d) None of the above

2. Population is also known as:

a) Universe b) Range
c) Area d) Region

3. A portion of the population which is examined with a view to estimating the characteristics of the population are known as:

a) Sample b) Universe
c) Population d) Statistic

4. If a sample is known to have been drawn from a continuous probability distribution then the population is .

a) Large b) Finite
c) Infinite d) Nothing can be said about the population

5. The population of tea drinkers in Kolkata City is an example of:

a) A hypothetical population
b) An infinite population
c) A finite population
d) Nothing can be said about the population

6. The possibility of reaching valid conclusions concerning a population by means of a population by means of a properly chosen sample is based on which of the following laws?

a) Law of Inertia b) Law of Large Number
c) Law of Statistical Regularity d) All of the above

7. A sample is a study of a of the population.

a) parameters b) statistics
c) part d) none of the above

8. A population is the of limits under study.

a) totality b) part
c) subset d) none of the above

9. "A sample is less expensive than a census"
   a) The statement is incorrect.
   b) The statement is correct.
   c) The given statement is based on nature of sample.
   d) None of the above.

10. When the population is infinite we should use the:
   a) Sample Method
   b) Census Method
   c) Either Sample or Census Method
   d) None of the above

11. A complete list of all the units in a finite population, properly numbered for identification, is called a:
   a) Universe
   b) Sampling Data
   c) Sampling Units
   d) Sampling Frame

12. Statistical data may be collected by complete enumeration called
   a) Sample Enquiry
   b) Census Enquiry
   c) Both a) and b) above
   d) Neither a) nor b) above

13. A border patrol checkpoint which stops every passenger van is utilizing :
   a) simple random sampling.
   b) systematic sampling
   c) systematic sampling.
   d) complete enumeration

14. A population consisting of all the items which are physically present is called :
   a) hypothetical
   b) normal population
   c) existent population
   d) none of the above

15. A population consisting of all real numbers is an example of :
   a) an infinite population
   b) a finite population
   c) an imaginary
   d) none of the above

16. The population of roses in Salt Lake City is an example of
   a) a hypothetical population
   b) an infinite population
   c) a finite population
   d) an imaginary population

17. Value of _____ is different for different sample

a) Statistic b) Population
c) Both a) and b) above d) None of the above.

18. A statistic is a variable.

a) Compound b) Simple
c) Random d) Both a) and c) above

4. Basic principle of Sample Survey

a) **Law of Statistical Regularity :** It states that a reasonably larger number of items selected at random from a large group of items, will on the average, represent the characteristics of the group.

b) **Law of Inertia of Large Numbers :** This law states that other things same, as the sample size increases, the results tend to be more reliable and accurate.

c) **Principle of Optimization :** The principle of optimization ensures that an optimum level of efficiency at a minimum cost or the maximum efficiency at the given level of cost can be achieved with the selection of an appropriate sampling design.

d) **Principle of Validity :** The principle of validity states that a sampling design is valid only if it is possible to obtain valid estimates and valid tests about population parameters. Only a probability sampling ensures this validity.

Related MCQ's:

19. Law of Statistical Regularity states that:

a) A sample of reasonably small size when selected at random, is almost not sure to represent the characteristics of the population
b) A sample of reasonably large size when selected, is almost not sure to represent the characteristics of the population.
c) A sample of reasonably large size when selected at random, is almost sure to represent the characteristics of the population, on an average
d) None of the above

20. Law of Inertia is also known as:

a) Law of Statistics. b) Law of Large Number
c) Law of Balance d) None of the above.

21. Law of Inertia states that:

a) Sample of high size show a high degree of stability.

b) Sample of low size shows a high degree of stability.

c) Results obtained from sample of high size are expected to be very far.

d) None of the above.

22. Increase in reliability and accuracy of results from a sampling study with the increase in sample size is known as the principle of:

a) statistical regularity b) optimization.

c) law of increasing returns. d) inertia of large numbers.

23. Sampling error increases with an increase in the size of the sample.

a) The above statement is true.

b) The above statement is not true.

c) Sampling error do not depends upon the sample size

d) None of the above

24. Two basic Statistical laws concerning a population are

a) the law of statistical irregularity and the law of inertia of large numbers.

b) the law of statistical regularity and the law of inertia of large numbers.

c) the law of statistical regularity and the law of inertia of small numbers.

d) the law of statistical irregularity and the law of inertia of small numbers.

