Answer of Question paper 2017-18

Q1(A). Define simple random sampling. Differentiate between SRSWOR and SRSWR.

Ans. The simplest of the methods of probability sampling which is usually called the method of random sampling. In this method, an equal probability of selection is assigned to each available units of the population at the first and each subsequent draw. Thus, if the number of units in the population is N, then the probability of selection of any unit at first draw is 1/N and at the second draw is 1/N-1 etc, which are ultimately equal to 1/N. The sample obtained using the above method is called “ Simple Random Sampling”. Since this result is independent of the specified unit it follows that every one of the units in the population has the same chance of being included in the sample under the procedure of simple random sampling.

Difference Between SRSWOR and SRSWR:

(i) If the selected units are not being replaced back in the population before the second draw, it is called SRSWOR and if the selected units are being replaced back in the population before the second draw, it is called SRSWR
(ii) In SRSWOR, at each draw ,new information on the units will be generated while it may be possible to have the same kind of information on the units in SRSWR.
(iii) SRSWOR method will cover the whole population units while it is not true in the case of SRSWR.
(iv) The variance of sample mean in SRSWOR is found to be smaller rather than that of SRSWR providing more efficiency in SRSWOR.

(B) Define systematic sampling with four examples.

we have considered methods of sampling in which the successive units (whether elements or clusters) were selected with the help of random numbers. We shall now consider a method of sampling in which only the first unit is selected with the help of random numbers, the rest being selected automatically according to a predetermined pattern. The method is known as systematic sampling. The pattern usually followed in selecting a systematic sample is a simple pattern involving regular spacing of units. Thus, suppose a population consists of N units, serially numbered from 1 to N. Suppose further that N is expressible as a product of two integers k and II, so that N = kn. Draw a random number less than k, say i, and select the unit with the corresponding serial number and every k-th unit in the population thereafter. Clearly, the sample will contain the n units i, i + k, i + 2k, ... ,(i + n -1)k, and is known as a systematic sample. The selection of every k-th strip in forest sampling for the estimation of timber, the selection of a corn field, every k-th mile apart, for observation on incidence of borers, the selection of every k-th time-interval for observing the number of fishing craft landing on the coast, the selection of every k-th punched card for advance tabulation ,after the first unit is chosen with the help of random numbers less than k, are all examples of systematic sampling.

(C)Sampling and Non –Sampling errors:

Sampling error:

If complete accuracy can be ensured in the procedures such as determination, identification and observation of sample units and the tabulation of collected data, then the total error would consist only of the error due to sampling, termed as sampling error. Measure of sampling error is mean squared error (MSE). The MSE is the difference between the estimator and the true value and has two components: - square of sampling bias. - sampling variance.

Non Sampling Errors

It is a general assumption in the sampling theory that the true value of each unit in the population can be obtained and tabulated without any errors. In practice, this assumption may be violated due to several reasons and practical constraints. This results in errors in the observations as well as in the tabulation. Such errors which are due to the factors other than sampling are called non-sampling errors.

Sources of non-sampling errors:

Non sampling errors can occur at every stage of planning and execution of survey or census. It occurs at planning stage, field work stage as well as at tabulation and computation stage. The main sources of the non-sampling errors are

(i) lack of proper specification of the domain of study and scope of investigation,
(ii) incomplete coverage of the population or sample,
(iii) faulty definition,
(iv) defective methods of data collection and
(v) tabulation errors.

(D) Explain the situations when cluster sampling is used.

A sampling procedure, pre-supposes the division of the population into a finite number of distinct and identifiable units called the sampling units. Thus a population of fields under wheat in a given region might be regarded as composed of fields or groups of fields on the same holdings, villages, or other suitable segments. A human population might similarly be regarded as composed of individual persons, families, or groups of persons residing in houses and villages. The smallest units into which the population can be divided are called the elements of the population, and groups of elements the clusters. When the sampling unit is a cluster, the procedure of sampling is called cluster sampling. When the entire area containing the population under study is subdivided into smaller areas and each element in the population is associated with one and only one such small area, the procedure is alternatively called area sampling. For many types of population a list of elements is not available and the use of an element as the sampling unit is therefore not feasible. The method of cluster or area sampling is available in such cases. Thus, in a city a list of all the houses is readily available, but that of persons is rarely so. Again, lists of fields are not available, but those of villages are. Cluster sampling is, therefore, widely practiced in sample surveys. The size of the cluster to be employed in sample surveys therefore requires consideration. In general, the smaller the cluster, the more accurate will usually be the estimate of the population character for a given number of elements in the sample. Cluster sampling is used for conducting a larger survey when the sampling frame (List of the units in the population) is not available.

(E) Define PPS sampling. Explain about cumulative total method.

In simple random sampling without replacement, we assume that samples are selected with equal probability for all the units in the population. If the units vary considerably in size, SRSWOR may not be appropriate since it does not take in to account the possible importance of the varying sizes of the units in the population. In fact, a larger unit for variable Y may contribute more to the population total rather than the smaller units. For example, villages having larger geographical areas are likely to have larger population and larger areas under food crops. It is therefore natural to expect that a scheme of selection which provide the probability of selection in the sample to larger units also than to smaller units giving more efficient estimators in comparison to equal probability. Such type of sampling which vary from probability to probability ccording to the size of units. It is called “Probability Proportional to Size” (PPS) sampling.

