Introduction and Data collection

Statistics

Statistics is the science and art of collecting, summarizing, analyzing and interpreting of data that are subject to random variation. The word statistics derived from Latin word status, Italian word statista, and German word staistik which means political state.

The word statistics used two different senses

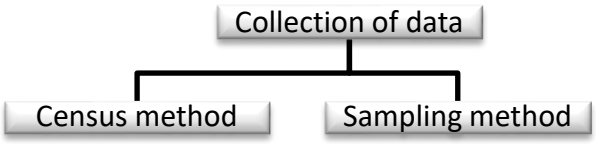
1. Singular sense
2. Plural sense

1) Singular sense

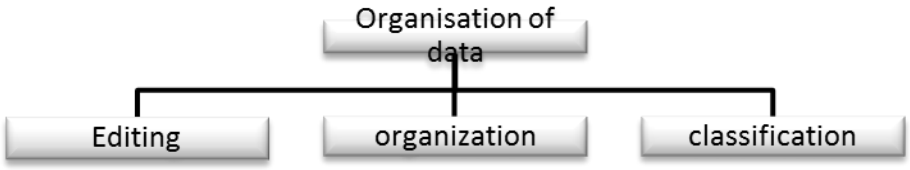
It is known as "Statistical Method", these statistical methods are collection, classification, presentation, analysis, interpretation of data and forecasting.

Different steps of statistical method in singular statistics.

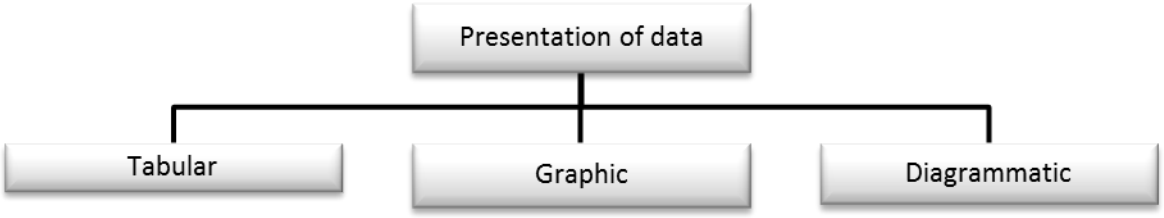
Step 1



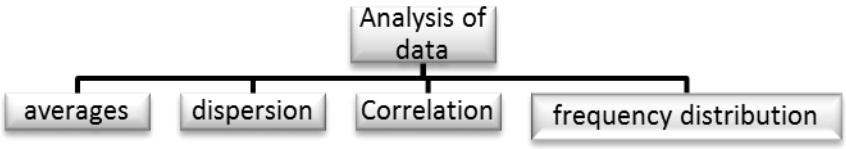
Step2



Step3



Step 4



2) Plural form :

In this sense, it refers to the figure of data (numerical data) collected in systematic manner.

It is used as plural noun "statistic" which refers any one of many computed or estimated statistical quantities, such as the mean, standard deviation or the correlation coefficient.

1) **Descriptive statistics**

It is the statistical procedure which deal with the collection, representation, calculation and processing. It involves graphical and tabular approaches to describe, summarize and analyze the data

It describes the large data using only few numbers (like mean, median, range, standard deviation etc). The large number of census data of the country can be summarized in just a few numbers like, total population, sex ratio, age distribution etc.

2) **Inferential statistics**

Inferential statistics is the branch of statistics that generalized the result of small sample to the larger population. It is also known as sampling statistics.

For example;

To test the efficacy of drug of cancer disease, we take small sample from the cancer patients and summarize it to the whole cancer patients.

Applied statistics

This consists of massive application of theoretical or mathematical statistics in the areas of biology, physics, astronomy, meteorology, chemistry, medicine, sociology, business, economics and so on. The statistical tools and methods are used in order to solve many more practical problems in diversified area. moreover, applied statistics has played very important role in decision making problems as well.

Functions of statistics

The main functions of biostatistics are given below:

- To simplify complexities

Statistics consists of aggregate of numerical facts. Huge facts and figures are difficult to remember. The complex mass of figures can be made simple and understandable with the help of statistical methods.

- To present facts in definite form

One of the important functions of statistics is to present the general statements in precise and definite form. The conclusion stated numerically is definite and hence more convincing than the conclusions stated qualitatively.

- To estimate for the present and forecast for the future

While preparing suitable policies and plans, it is necessary to have the knowledge of future tendency. Statistical methods provide helpful means in forecasting the future by studying and analyzing the tendencies based on past records.