5. Sampling and 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 inference about them. As such the sampling errors are totally absent in a census enumeration.

Sampling errors can never be completely eliminated but can be minimize by choosing a proper sample of adequate size.

ii) **Non Sampling Errors or Bias:** As distinct from sampling errors, the non-sampling errors primarily arise at the stages of observation, approximation and processing of the data and are thus **present in both the complete enumeration and the sample survey.** These error usually arise due to faulty planning, defective schedule of questionnaire from non-response from the respondents.

iii) Sampling error is totally absent in "Complete Enumeration" or "Census" But, Non-Sampling errors are present in both "Complete Enumeration" and "Sample survey"

- Parameter is a statistical measure on population. Statistic is a statistical measure on sample.

Related MCQ's:

25. How many different kind of errors can one find in sampling process?

a) One b) Two
c) Three d) Many

26. A sample survey is prone to:

a) Non-sampling errors b) Sampling errors
c) Either a) or b) d) Both a) and b)

27. Bias is also known as:

a) Sampling Error b) Non-Sampling Error
c) Error d) None of the above

28. Sampling error are:

a) Particularly detectfull
b) Can be corrected
c) Arise because the information collected relates only to a part of the population.
d) All of the above.

29. Sample Error is completely absent is:

a) Complete Enumeration b) Census
c) Both a) and b) above d) None of the above

30. _Can occur in census.

a) Standard Error b) Sampling Error
c) Bias d) None of the above

31. Non- Sampling Errors include :

a) bias b) mistakes
c) both bias and mistake d) none of these

32. Errors are likely to be more in case of complete enumeration:

a) Sampling errors
b) Probability errors
c) Non sampling errors
d) None of the above109.

33. "Sampling errors are present both in census as well as a sample survey." -State whether the given statement is correct or not.

a) Correct
b) Incorrect
c) Nothing cannot be said
d) None of the above

34. There are more chances of non-sampling errors than sampling errors in case of :

a) studies of large samples.
b) complete enumeration.
c) inefficient investigators.
d) all of the above

6. Sampling Distribution of a Statistic

From a population of size N, number of samples of size n can be drawn. These samples will give different values of a statistic. E.g. if different samples of size n are drawn from a population, different values of sample mean are obtained. The various values of a statistic thus obtained, can be arranged in the form of a frequency distribution known as Sampling Distribution. Thus we can have sampling distribution of sample mean x , sampling distribution of sample proportion p etc.

Errors in Sampling

Any statistical measure say, mean of the sample, may not be equal to the corresponding statistical measure (mean) of the population from which the sample has been drawn. Thus there can be discrepancies in the statistical measure of population, i.e., parameter and the statistical measures of sample drawn from the same population i.e., statistic. These discrepancies are known as Errors in Sampling.

Standard Error of a Statistic

Standard error is used to measure the variability of the values of a statistic computed from the samples of the same size drawn from the population, whereas standard deviation is used to measure the variability of the observations of the population itself.

The standard deviation of the sample statistics is called standard error of that statistic. E.g. if different samples of the same size n are drawn from a population, we get different values of sample mean $\bar{x}$. The S.D. of $\bar{x}$. is called standard error of $\bar{x}$ . . It is obvious that

the standard error of $\bar{x}$. will depend upon the size of the sample and the variability of the population.

i) Standard error of sample mean SE ($\bar{x}$) = $\frac{\sigma}{\sqrt{n}}$ or $\frac{s}{\sqrt{n}}$

σ=Population S.D

and s=Sample S.D

ii) Standard error of proportion SE (p) = $\sqrt{\frac{P\ (1\ -\ P)}{n}}$ or $\sqrt{\frac{p(1-p)}{n}}$

Where P=Population proportion P=Sample proportion

If i) Population size is Finite and the **Sampling Fraction** $\frac{n}{N} \geq .05$

And ii) **Samples are drawn Without Replacement(SRSWOR)**

Then , each of the above formula for Standard Error will be multiplied by the factor

$\sqrt{\frac{N-n}{N-1}}$ (Finite Population correction or Finite Population Multiplier)FPC

