**Sampling and Sample survey**

A sample survey is a survey which is carried out using a sampling method, i.e. in which a portion only, and not the whole population is surveyed.

For example, suppose we want to know the percentage of total cultivable land of Bangladesh or the percentage of farmers using organic manures in Bangladesh. In such cases, a part of the total cultivable area or a fraction of all farmers is selected using statistical techniques for collection of necessary data. Here, the selected area or a fraction of farmers is a sample and the method of selecting a sample is called sampling.

**Errors in sample survey:**

* Faulty selection of the sample: Use of a defective sampling technique introduces some bias.
* Substitution: Sometimes investigators deliberately substitute a convenient member of the population for a difficult sampling unit.
* Faulty demarcation of sampling units: This type of bias is particularly significant in area surveys, such as agricultural experiments.
* Improper use of statistics for parameter estimation.

**Simple Random Sampling:** It is a scientific sampling method based on probability theory. In this method, each population unit has the same probability for being included in the sample if the population is homogeneous.

**Advantages and Disadvantages of Simple Random Sampling:**

**Advantages:**

* In simple random sampling each population element has equal probability to be included in the sample.
* Investigator’s personal bias is completely eliminated in this sampling process.
* Most representative sample can be obtained if the population is homogeneous.

**Disadvantages:**

* For very large population the process of numbering may be difficult and complicated.
* If the population is not homogeneous, sample drawn in this method may not be representative.
* If the sample units are geographically far away from each other, data collection often becomes expensive and troublesome.

Probable samples and the corresponding sample means will be,

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample  Elements | 1,2 | 1,3 | 1,4 | 1,5 | 2,3 | 2,4 | 2,5 | 3,4 | 3,5 | 4,5 |
| Sample  Mean | 1.5 | 2 | 2.5 | 3 | 2.5 | 3 | 3.5 | 3.5 | 4 | 4.5 |

Therefore, mean of the sample means and population mean are equal. Hence, the estimate is unbiased.

**H.W:**

1) Suppose, you have a population whose elements are 3, 4, 5, 6, 7 and 8.

**Stratified Random Sampling:** If the population is not homogeneous (the population elements are not similar) in respect of the characteristic under study, a simple random sample may not properly represent the population.

In such cases, the whole population is divided into a number of more or less homogeneous subdivisions, these subdivisions are called strata. From each of these subdivisions, separate random selections of elements are made to constitute a sample. This method of sampling is known as stratified random sampling.

The strata should be such that:

* Elements included in each stratum should be as far as possible of homogeneous nature.
* Elements of different strata should be as far as possible of different nature.

**Solution:** We have,

The required sample will be of size,

Therefore, the sample of size 60 will consist of 40 small farmers, 15 medium farmers and 5 big farmers.

**Solution:** Here,

The population is to be stratified on the basis of profession and income as shown below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Profession** | **Income Group** | | **Total** |
| **Mid income** | **Low income** |
| Farmer |  |  | 300 |
| Businessman |  |  | 100 |
| Factory worker |  |  | 200 |
| Service holder |  |  | 100 |
| **All** | 210 | 490 | 700 |

So the composition of the sample will be as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| **Profession** | **Income Group** | | **Total** |
| **Mid income** | **Low income** |
| Farmer |  |  | 60 |
| Businessman |  |  | 20 |
| Factory worker |  |  | 40 |
| Service holder |  |  | 20 |
| **All** | 42 | 98 | 140 |

The sample will be of size 140 of which 42 and 98 represent the medium and low income group respectively.

Among the 42 mid income people, 12 are farmers, 14 are businessmen, 8 factory workers and 8 service holders.

Among the 98 low income people, the corresponding numbers will be 48, 6, 32 and 12 respectively.

**Sampling of Attributes**

In the sampling of attributes we are concerned with the possession or non-possession of some attribute by the individual selected in sampling. A simple sample of n individuals may be regarded as a series of n independent trials with constant probability p of success. We have the following two tests:

If the difference between expected and observed values is more than 2.58 (1% level) there is a reason to doubt the hypothesis. If the difference is less than 1.96 (5% level) the difference in proportion may be regarded as due to random sampling variation.

**Ex.1) A coin is tossed 10,000 times and the head turns up 5195 times. Would you consider the coin biased?**

**Solution:** Let us take the hypothesis that the coin is unbiased.

Since the difference between expected and observed values is more than 2.58 (1% level) there is a reason to doubt the hypothesis. Hence, we will consider the coin biased.

**Ex.2) In a sample of 500 persons 310 were found to be wheat-eaters and the rest rice-eaters. Can we conclude that both the food articles are equally popular?**

**Solution:** Let us take the hypothesis that, both the food articles are equally popular.

Since the difference between expected and observed values is more than 2.58 (1% level) there is a reason to doubt the hypothesis. Hence, we conclude that both the food articles are not equally popular.

**Ex.3) The figures of deaths in a hospital from general anesthesia and local anesthesia are given below:**

|  |  |  |
| --- | --- | --- |
|  | **Cases** | **Deaths** |
| **General anesthesia** | 200 | 10 |
| **Local anesthesia** | 480 | 12 |

**Can it be concluded that there is a significance difference in general anesthesia and local anesthesia.**

**Solution:** Let us take the hypothesis that there is no difference in the number of deaths from general anesthesia and local anesthesia.

Since, the difference is less than 1.96 (5% level) it could have arisen due to fluctuations of sampling. Hence, there is no evidence to doubt the hypothesis.

**Ex.4) 500 articles from a factory are examined and of these 4% are found defective. 1300 similar articles from another factory are examined and of these 3% are found defective. Can we conclude from the data given that the products of the first factory are inferior to the second?**

**Solution:** Let us take the hypothesis that there is no difference in the percentage of defective items in the two factories.

Since, the difference is less than 1.96 (5% level) it could have arisen due to fluctuations of sampling. Hence, there is no evidence to conclude that, the products of the first factory are inferior to the second.

**H.W:**

3) In town A 10000 persons were observed and 20% of them were found to bear spectacles. In town B 25000 persons were observed and 18% of them were found to bear spectacles. Does this data lead you to infer that, there is a significant difference in the two towns with regard to the percentage of persons bearing spectacles is concerned?

4) In a random sample of 700 from a particular district 200 are found to be smokers. In another district out of 1300 people 400 are found to be smokers. Can you conclude that, there is a significant difference between the two districts with regard to the smoking habit?