- To facilitate comparison

The science of statistics does not mean only counting but also comparison. Unless the figures are compared with other figures with the same kind, they are meaningless. Statistical methods such as average, ratio, rates, percentage, coefficients etc offer the best way of comparison between two phenomena which will enable to draw valid conclusion.

- To help in formulation of policies;

Statistics helps in formulating the policies in different fields. The government policies are also framed on the basis of statistics. In fact, without statistics, suitable policies can not be framed.

- To help in formulation and testing the hypothesis

Statistical methods are helpful not only in estimating the present, forecasting the future but also helps in formulating and testing the hypothesis for the development of new theories.

Limitations

Statistics, with its wide applications in almost every sphere of human activity is not without limitations. The followings are some of its important limitations:

- Statistics is not suited to the study of qualitative phenomenon.

Qualitative phenomenon can not be expressed numerically are not capable of direct statistical analysis. However statistical techniques may be applied indirectly by first reducing the qualitative expressions to precise quantitative terms.

- Statistics does not study individuals

Individual items, taken separately do not constitute statistical data and are meaningless for any statistical enquiry.

- Statistical laws are not exact

Unlike physical and natural sciences, statistical laws are only approximation and not exact. Statistical conclusions are not universally true – they are true only on average.

- Statistics is liable or be misused

Statistics must be used by experts. The use of statistical tools by in experiences and untrained persons might lead to very fallacious conclusion. As King says “Statistics are like clay of which one can make a god or devil as one pleases.”

Variable

The quantitative characteristics which can take different values under study is termed as variable.or The characteristics of the measurement differ from one biological entity to another or place to place, or time to time is known as variable. There are different kinds of variable encountered by biologist, the example are sex, cast , weight, blood pressure, pulse rate, occupation, blood group, disease, hair color, religion, education level etc,

Generally variables are denoted by X, or Y.

Types of variable (statistical Data)

There are two types of variable, such as

1. Categorical (Qualitative) variable.
2. Numerical (Quantitative) variable.

Qualitative variable

Such types of variable which have no notation of magnitude of size of characteristics, it can just categorized is known as qualitative variable.

For example disease: Present/ absent

Blood pressure: Low / high

Gender: Male / female

Vaccine: Vaccinated /not vaccinated

Smoking: smoking / non-smoking

These types of variable are known as dichotomous variable or binary variable which is also known as nominal data.

Smoking: non- smoking/ ex-smoker/light smoker/ heavy smoker

Quantitative variable

The types of variable or data which can be express as a number or magnitude are known as quantitative variable. Which have numerical measure.

For example:-

- No of child in a family
- Size of house hold
- Pulse rate
- Blood pressure
- Action of drug
- Profit
- Death
- Temperature
- Number of patients
- Weight

Types of quantitative variable

- Discrete variable
- continuous variable

Discrete variable

A variable is said to be discrete if it countable many values (whole number) Discrete variable can only have a finite number of values over a certain range.

No of children in a family: - 3, 3, 2, 6, 7

Size of the house hold: - 7, 4, 3, 2

WBC count: - 4050, 3010, 5005

Bed in a hospital: -50, 75, 100, 500

Continuous variable

A continuous value of variable which can take any value of variable over a particular range is known as continuous variable, it contains fractional or integral value.

Most of the biological variables are continuous in nature,

For example;

Wight of the children, Height of the patients etc.

Data

The collective information or numbers is called as data. Data are collected to getting the necessary information from the units under investigation. In the health field data are collected for planning, policies making etc. data are the raw materials for final statistical conclusions. Generally data can either qualitative or quantitative.

Types of data

1. Primary data
2. Secondary data

1) Primary Data

Primary data are those fresh and original forms of data, which are collected and recorded by investigator or researcher. They are first had and collected for specific purpose of study. Such data are to be unique, original, reliable and accurate in character. These types of data are obtain in the survey and enquires conducted by government, some individuals, institutions and research bodies.

Primary data collection technique

- ✓ Observation method
- ✓ Questionnaires survey (post/ mail questionnaires)
- ✓ Interview (personal / telephone)
- ✓ Case studies
- ✓ Information through key persons.
- ✓ Focus group discussion.
- ✓ Schedule sent through enumerator. Etc

2) <u>Secondary data</u>

Any data which have been collected earlier for some purpose are secondary data for the individuals who are using them. The data is a primary for those persons or institution's that collect them but the same data become secondary for another. Actually data are the data. Which are borrowed from other who have collected them for some other purpose. The degree of accuracy of this type of data is comparatively less than that of the primary data.