- Formula for standard Error when i) n<30(small sample)

ii) Population S.D σ is unknown in such a case SE ($\bar{x}$)= $\frac{s}{\sqrt{n-1}}$

The following table will provide us a better understanding of the situations while calculating SE ($\bar{x}$)

Sample Size	Parameter	Formula
Large (n ≥ 30)	SD is known	$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
Large (n ≥ 30)	SD is unknown	$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$
Small (n < 30)	SD is known	$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
Small (n < 30)	SD is unknown	$SE_{\bar{x}} = \frac{s}{\sqrt{n-1}}$
Rule of multiplying FPC will remain unaltered in a cases		

Summary

Concept of Sampling Distribution of Statistic and Standard Error:

⇨ Samples can be drawn with or without replacement

- ⇨ Probability distribution of a statistic is called sampling of statistic. Example: sampling distribution of ($\bar{x}$)., sampling distribution of (p)
- ⇨ Standard deviation of the sampling distribution of the sampling is called Standard Error of statistic
- ⇨ As sample size increases standard error decreases proportionately.
- ⇨ Precision of the sample is reciprocal to standard Errors..
- ⇨ Standard Error measures sampling fluctuations. i.e fluctuations in the value of statistics due to sampling

Related MCQ's:

35. Concept of Sampling Distribution is offer talked about in context of.
    a) Statistical
    b) Quantitative Analysis
    c) Sampling Analysis
    d) None of these

36. Values of a particular statistic with their relative frequencies will constitute the of the concerned statistic.
    a) Probability Distribution
    b) Sampling Distribution
    c) Theoretical Distribution
    d) None of these

37. We can have sampling distribution of:
    a) mean only.
    b) standard deviation
    c) both mean and standard deviation
    d) any statistical measure

38. In general mean of the sampling distribution is _ as the mean of the population.
    a) more than
    b) less than
    c) same
    d) none of the above

39. The standard deviation of sampling distribution is commonly known as:
    a) probability error
    b) human bias
    c) simple error
    d) standard error

40. The population standard deviation describes the variation among elements of the universe, whereas, the standard error measures the:

a) variability in a statistic due to universe
b) variabillity in a statistic due to sampling
c) variablity in a parameter due to universe
d) variablity in a statistic due to parameter

41. As the units selected in two or more samples drawn from a population are not the same, the value of a varies from sample to sample, but the always remains constant.

a) mean, standard deviation
b) statistic, standard deviation
c) statistic, parameter
d) parameter, statistic

42. The greater the value of standard error implies:

a) More the departure of actual frequencies from the expected ones.
b) More unreliability of the sample.
c) Both a) and b) above
d) None of the above

43. Standard error can be described as:

a) The error committed in sample survey
b) The error committed in estimating a parameter
c) Standard deviation of a statistic
d) The error committed in sampling.

44. The reciprocal of the standard error is:

a) Precision of the sample
b) Error of the sample
c) Error of the Universe
d) None of the above

45. Precision of random sample:

a) increases directly with increase in sample size
b) increases with the increase in sample size
c) increases proportionately with sample size
d) none of these.

46. The standard error of the is the standard deviation of sample means.

a) Population b) Sample c) Mean d) Median

47. Sampling Fluctuations may be described as :

a) the variation in the values of a statistic.

b) the variation in the values of a sample.

c) the differences in the values of a parameter.

d) the variation in the values of observations.

7. Types of Sampling

A sample can be selected from a population in various ways. Different situations call for different methods of sampling. There are three methods of Sampling:

1. Random Sampling or Probability Sampling Method

2. Non-Random Sampling or Non-Probability Sampling Method.

3. Mixed Sampling

1 Random Sampling or Probability Sampling

Random Sampling: Random or Probability sampling is the scientific technique of drawing samples from (he population according to some laws of chance in which each unit in the universe or population has some definite pre-assigned probability of being selected in the sample. It is of two types.

(a) Simple Random Sampling (SRS):

It is the method of selection of a sample in such a way that each and every member of population or universe has an equal chance or probability of being included in the sample. Random sampling can be carried out in two ways.

1. **Lottery Method:** It is the simplest, most common and important method of obtaining a random sample. Under this method, all the members of the population or universe are serially numbered on small slips of a paper. They are put in a drum and thoroughly mixed by vibrating the drum. After mixing, the numbered slips are drawn out of the drum one by one according to the size of the sample. The numbers of slips so drawn constitute a random sample.

