Paper ECOH-502 STATISTICS FOR ECONOMICS

Syllabus

(2)

Unit-I

Collection and Presentation of Data - Methods of Data Collection - Population & Sample - Questionnaire - Tabulation of Data - Diagrammatic Representation

Unit-II

Measures of Central Tendency - Dispersion - Moment and Skewness

Unit-III

Bivariate Frequency Distribution - Linear Correlation - Scatter Diagram - Pearson's Co-efficient of Correlation Rank Correlation.

Unit-IV

Concept of Regression Analysis and its Properties - Linear Regression between two variables

Unit-V

Concept of Probability - Addition Theorem & Multiplication Theorem - Mathematical Expectation Theorems - Problems.

Unit 2 (Stats

1. Define statistical data

Ans → By statistical data we mean facts represented by numerical figures relating to a particular phenomenon under study.

2. What are the types of data?

Ans → Statistical data are mainly of two types –
(i) Primary data and
(ii) Secondary data.

(i) Primary Data :- Primary data are those data which are collected by the investigator for the first time for a specified purpose. They are also called original data. These data are fresh right from the beginning to the end. They are called primary because these are collected from the very source of information. for example, national income data collected by central Statistical Organisation (a govt. body) are primary data, RBI bulletin (monthly) [Issued by RBI, Mumbai]

(ii) Secondary Data :- Data collected by someone and used by someone else is called secondary data. Secondary data are ~~already in existence and which~~ numerical information which have been previously collected as primary data by some agency for a specific purpose but are now compiled from that source for use in a different connection. For example, national income data collected by the government are primary data, but the same data will become secondary when used by different investigators. Similarly, data in the

Bulletin (collected by the Research Department of the Reserve Bank) are primary but will become secondary in the hands of other investigators.

What are the main sources of secondary data?
The main sources of secondary data can broadly be classified into two ~~types~~ parts :-
(i) Published Sources, and
(ii) Unpublished Sources.

(i) Published Sources :- The various sources of published data are given below :-

A) Government Publications :- Various government departments regularly collect and publish data which are very useful and reliable. Some of the important government publications are as under :-
a) The Statistical Abstract of India published by the Central Statistical Organisation.
b) Five Year Plans
c) Economic Surveys.
d) Annual Budgets
e) RBI bulletins
f) Report on Currency & Finance
g) Census of India.
h) National Sample Survey

B) Semi Government Publications :- It includes semi-official publications of various local bodies such as municipal committees, corporations, panchayats etc.

c) International Publications :- The Statistical Yearbook and the Demographic Yearbook are two important UN publication containing information on socio-economic topics and on population respectively.

Similarly, Epidemiological and Vital Records is still another important monthly periodical on health and mortality published by WHO.

D) Private Publications such as publications of
a) Stock Exchanges
b) Institute of Chartered Accountants
c) Labour Unions
d) Newspapers such as the Economic Times, The Financial Express etc

E) Research Publications :- Research institutions also publish their research work. Publications of universities, research institutes and individuals fall in this category.

F) International Institutions such as IMF, World Bank etc also provide useful statistical information about various countries of the world.

(ii) Unpublished Sources :-
Secondary data are also available from unpublished sources, because all statistical material is not always published. For example, information recorded in various govt. & private offices, studies made by research institutions, etc.

4 What are the precautions in the use of secondary data?

Before using the secondary data, the following precautions are to be taken into consideration :-
i) Firstly, the investigator must ensure that

or

no data are suitable for the purpose of enquiry. He must see the purpose for which the data were originally collected and its suitability for the purpose of investigation now in view.

ii) Secondly the investigator should also see whether the data are adequate for the investigation. Adequacy of the data is to be judged ~~on the content of the requirements of the survey by the available data~~ on the basis of three considerations:-

a) Requirements of the present investigation
b) Geographical area covered by secondary data
c) Time period for which secondary data are available

iii) Thirdly, the investigator must ensure whether the data are reliable. For this purpose, the following points need to be taken into consideration:

a) The status of the agency or authority which collected the data.
b) Was the collecting agency unbiased?
c) The method used for collecting the data.
d) Were the employed enumerators efficient and trained?
e) Was the editing, tabulation, and analysis of the data carefully and properly done?
f) What degree of accuracy was desired & achieved by the compiler etc.