The data obtained from some secondary source (published or unpublished) such as journals, magazines, newspaper etc. is known as secondary data. These types of data are not in original character. For example, the data of AIDS for world reported by WHO becomes secondary for researcher.

Sources of secondary data

- ✓ Census of household and population
- ✓ Budgets
- ✓ Demographic health survey
- ✓ Nepal living Standard survey

- ✓ UN publication
- ✓ Who publication
- ✓ News paper
- ✓ Journals
- ✓ Magazines etc.

Advantages of secondary data

Following are the advantages of secondary data.

- It saves time and cost.
- If specially trained persons collect it, the quality of secondary data is better.
- The scope of inquiry can be increased in terms of area a d time period to be covered.

Disadvantages

- Any a time the exact definition of terms and units used in a secondary data are not known.
- When the secondary data are obtained from two sources, it may not be comparable in terms of definitions, units and time period covered.
- Some information is often omitted or some categories are pooled
- Data may not be in the exact form of the requirement of the researcher.

Method of data collection

The data can be collected by two ways

1) Census method
2) Sample method

Census method:

A census defined by united nations as “the total process of collecting, compiling and publishing demographic, economics and social data pertaining at a specified time or times, to all persons in a country or delimited territory.

Or

The complete enumeration of all units of population is known as census survey. It is the process of complete enumeration in which every member of the defined population is included.

Advantages

- ✓ The data high degree of accuracy
- ✓ More representative and true
- ✓ Result are more reliable
- ✓ Minimum biased

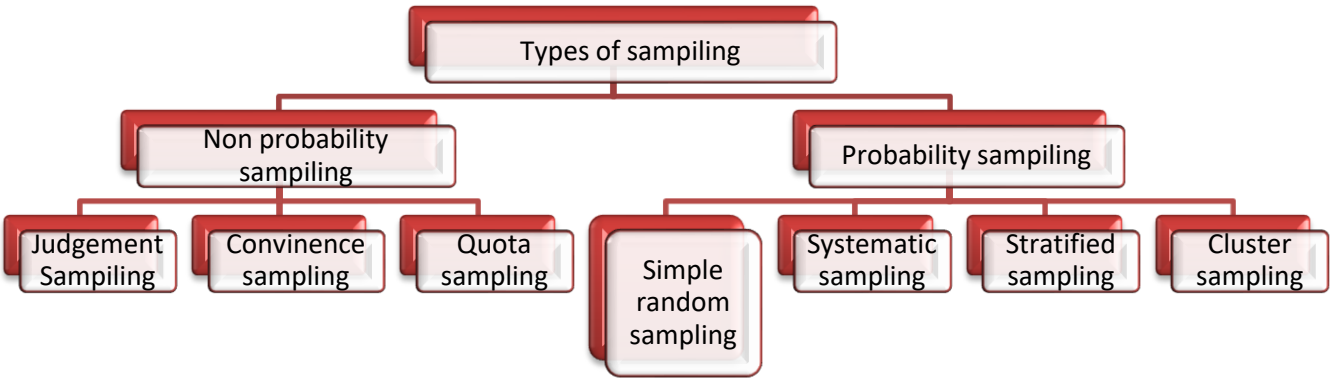
Disadvantage

- ✓ More cost, labor, and energy required
- ✓ Large number of investigator or enumerators are required
- ✓ It can’t be applied in all situations.

Sampling

The data is collected from small group of population is termed as sample. The process of getting sample from the population is known as sampling.

For example: A doctor examines a few drops of blood and draws conclusion about the blood constitution of whole body.



Advantage of sampling:

a. This method is more economical than census because it consumes less time less expenditure etc.

b. It requires less number of investigators.

Disadvantages.

a. It requires skill manpower

b. Appropriate sampling technique may not select.

The experimental method

Experiment is a controlled observation done to test the truth fullness of the hypothesis relating to some research problem. In experiment, the experiment has control over the circumstances. That is varies only the variable whose effect is to be studied keeping other constant and the outputs are noted down. Mathematically, an experiment can be described by an equation [functional relation];

$Y = f(X_1, X_2, \ldots \ldots X_n)$; Where Y = dependent variable or output, f= function and x_1, x_2…….x_n are independent variables also known as inputs.