2. **Random Number Method:** In this method, sampling is conducted on the

basis of random numbers which are available from the random number tables. The various random number tables available are:

a. Trippet's Random Number Series;

b. Fisher's and Yales Random Number Series;

c. Kendall and Badington Random Number Series;

d. Rand Corporation Random Number Series;

One major disadvantage of random sampling is that all the members of the population must be known and be serially numbered. It will entail a lot of difficulties in case the population is of large size and will be impossible in case the population is of infinite size.

(b) Restricted Random Sampling:

It is of three types

- Stratified Sampling
- Systematic Sampling
- Multi-stage Sampling

Stratified Sampling: In stratified random sampling, the population is divided into strata (groups) before the sample is drawn. Strata are so designed that they do not overlap. An elementary unit from each stratum is drawn at random and the units so drawn constitute a sample. Stratified sampling is suitable in those cases where the population is hetrogeneous but there is homogeneity within each of the groups or strata.

Advantages

(i) It is a representative sample of the hetrogeneous population.

(ii) It lessens the possibility of bias of one sidedness.

Disadvantages

(i) It may be difficult to divide population into homogeneous groups.

(ii) There may be over lapping of different strata of the population which will provide an unrepresentative Sample.

Systematic Sampling: In this method every elementary unit of the population is arranged in order and the sample units are distributed at equal and regular intervals. In other words, a sample of suitable size is obtained (from the orderly arranged population) by taking every unit say tenth unit of the population. One of the first units in this ordered arrangement is chosen at random and the sample is computed by selecting every tenth unit (say) from the rest of the lot. If the first unit selected is 4, then the other units constituting the sample will be 14, 24, 34, 44, and so on.

Advantages: It is most suitable where the population units are serially numbered or serially arranged.

Disadvantages: It may not provide a desirable result due to large variation in the items selected.

Multi-stage Sampling: In this sampling method, sample of elementary units is selected in stages. Firstly a sample of cluster is selected and from among them a sample of elementary units is selected. It is suitable in those cases where population size is very big and it contains a large number of units.

2 Non-Random Sampling or Non-Probability Sampling Method

A sample of elementary units that is being selected on the basis of personal judgment is called a non-probability sampling. It is of four types.

- Purposive Sampling;
- Quota Sampling;
- Convenience Sampling;
- Sequential Sampling.

Purposive Sampling: Purposive sampling is the method of sampling by which a sample is drawn from a population based entirely on the personal judgement of the investigator. It is also known as Judgement Sampling or Deliberate Sampling. A randomness finds no place in it and so the sample drawn under this method cannot be subjected to mathematical concepts used in computing sampling error.

Quota Sampling: In quota sampling method, quotas are fixed according to the basic parameters of the population determined earlier and each field investigator is assigned with quotas of number of elementary units to be interviewed.

Convenience Sampling: In convenience sampling, a sample is obtained by selecting convenient population elements from the population.

Sequential Sampling: In sequential sampling a number of sample lots are drawn one after another from the population depending on the results of the earlier samples draw from the same population. Sequential sampling is very useful in Statistical Quality Control. If the first sample is acceptable, then no further sample is drawn. On the other hand if the initial lot is completely unacceptable, it is rejected straightway. But if the initial lot is of doubtful and marginal character falling in the area lying between the acceptance and rejection limits, a second sample is drawn and if need be a third sample of bigger size may be drawn in order to arrive at a decision on the final acceptance or rejection of the lot. Such sampling can be based on any of the random or non-random method of selection.

Advantages of Random (OR Probability) Sampling

1. Random sampling is objective and unbiased. As a 'result, it is defensible before the superiors or even before the court of law. 8
2. The size of sample depends on demonstrable statistical method and therefore, it has a justification for the expenditure involved.
3. Statistical measures, i.e. parameters based on the population can be estimated and evaluated by sample statistic in terms of certain degree of precision required.
4. It provides a more accurate method of drawing conclusions about the characteristics of the population as parameters.
5. It is used to draw the statistical inferences.
6. The samples may be combined and evaluated, even though accomplished by different individuals.
7. The results obtained can be assessed in terms of probability, and the sample is accepted or rejected on a consideration of the extent to which it can be considered representative.

