بسم الله الرحمن الرحيم

L   XI

Biostatistics

METHODS IN DATA COLLECTION

L XI
17th -8-2025

Prof. Dr. Waqar Al-Kubaisy

Biostatistics consist of

1-Collection of data .

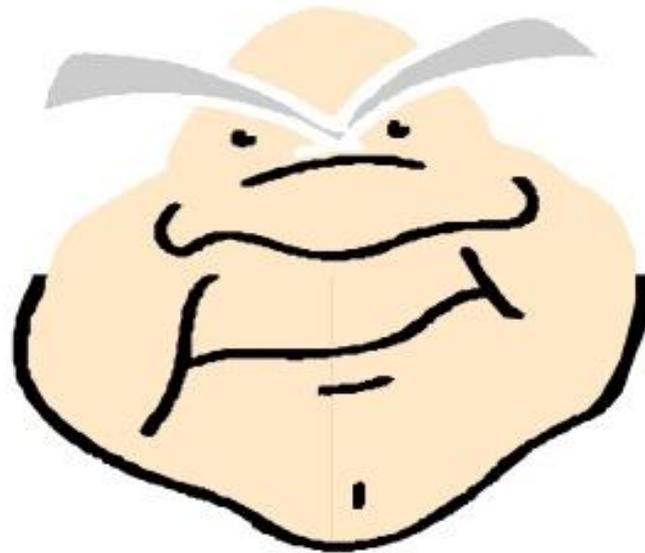**2-Presentation of data .**
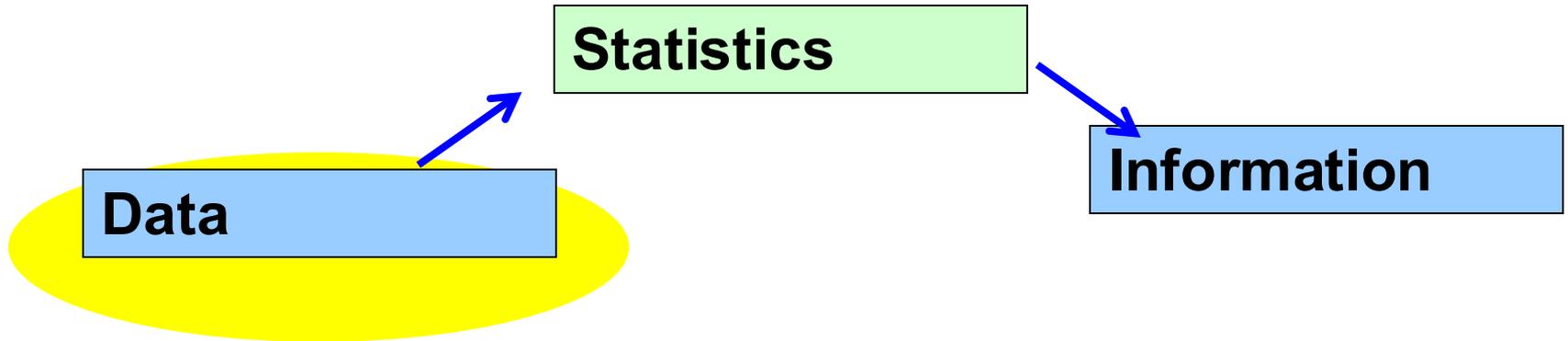
3- Estimation of data

DATA

Data is

➢ one of the most important and vital aspect of any research studies.

➢ -the basic unit in statistical studies

➢ every research is based on data which is analyzed and interpreted to get information.

❖ **Data can be quantitative or qualitative values of a variable**

Need of Data Collection

- To get information for analysis.
- To get idea about real time situation.
- For comparison between two situation.

Statistics is a tool for converting *data* into *information*:

Statistics

Data

Information

where does *data* come from?

How is it **gathered?**

How do we ensure its **accurate?**

Is the data **reliable**?

Is it **representative** of the population from which it was drawn?

Factors Should To Be Considered Before Collection Of Data

❖ **Objective and scope of the enquiry**

❖ **Sources of information**

❖ **Technique of da collection**

❖ **Unit of collection**

A COLLECTION TECHNIQUES

Primary data collection

- Data is collected by researcher himself

- Data is gathered through questionnaire, interviews, observations etc.