Thus, we can conclude that if data are found satisfactory in all respects, then only they can be used. They should not be taken as they are. It is necessary to verify them for any inaccuracy.

unsuitability or inadequacy

5 What are the differences between primary data & secondary data?

Ans ⇒ Data are primary to an individual or institution collecting them, while they are secondary in the hands of others. Infact, there is not much difference between primary & secondary data.

However, the main points of difference between primary & secondary data are as follows:-

Primary Data	Secondary data
(i) Primary data are those which are collected originally.	(i) Secondary data refer to those data which are collected through other sources.
(ii) Primary data are in the form of raw material to which statistical methods are applied.	(ii) Secondary data are in the form of finished product as they have already been statistically applied.
(iii) Primary data have been collected for a definite purpose.	(iii) Secondary data are collected from published or unpublished sources.
(iv) A lot of time and money is required for collection of primary data.	(iv) Since secondary data are collected through other sources, so both time & money are saved
(v) Primary data are strictly in accordance with investigation	(v) Secondary data are made according to the object of investigation.

tween primary & secondary data, which of the two kinds of data are better?

Between primary and secondary data, the choice depends upon the following considerations.

(i) Nature and scope of the enquiry
(ii) Availability of time & money
(iii) Degree of accuracy
(iv) The status of the investigator, i.e individual, private agency & government etc.

However, most of the statistical analysis rests upon the secondary data. Primary data are generally used only in those cases where secondary data does not provide adequate & accurate information. Nowadays secondary data are increasingly being used as they are easily available. In certain cases both primary and secondary data may be used.

1. Advantages & Disadvantages of Primary Data Over Secondary Data:-

Advantages:-

1. Primary data contain more detailed information while some information are compressed or condensed in secondary data
2. Secondary data may contain errors like ~~transcr~~ transcribing errors etc while primary data cannot have such errors.
3. In the primary data, precise definitions of the terms used and the scope of the data are mentioned while this is is not available in secondary data

4. Primary data are collected with the specific requirement of the research in mind and hence, such data are likely to be more relevant than secondary data.

5. Primary data are current information whereas secondary data may not be very current & thus may not reflect the present scenario.

Disadvantages :-

1. Collection of primary data generally needs a lot of time and involvement of a number of investigators depending on the nature of the survey which is not required for the collection of secondary data.
2. Collection of primary data is an expensive process whereas secondary data collection are cheap in monetary terms.
3. If the information is to be collected from a large number of places or respondents, it is difficult to perform the same for an individual or small organisation. In such a case the individual or organisation has to depend on secondary data.

8. Methods of Collecting Primary Data :-

Primary data may be collected by using any of the following methods :-

i) Direct Personal Investigation.
ii) Indirect Oral Investigation.
iii) Information through Correspondents
iv) Information through Mailed Questionnaire.
v) Questionnaire through enumerators

Direct Personal Investigation :-

Under this method data are collected personally by the investigator. Data are collected by asking questions related to the enquiry. The investigator has to go to the spot and meet the people concerned.

For example suppose that an investigator wants to collect data about the wages of the workers of a particular factory, he would go to that factory and obtain required information by contacting the workers personally.

Merits:-

a) Since the investigator approaches the people personally they would supply the desired information without much hesitation.

b) The information obtained is most likely to be reliable and accurate because the interviewer can remove doubt of the informants.

c) Direct personal investigation is more appropriate where the scope of enquiry is limited & accurate information is highly required.

Demerits:-

a) This method is unsuitable where the number of persons to be interviewed is large and they are spread over a large area.

b) This method suffers from the chances of personal bias.

c) It involves unnecessary wastage of time & money.

Data can also be collected through telephonic interview by the investigator.

(ii) Indirect Oral Investigation :-

Under this method, the investigator collects the information by interviewing persons not those from whom information is to be collected instead he makes enquiries with some other persons who are directly or indirectly in touch with them.

For example, in case of a road accident, the investigator (police) can enquire the persons present at or near the place of accident.

Merits :-

a) Indirect oral investigation is economical. There is not much wastage of time and money.

b) This method is most suitable where the scope of enquiry is very wide.

c) Data or information collected under this method would not be biased.

Demerits :-

a) Under indirect oral investigation there is always a possibility of data being wrong, if the informants do not possess true knowledge.

b) Since data are not collected from the concerned persons directly, they are liable to be wrong.

c) There is fear of biased information. Due to bribery or certain other reasons, the informants can give wrong witness.