Example: - In an agricultural research, suppose, Y = yield of crop/ output

It is affected by a number of independent variables, say

X_1 = Seyed quantity

X_2=Irrigation amount

X_3= Amount of manure

X_4= Labour input

The researcher may be interested in finding the effect of change in any of these variables, say X_2, on the output Y. Here, he keeps X_1, X_3&X_4 constant and controls X_2 as per his will.

Types of experiments:-

There are two types of experiments, they are;

1. Laboratory experiment: - The experiment related to a research problem that can be isolated from real world situation is called laboratory experiment. Here, the experimenter has full control over the experiment.

2. Field experiment: - If the problem is to be studied in its real setting, the experiment done is called field experiment. Here, the experimenter uses some statistical control method [e.g. replication, randomization etc.] to manipulate several variables at once.

In field experiments, the researcher often uses *experimental and control group* to control such experiments. In experimental group, there is group of people [experimental units] subjected to treatment [experimental variable] and the control group is the group of people which is not subjected to the experimental treatment, e.g. to study the effect of certain training, data on two groups of people before trainings collected. One group is, then, trained and again data on two groups is collected and effectiveness of training is studied.

Problems in experimentation with human population: - Human being is the most dynamic, and conscious object to be taken as experimental unit. There is continuous interaction of human being with their environment and change in situation. So, the human population may constantly change the responses causing problem in experimental on them.

The survey method

The literal meaning of 'survey' is "to oversee" or to look from a high place. It can be defined as the technique of the investigation or collection of information about social aspect of a community's position and activities by direct observation of the phenomena or through interviews.

Although, the approach and technique of survey research can be used on any set of objects that can be well defined, survey research focuses on people, the vital facts of people and their belief, opinions, attitudes, motivations and behaviours. Sociological facts are attributes of individuals that spring from their membership in social group: sex, income, political and religious affiliations, socio- economic status, education, age, living expenses, occupation, race and so on.

Survey is concerned with specific persons, specific problems and situations.

The objective is to fulfil immediate needs and to use knowledge available at given time. It is the practical in nature.

Purpose of survey research:-

1. To describe current practices and events. In this sense it can be termed as polls .A polling survey is concerned mainly with the distribution of responses to any particular item.
2. To analyse the fact: - The result obtained from polling survey can be analysed and we may answer the question regarding the relationship between the variable involved.

Questionnaire;

A questionnaire is mailed to the respondents expecting them fill it and return to the sender. If population list is available, questionnaire method of data collection saves both time and money but it cannot be used if population list is not available.

Merits: - If the respondents are literate, mailed questionnaire reduces the time & cost of survey to a great extent. In this method, the respondents freely answer even the sensitive questions and they are free from enumerator's bias.

Demerits: - Questionnaire method of data collection suffers from a number of drawbacks. Some of them are,

1. It can't be used in case of illiterate audience.
2. Majority of respondents do not reply. So, non- response rate is considerably high.
3. Any confusion about the question [enquiry] can't be clarified.

Question

A question is an expression pertaining to the field of enquiry in the interrogative sentence that is related to facts, figures, knowledge and opinions of relevant to the study.

Keshav Lamichhane

Types of questions

The different types of questions are framed by social scientist in order to elicit different types of information on various aspects of problem. The following are main types of questions used in schedule.

1. Structured question: - A question provide with prefixed possible alternative replies is a structured question. It may be dichotomous [yes/no] or multiple choice question. Eg.

- Are you a literate or an illiterate? Employed?

- What is your marital status?

Married/ Unmarried / Widower / Widow/ Separated/ Divorced `

2. Open end questions: - The questions where the possible alternative replies cannot be exhausted are to be answered by the respondent after due thinking. Opinions, suggestions etc. are sought through such question. E.g.

- What measures would you suggest for national integration?

- What do you think are the main causes of present indiscipline among students?

3. Leading question: - The question, where the preferred answer is hinted, though, the respondent is free to supply the answer other than that mind. E.g.

-Should not so\mething be done?

- Is it not desirable that teachers must keep themselves fee from politics?

4. Ranking item question: - Here various alternatives, unlike multiple choice questions, are given but they have to be ordered based on the preference. E.g. which of the following professions you like best? Rank those in order of your preference, '1' for the most liked and so on.

- Administration ( )
- University teacher ( )
- Lawyer ( )
- Doctor ( )
- Engineer ( )
- Other (specify)

5. Ambiguous questions: - When the language of the question is such that it may be interpreted in more than one way, it is called ambiguous question. E.g.