Secondary data collection

- Data collected, compiled or written by other researchers

eg. books, journals, newspapers
- Any reference must be acknowledged

Each type has its own weaknesses and strengths

PRIMARY DATA

1-Primary data means original

2--collected for the first time

3-collected from the original source first hand.

4- It is real time data,

5- collected by researcher himself .

 6-collected specially for the purpose in mind to address

the research problem

7-questions are **meaningful to the purpose**

8-uses surveys, experiments or direct observations

9-has **not been published yet**

10-has not been changed or altered by human beings;

11-its **validity is** greater than secondary data.

12--more reliable, authentic and objective.

•IMPORTANCE OF PRIMARY DATA

Importance of Primary data cannot be neglected.
A research can be conducted without secondary data
<div style="text-align:center">but</div>
a research based on only secondary data is least reliable and may have biases because secondary data has already been manipulated by human beings.

In statistical surveys it is necessary to get information from primary sources and work on primary data

Validity is

one of the major concerns in a research.
The quality of a research that makes it trustworthy and scientific. **the use of scientific methods in research to make it logical and acceptable**.
Validity" refers to the extent to which something is well-founded, sound, or legally or logically correct
For example,
• **does an intelligence test really measure intelligence?**
•**Does a self-esteem scale really measure self-esteem**

Reliability

is the **certainty that the research is enough true to be trusted on**. Reliability is the **certainty that the research is enough true to be trusted on**.

Primary data can be relied on because you know where it came from and what was done to it. Its like cooking something yourself. You know what went into

Authenticity is اصالة

•**the genuineness of the research trustworthiness**, and **accuracy** of the data collected, analyzed, and reported in a research study.

METHODS USED TO COLLECT PRIMARY SOURCE DATA

1. **Interviews**
2. **Questionnaires**
3. **Survey**
4. **Experimentation**
5. **Case Study**
6. **Observation**

SOURCES OF PRIMARY DATA

- **Sources for primary data are**

❖ limited and at times

❑ it becomes **difficult to obtain** data from primary source

•**because of either**

➢ scarcity of population or

➢ lack of cooperation.

•it is the most **authentic** and **reliable** data source.

•Following are some of the sources of primary data.

- •Experiments
- •Survey
- •Interview

SOURCES OF PRIMARY DATA

•Experiments

❖ Experiments **require an artificial or natural setting** in which to perform logical study to collect data.

❖ **more suitable for medicine**, and for other scientific studies.

❖ In experiments **the experimenter has to keep control over the influence of any extraneous variable** on the results.

SOURCES OF PRIMARY DATA

•**Survey**

•**Survey is most commonly used method in social sciences,**

management,

•**Surveys can be conducted in different methods.**

•**Questionnaire**

• is the most commonly used method in survey.

Questionnaires are **list of questions either**
an open-ended
close –ended
for which the respondent give answers.

Questionnaire can be conducted via
➢ **telephone,**
➢ **mail,**
➢ **live in a public area, or**
➢ **in an institute,**
➢ **electronic mail or through**
➢ **fax and**
➢ **other methods**.

Health Surveys

❖ **should be population-based Survey**

1-for investigating of factors affecting H and disease , environment, nutrition etc..

➢ **studying the natural history of disease and factors affecting**

2 Surveys related to administrative of H. Services
uses of H. services,
evaluation of population needs &
evaluation of medical care .

Population surveys

H survey needs sample technique

H survey could be

cross sectional,

     longitudinal ,

           descriptive,

               analytic or both

Survey Methods

1-H. interview survey face-to-face survey .
2-H. examination survey .
3-H. records survey .
4-Mailed questionnaire survey

The most common
data collection instrument

Useful to collect
information

**Survey
Questionnaire**

Should contain 3 elements:
1. Introduction – to explain the objectives
2. Instructions – must be clear, simple language & short
3. User-friendly – avoid difficult or ambiguous questions

2 Basic Types of survey questions:

1. **Open-ended Questions**
   - Free-response

     (Text Open End)
   - Fill-in relevant information

2. **Close-ended Questions**
   - **Dichotomous question**
   - **Multiple-choice**
   - **Scale**
   - **Categorical**
   - **Numerical**

Steps To An Effective Survey Questionnaire

Prepare your survey questions
(Formulate & choose types of questions, order them, write instructions, make copies)

⬇

Select your respondents/sampling
Random/Selected

⬇

Administer the survey questionnaire
(date, venue, time)

⬇

Tabulate data collected
(Statistical analysis-frequency/mean/correlation/%)

⬇

Analyze and interpret data collected

ADVANTAGES OF PRIMARY RESEARCH

- Targeted Issues are addressed

- Data interpretation is better

- Efficient Spending for Information

- Decency of Data.