In the cumulative total method , the following steps are considered:

(i) A random number M is selected from" Random Number Table" from 1 to T_N where T_N is the total of all units in the population.

(ii) If this random number M falls between $T_{i-1} < M \leq T_i$, then ith unit is selected.

(iii) This process is repeated for n times for getting n samples from the lot.

(iv) It is to be noted that here if a random number gets repeated in the procedure, the corresponding same unit is taken in the sample, then the selection of the units is called sampling with replacement.

Q2(A) Explain the important points for planning and organization of a sample survey.

Write three –four lines in each of the following points:

(i) Objectives
(ii) Data to be gathered
(iii) Population under investigation
(iv) Sampling frame
(v) Methods of collecting data
(vi) Organization and supervision of field work
(vii) Tabulation of data
(viii) Analysis of data
(ix) Precision
(x) Writing reports and conclusions

2(B) Derive a relationship between mean square error, sampling variance and bias.

Let us suppose that $\hat{\theta}$ be the estimator of the parameter θ. Then mean square error is defined as MSE= $E(\hat{\theta}-\theta)^2$

subtracting & adding $E(\hat{\theta})$ in the right hand side, we have

MSE= $E[\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta]^2$

Squaring R.H.S. taking two terms together , we have

MSE= $E[\hat{\theta}-E(\hat{\theta})]^2+E[(\hat{\theta})-\theta]^2+2E[\hat{\theta}-E(\hat{\theta})(\hat{\theta}-\theta]$

The third term in R.H.S. vanishes equal to zero,then

MSE= $E[\hat{\theta}-E(\hat{\theta})]^2+E[(\hat{\theta})-\theta]^2$, which is equal to

MSE= S.V. + Bias^2 ,Since S.V.= $E[\hat{\theta}-E(\hat{\theta})]^2$ and Bias= $E[(\hat{\theta})-\theta]^2$

Q3. Define stratified random sampling and write the advantages of stratification. What are the choice of sample sizes in different allocations?.Derive the expression of variance of sample mean in proportional and Neyman allocations.

STRATIFIED SAMPLING (definition)

We know that the precision of a sample estimate of the population mean depends upon two factors: (1) the size of the sample, and (2) the variability or heterogeneity of the population. Apart from the size of the sample, therefore, the only way of increasing the precision of an estimate is to devise sampling procedures which will effectively reduce the heterogeneity. One such procedure is known as the procedure of stratified random sampling. It consists in dividing the population of N units into K groups of sub population of $N_1,N_2,\dots N_k$ units respectively. These sub population are non-overlapping and together they can comprise the whole of the population, so that $N_1+N_2+\dots N_k=N$. These sub-population are called strata and the single group is called stratum. The sample size within the strata are denoted by $n_1,n_2,\dots n_k$ respectively such that $n_1+n_2+\dots n_k=n$. K is called the number of strata or groups.

If a simple random sample is taken in each stratum using SRSWOR, the whole procedure is described as" Stratified Random Sampling"

Advantages of stratification:

(i) If the admissible error is given, a small sample should be taken so that our expenditure may be reduced.
(ii) There is a reduction of error due to stratification, if the cost of survey is fixed.
(iii) Stratification provides the individual means of stratum and then for the whole population.
(iv) Stratification may provide the administrative convenience.

Choice of Sample Sizes in Different Strata

There are four types of choice of sample sizes in different strata.

(i) Equal allocation: ni=n/k where n is the total sample size and k is the number of strata.
(ii) Proportional allocation: ni=npi where pi= Ni/N stratum weight.
(iii) Optimum allocation: It involves cost per unit in the stratum
(iv) Neyman allocation: A particular case of optimum allocation when ci=c ,a constant cost for all the units ni= npisi/$\sum pisi.$

Derivation of the variance of sample mean in proportional and Neyman allocation:

We know that $V(\bar{y}_{st}) = (\frac{1}{n_i} - \frac{1}{N_i})\sum_{i=1}^{k} p_i^2 S_i^2$

Where pi=Ni/N stratum weight and Si^2 is the population mean square.

Substituting ni=npi in the above expression, we have the variance of sample mean in proportional allocation.

$V(\bar{y}_{st})_p = (\frac{1}{np_i} - \frac{1}{N_i})\sum_{i=1}^{k} p_i^2 S_i^2 = (\frac{\sum_{i=1}^{k} p_i^2 S_i^2}{n} - \frac{1}{N}\sum_{i=1}^{k} p_i^2 S_i^2)$ which is written in simplified form as $V(\bar{y}_{st})_p = (\frac{1}{n} - \frac{1}{N})\sum_{i=1}^{k} p_i S_i^2$.

Similarly, substituting ni= npisi/$\sum pisi.$ in the above expression, we have the variance of sample mean in Neyman allocation .

$$V(\bar{y}_{st})_{Ney} = (\frac{1}{ni} - \frac{1}{Ni})\sum_{i=1}^{k} p_i^2 S_i^2 = [\frac{(\sum pisi)^2}{n} - \frac{1}{Ni}\sum_{i=1}^{k} p_i^2 S_i^2] = [\frac{(\sum pisi)^2}{n} - \frac{1}{N}\sum_{i=1}^{k} p_i S_i^2]$$