3 Mixed Sampling

Cluster Sampling: Cluster Sampling involves arranging elementary items in a population into hetrogeneous subgroups that are representative of the overall population. One such group constitutes a sample for study.

Related MCQ's:

48. Simple random sampling is very effective if
    (a) The population is not very large
    (b) The population is not much heterogeneous
    (c) The population is partitioned into several sections.
    (d) Both (a) and (b)

49. Simple random sampling is
    (a) A probabilistic sampling (b) A non- probabilistic sampling
    (c) A mixed sampling (d) Both (b) and (c).

50. According to Neyman's allocation, in stratified sampling
    (a) Sample size is proportional to the population size
    (b) Sample size is proportional to the sample SD
    (c) Sample size is proportional to the sample variance
    (d) Population size is proportional to the sample variance.

51. Which sampling provides separate estimates for population means for different segments and also an over all estimate?

(a) Multistage sampling (b) Stratified sampling
(c) Simple random sampling (d) Systematic sampling

52. Which sampling adds exibility to the sampling process?

(a) Simple random sampling (b) Multistage sampling
(c) Stratified sampling (d) Systematic sampling

53. Which sampling is affected most if the sampling frame contains an undetected periodicity?

(a) Simple random sampling (b) Stratified sampling
(c) Multistage sampling (d) Systematic sampling

54. Which sampling is subjected to the discretion of the sampler?

(a) Systematic sampling (b) Simple random sampling
(c) Purposive sampling (d) Quota sampling.

8. SAMPLING WITH REPLACEMENT (SRSWR)

While selecting the units for a sample, when a unit of sample selected is replaced before the next unit is selected then it is called sampling with replacement.

In this case the total number of samples that can be drawn = $(N)^n$ For E.g.: Let Population = {a, b, c}

N = 3, let n = 2

No. of samples = (N)n = $(3)^2$ = 9

No. of samples = {(a, b) (a, c) (b, c) (b, a) (c, a) (c, b) (a, a) (b, b) (c, c)}

9. SAMPLING WITHOUT REPLACEMENT (SRSWOR)

While selecting the units for a sample, when a unit of sample is selected but not replaced before the next unit is selected then it is called Sampling Without Replacement.

In this case the total number of samples that can be drawn = For E.g.: Let population = {a, b, c}

N = 3, let n = 2

No. of samples = ${}^{N}C_{n} = {}^{3}C_{2} = {}^{3}C_{1} = 3$

No. of samples = {(a, b), (a, c), (b, c)}

Related MCQ's:

55. A population comprises 3 members 1, 5, 3. Draw all possible samples of size two (i) with replacement (ii) without replacement

Find the sampling distribution of sample mean in both cases.

56. In simple random sampling with replacement, the total number of possible sample with distinct permutation of member is:

(N = Size of Population, n = Sample size)

a) N x n b) N^n c) N d) n

57. In simple random sampling without replacement, the total number of possible sample with distinct permutation of member is:

(N = Size of Population, n = Sample size)

a) N^n b) P(N, n) c) C(N,n) d) None of the above

58. If from a population with 20 members, a random sample without replacement of 2 members is taken, the number of all such samples is :

a) 400 b) 190 c) 210 d) 200

59. If drom a population with 25 members, a random sample with replacement of 2 members is taken, the number of all such samples is:

a) 50 b) 300 c) 625 d) 125

60. If from a population with 25 members, a random sample without replacement of 3 members is taken, the number of all such sample is:

a) 3400 b) 1250 c) 3400 d) 2300

Theory Answers

1	B	11	D	21	A	31	C	41	C	51	B
2	A	12	B	22	D	32	C	42	C	52	D
3	A	13	D	23	B	33	B	43	C	53	D
4	C	14	C	24	B	34	D	44	A	54	C
5	C	15	A	25	B	35	C	45	C	55	Read note
6	D	16	B	26	D	36	B	46	C	56	B
7	A	17	A	27	B	37	D	47	A	57	C
8	A	18	C	28	D	38	C	48	D	58	B
9	B	19	C	29	C	39	D	49	A	59	C
10	A	20	B	30	C	40	B	50	B	60	D

Note : Students shall workout in the class for prof