(iii) Information Through Correspondents :-

In this method, the investigator does not get information personally rather he appoints local agents or correspondents in different parts of the field of enquiry. These agents or correspondents would collect & send the required

tion to a central office where the data are essed Radio, television and newspaper agencies generally dopt this method This method is also adopted by various government departments where regular information is to be collected from a wide area.

Merits :-

a) Information through correspondents is economical
b) This method provides information quickly
c) It is most suitable in those cases where data is to be collected from a large area

Demerits :-

a) The job for obtaining information is left to the correspondents and, thus, collected data lack accuracy and originality
b) Data are likely to be biased.
c) Data is send by correspondents through using different methods. There is lack of uniformity of data.

(iv) Information through Mailed Questionnaire :-

This method of collecting primary data is becoming very popular these days. Under this method, a questionnaire (a list of questions to be asked) is sent to the informants by post along with a covering letter to fill up the questionnaire and send it back within a specified time.

Merits :-

(a) This method is most suitable where field of enquiry is wide
(b) It is also economical. Fairly large amount of information can be received by spending less time & money
(c) Data are reliable since they are directly given by information.

Demerits:-
a) This method is obviously limited to literate persons.
b) Informants generally do not send this questionnaire duly filled up or they are send back incomplete.
c) This method is not flexible.

(v) Questionnaire through Enumerators:-
Under this method, information is collected on the basis of a questionnaire through the enumerators (or interviewers). The enumerator himself goes to the ~~informator~~ informants. Questionnaire is filled up by him on the basis of the answers given by the informants. ~~Respondent~~

Merits:-
(a) It can be adopted even in those cases where informants are illiterates.
(b) Complex and difficult questions can be easily asked
(c) There is very little possibility of no reply.

Demerits:-
(a) This method is very costly because a lot of money is spent on enumerators in form of wages, travelling allowances, dearness allowances.
(b) If enumerators are not efficient and trained, the informed received will be misleading
(c) The enumerators may be biased.

9. Differences between Questionnaire & Schedule:-
In the collection of primary data, questionnaires & schedules are used Both of them, show a set of questions But there is a difference between the two. A questionnaire is, filled up by the informant whereas in case of a schedule answers are to be filled up by enumerators.

questionnaire

questionnaire may be sent either by post or through enumerators, but in case of schedule both investigator and informants have personal contact. The use of questionnaire is suitable where the informants are literate, whereas, schedule can be used both for literate and illiterate persons.

10 Define the terms :- (i) Investigator (ii) Enumerator (iii) Respondent

(i) Investigator :- The person who plans and conducts the statistical investigation is called the investigator. He may conduct the investigation independently or with the help of others.

(ii) Enumerator :- Enumerator is the person who goes out to the field and collects the desired statistical information. He is the person who actually collects the data.

(iii) Respondent :- Respondent is a person who provides the actual data. He responds by answering the questions included in the questionnaire.

What are the essentials of a good questionnaire?

A questionnaire is a list of questions relating to the field of enquiry and provides space for answer. It may be defined as an instrument of collecting primary data from a large number of persons.

The success of the investigation largely depends upon the proper drafting of the questionnaire. A good questionnaire must possess the following essentials :-

1. Covering letter :- The person conducting the survey must introduce himself and make the aims and objectives of the survey clear to the informants. This can be done by attaching a covering letter with the questionnaire.

2. Questions should be Simple & Clearly Worded
questions should be short, simple and clearly worded. Unless the person (to be interrogated) is technically qualified, difficult and technical terms should not be used

3. The number of questions should be small as far as possible. Only those questions should be asked which are related to the investigation. The questionnaire should be ~~concise~~ concise. The precise number of question would depend upon the objective of the investigation

4. Questions should be properly arranged :- The questions should be arranged in a logical manner. The questions may be divided and sub-divided into groups & sub-groups

5. Personal Questions should be avoided :- As far as possible, personal questions such as questions about sources of income, volume of sales etc. should not be asked

6. Questions should be strictly related to the investigation

7. 'Yes' or 'No' Questions :- As far as possible & practicable, the questions should be answered in 'Yes' or 'No' so that informants do not find much difficulty in answering them

8. Prohibited Questions should not be asked. For example, in many cases, people are generally reluctant to disclose the truth

9. Questions which require calculations should also be avoided. For example, informants should not be asked yearly income, whereas he gets his income on a monthly basis

10. If possible, some cross-questions should be made to find out the truth

11. If a particular question is of complex nature,

same should be made clear in footnote.