- Do you like travelling on trains or buses?

- Do you give lectures?

- What kind of home do you have?

6. Presuming question: - On this question some presumption about the respondent is made.E.g.

a. How many cigarettes do you smoke a day? [Here presumption is that he does smoke]

b. Whom did you vote in the last election? [He has voted].

7. Hypothetical question: - Questions used to predict future behaviors of the respondent belong to this type. E.g.

c. What would you do if certain thing happens?

d. Would you buy a motor car, if the price is lowered by 25%?

8. Personalized questions: - Personalized question is asked to differentiate group action from personal action. Eg. The respondent may say that it is good to have every one's chest regularly checked by X-ray but to the question 'Have you ever had your chest X-rayed' he may say 'no'.

9. Behavior question: - Question eliciting [producing as response] the behaviour of the respondent is called behaviour question. E.g. How often do you go to cinema?

10. Memory question: - For the questions dealing with past, serious attention should be given to the respondent ability to recall the required information. A question whose answer requires the call of the memory of the respondent is memory question e.g. when did you buy your television?

11. Embarrassing question: - Questions whose answers are related to private life of the respondent and it is embarrassing to answer. In this case, the best method is to ask views of others. E.g. some women who use this cleaner find a lot of faults with it. I wonder, if you can guess what they are objecting to it.

Requisites of a good question

Nature of the question depends on the nature of the study, types of the respondents,quality of the field workers etc. Some general guidelines for good question are:

- The question should be few, short, clearly worded, simple and easy to reply.
- The question should be within the information scope of the respondents i.e., the respondent is in the position to reply.
- The question should have direct bearing with objectives of the research problem/ investigation.
- As far as possible, the units or technical terms used in schedule should be as in previous inquires.
- The question should be inter-related with each other.

- Cross checking questions should be given to avoid any manipulation by the respondents.
- The question should be impersonal and it should demand minimum of writing works.
- The question should be free from ambiguity.
- For the question dealing with the degree of intensity of feeling or convistion, the words- Why, What, When, How question should be included.
- The question should be such that it's answer facillates the classification and tabulation of the data.

The Questions To Be Avoided

The questions which are: too long, complex, ambiguous, leading, not relevant, embarrassing etc. should be avoided. These questions do not follow the requisites of good question too.

The words to be avoided

The words used in questions play important role in increasing response rate and accuracy of the responses. The words used should be simple, clear and unambiguous. The following are the words to be avoided in the questions.

- Technical terms, unless the survey is about the technical subjects.
- Native and unused words.
- Words carrying emotional connotions.
- Subjective or qualitative words.
- Ambiguous words.

Sequence of question

The guidelines for the sequencing the questions are as follows:

- They should be arranged from simple to complex and from general to specific.
- Subject matter sequence should be given priority than time sequence.
- Each question and each group should appear in a chain.
- The change from one subject to another should be very smooth.

Mailed Questionnaire: - If respondents are literate and their list is available mailed questionnaire is used to collect information. This helps greatly in reducing the field cost and saves time. It is sent to

selected respondents by post. It contains a request letter to fill up the questionnaire and send back to the sender.

Problems in questionnaire method: - The greatest problem in questionnaire method is high proportion of non- responses. It may be because the respondents do not clearly understand the question or they do not think necessary to answer it or though even filled up many respondents fail to return it. Some other factors affecting the response rate are as follows.

1] Special characteristics of the group: - the response rate is found to be linked with the status of the respondents [sex, age, income level etc.]. It is found that lower and upper income groups usually give low responses.

2] Prestige of sponsoring agency: - It is found that higher the reputation of the institute carrying the survey, higher the response rate is.

3] Importance of the problem: - For the enquiry on an important and significant problem, a higher responses are expected than in case of problem of general nature.

4] Nature of reaction of the respondents: - If the respondent possesses strong feeling in favour or against the problem, they are indifferent about the problem, they usually do not respond.

5] Inducement to reply; - Must of the respondents respond if they get benefit from it self-addressed envelope is an example of inducement.

6] Follow up: - Follow up process increases the response rate. If successive reminders are sent to non-respondents they are likely to respond ultimately.

Format of a questionnaire: - A questionnaire is sent to the respondents by post expecting that they duly fill it up and send it back. For this reason it should have certain characteristics that appear appealing to the respondent. These characteristics are:

- It is smaller than schedule [size]
- It is attractive in appearance
- The questions are clear.
- The questions are in natural sequence.

Its overall design should be such that it catches interest of respondents