12 Clear cut instructions should be given to the informa so that they can fill up the questionnaire correctly and neatly.

13 Some incentive for filling up the questionnaire may be given.

14 Pre-testing of the questionnaire :- Before sending the questionnaire to the informants, it should be pre-te As a result of it, shortcomings, if any can be removed.

12 What is pilot survey ?

Once the questionnaire is prepared, it is advisable to conduct a survey on a small group for pre testing the questionnaire. This is known as pilot survey. Th this pilot survey, the investigator comes to know the drawbacks of the questions included in the questionna It also helps in assessing the performance of enumerator and cost and time involved in the actual survey. Finally it also helps in formulating effective methods of askin questions and also in the improvement of the questio

Use of Pilot Survey :-

(i) They can be used for obtaining information regarding the variability in the population.

(ii) The cost of the survey under various sampling techniques can be known.

(iii) This survey is also used to build up cost and variance functions which can be used for planning future surveys.

(iv) To train the investigator as to how they should fill up the schedule.

13 What are the Sources of Data?

Ans = There are two main sources of data:

(i) Internal Sources.

(ii) External Sources.

Internal sources of data are those which are obtained from the internal reports of an organisation. For instance, Hindustan Level publishes its annual report on total production, total profit and loss, total sales, loans, wages to employees, bonus and other facilities to employees etc. Information so received is the internal source of data

External sources refer to those information collected from other sources. They can be collected from primary as well as secondary sources For instance, to study the problem of transportation in 'Haryana' if we obtain the information from Haryana Roadways, it would be known as external sources of data

14 Distinction between Internal & External Data:-

Ans =

Internal Data	External Data
(i) Internal data refer to those data which have been obtained from the internal records of an organisation	(i) External data refer to those which have been obtained from other sources.
(ii) Reliance Industry Ltd. publishes its annual report which would constitute an internal source of data for the company.	(ii) When Reliance Industries Ltd obtains data from sources other than its own records, this would be external source of data.
(iii) These are less expensive.	(ii) These are more expensive.
(iv) These data are more reliable.	(iv) These data are comparatively less reliable.

What do you mean by population?

In statistics population may be defined as the aggregate of the entire group of items, units or subject which is under the reference of study. Population may consist of finite or infinite number of units which may be animate or inanimate. Population is also termed as universe. The total number of individuals in the population is called the population size.

The inhabitants of a region, no of wheat fields in a state, fruit plants in an orchard etc are examples of finite population. All real numbers, all stars in the night sky are examples of infinite population.

6. What do you mean by 'sample' and 'sampling'?

Ans = Sampling is simply the process of learning about the population on the basis of a part of the population that is drawn from the population. This part of the population is called the sample. A sample is that part of the population that is expected to exhibit the characteristics of the population. In sampling theory, instead of studying every unit of the population only a part of the universe is studied and the conclusions are drawn on the basis of the sample for the entire universe.

What is 'Parameter' and 'Statistic'?

Ans = The values obtained from the study of a sample, such as the mean, median, variance, dispersion etc is known as 'statistic' and the corresponding value of population is called as 'parameter'. The parameter being a population quantity or a function of population quantities is a constant quantity. The parameters are generally denoted by Greek letters like, ~~the~~ for

Statistic from the Sample may be the best for the parameter will be 1.

μ for population mean
σ^2 for population variance.
ρ for population correlation coefficient etc

A statistic on the other hand is a function of sample observations and hence is found to vary from sample to sample. Thus the value of the statistic is like a variable. Such values are generally represented by Latin letters in smaller case like,

$\bar{x}$ for sample mean
s for sample standard deviation
r for sample correlation co-efficient etc.

The process of deriving statistic is economical in time & cost than parameter.

18 What is Unbiased Estimator?

Ans: Let 't' be a statistic used for the estimation of the parameter 'θ' (say). The estimator 't' is said to be an unbiased estimator of θ if and only if, $E(t) = \theta$. If $E(t) \neq \theta$ then 't' is called as a biased estimator for θ.

19 What are the different types of Population?

Ans: Population can be fundamentally classified into four categories:-

(i) Finite population
(ii) Infinite population
(iii) Real population
(iv) Hypothetical population

(i) Finite population:- In this population the number of members of the population is fixed and limited. The population of babies born today in a particular health centre in a town is an example of finite population.

(ii) Infinite Population:- Here the population consists of

of stars in a night sky. The real numbers lying between 0 and 1.

(iii) Real Population: A population which consists of the items that are all present physically is termed as real population. for eg, the collection of all stars in a night sky, The real numbers lying between 0 & 1.

for eg, if we visit a health centre town and want to know the birth weights of all the newborn babies that were born in the last five years, we can get it from the Record Department of the Health Centre. Here each and every member of the population actually exist today.

(iv) Hypothetical Population - Let us consider another example in which one is interested in noting down the birth weights of newborn in the medical centre in the next five years. This is an example of hypothetical population because here the population does not exist in reality.

2) Basic Principles of Sample Survey:

1 Statistical Regularity: This principle is based on the theory of probability which may be considered as the basis for the selection of a random sample. The random sample is used in order to give each and every unit of the population equal chance of being included in the sample. In such a case the sample will start reflecting the heterogeneity present in the entire population and hence forth the sample becomes a representative of the population.

2. Validity:- This ask the sample design to be obtained in such a manner that it can be used for the purpose

obtaining valid estimates and tests about the p... further
The samples obtained by using a random technique
this principle

3. Optimisation - The word optimization means extreme case
i.e either maximization or minimization. Thus, the sampling
technique should be developed in order to achieve both
maximization and minimization. The maximization is done
by increasing the efficiency for a fixed cost and minimisation
is attained by reducing the expenditure incurred in the
survey for a fixed level of efficiency

2) Advantages of Sampling Over Complete Enumeration.

a) In sample survey only a part of the population is to be
investigated and hence it takes less time, less money
and less labour.

b) In census method, a large group of investigators are
employed - a thorough and intensive training of such
a large group of people can hardly be possible. But
in sample survey method, only a small group of
enumerators are employed - they are given specialised
training. Hence, though a little bit expensive, the output
is much more accurate.

c) Sample survey can with a limited budget can cover more
geographical area.

d) In sample survey, sampling error cannot be avoided. But
in census there is no question of sampling errors

e) Sometimes complete enumeration is not feasible or
practicable. This is especially the case when the subject
under study gets destroyed or cannot be used any

further on testing for instance, to examine the quality of sweets, in a sweet shop one has to restore to sampling methods, which is the only way out. Similarly for the blood test in the laboratory one has to resort to sample study only.

22 Limitations of Sampling -

The following are the chief disadvantages or limitations of sampling :-

i) In order to obtain accurate results a sample survey is to be properly planned and executed. Otherwise incorrect and misleading results may be obtained.

ii) Sampling requires the services of experts. If there is scarcity of such people, sampling may give unsatisfactory results owing to the use of faulty methods of selection inappropriate sampling design, or inefficient methods of estimation

iii) Complicated sampling plans require more labour than a complete count in the long run.

iv) There are various sources of errors in sample surveys. Every attempt must be made to minimise the chances of errors, otherwise correct conclusions cannot be expected.

23. Short notes:-

(i) Sampling Fraction :- The sampling fraction is the ratio between the sample size to the population size. Here, size means number of units. Thus population size refers to the number of units in the population and similarly sample size is the number of units in the sample.

Symbolically, Sampling Fraction = $n : N$ (i.e. n/N)

where, n = Sample Size.

N - Population Size.

Sometimes the sampling fraction is expressed in percentage.

(ii) Sampling Unit :- The individuals that are to be selected/sampled from the population to form the sample is called as sampling units.

For example, in order to know the consumption pattern of families of a town each family may be considered as a sampling unit but to know the income of individuals of a town each individual will now become a sampling unit.

(iii) Sampling Frame:- The sampling frame is an exhaustive list of sampling units. Such a list may be either in tabular form or in the form of maps in case of area surveys. The simplest example of sampling frame is the voter list, which provides the complete list of individuals along with their individuals and other details of a particular area.

(iv) Sampling Distribution :- If we consider all possible random samples of the same size drawn from the same population, and calculate the values of a statistic from each of them, the values will, in general, not be the same. The difference will occur due to difference in the composition of the samples. A statement of the distinct values of the statistic obtained from the different samples together with the corresponding probabilities is called sampling distribution.

(v) Sampling Fluctuations :- Since the composition of the different samples, drawn from the same population, are usually different, the value of a statistic is likely to vary from one sample to another. These differences in the values of a statistic are called its sampling fluctuations.