- Proprietary Issues

- Addresses Specific Research Issues

- Greater Control.

DISADVANTAGES OF PRIMARY RESEARCH

- High Cost

- **Time Consuming**

Inaccurate Feed-backs

- **More number of resources is required**

Secondary Data

- Secondary data are those that have already been collected by others.

- These are usually in journals, periodicals, dailies, research publications, official records etc.

- Secondary data may be available in the published or unpublished form. When it is not possible to collect the data by primary method, the investigator go for Secondary method.

- This Data collected for some purpose other than the problem at hand.

SECONDARY DATA

❖ Secondary data is the data that has been already **collected by others**

❖ and readily available from other sources.

❖ may be obtained from **many sources** is being **reused**

❑ Primary Data is gathered by people who can focus directly on the purpose in mind

❑ Secondary data doesn't have the privilege of this focus

IMPORTANCE OF SECONDARY DATA

• Secondary data can be less valid but its importance is still there

• Sometimes it is difficult to obtain primary data.

• Sometimes primary data does not exist in such situation one has to confine the research on secondary data.

• The Respondents are not willing to reveal information

• so such secondary data can suffice (enough)

•SOURCES OF SECONDARY DATA

•Secondary data is **often readily available**.

•Availability of secondary data has **become much easier after the expense of the electronic media and internet.**

•**Published Printed Sources**

•Their credibility depends on many factors.
on the writer, **publishing company** and **time and date** when published.

New sources are preferred and old sources should be avoided

SOURCES OF SECONDARY DATA cont

Census

Registration of Vital Events

Hospital and Health Record

Disease registration

Other H.S records

Environmental health data

H. man power statistic

Epidemiological surveillance.

<u>Census</u>

An important source of HI

<u>**defined by U.N**</u>.

➢ **The total process of collecting, compiling & publishing.**

Demographics ,Economic and social data pertaining at a specified time or times to all country

➢   so

➢ **census is a massive process**

➢ **undertaking to contact every member** of the population in a **given time,**

➢ and collect a **variety of information** .

It needs

It needs

- ❖ **a considerable organization**
- ❖ **vast preparation and**
- ❖ **several years to analyze the result**

It is taken in most countries of the world,
at a regular intervals usually every 10 years

Registration of Vital Events

    e.g. Birth, death and marriages

whereas census is intermitted counting of population vital events keeps a continuous cheek on demographic changes .

> Census
> **Registration of Vital Events**
> Hospital and Health Record
> Disease registration
> Other H.S records
> Environmental health data
> H. man power statistic
>
> Epidemiological surveillance.

Hospital and Health Record

Basic and primary source of information about disease which is prevalent in the community .

1. constitute only the tip of the iceberg of disease only those patients who seek medical care .
2. mild case may not attend
3. The admission policy may vary

drawback

Usefulness

- H.care activities,
- utilization hospital
- **discharging sheath** contain useful information
- diagnosis,
- Medical, Surgical procedures complication ,
- **length of stay**,
- laboratory data

Census
Registration of Vital Events
Hospital and Health Record
Disease registration
Other H.S records
Environmental health data
H. man power statistic

Epidemiological surveillance.

Disease registration;

- Give duration of illness,
- case fatality and survival.
- These information allow **follow-up of pt.**
- **Provide continues** account of the frequency of disease ,
Natural Course of disease , especially chronic disease

Other H.S records;

Hospital out-pt. department.
- MCH center, BW, W, Height, Immunization,
arm circumference, disease
- School H record,
- **DM & Hypertension** clinic

Draw Back;
it is only related to certain segment of population

Environmental health data;

identification and quantification of factors causative of disease . eg.

 Air, water,   noise,

food additive ,industrial toxicants,

inadequate west disposal

H. man power statistics;

no. of physicians,

    dentists,

     Pharmacies

      nurse….

 Classified into age , sex specialties

Epidemiological surveillance;

I n many countries where there is endemic diseases, special control/ eradication programs have been insulated. as part of these program, surveillance system are often set –up, to report on the
Occurrence new cases and
Efforts to control the disease
these program had yielded a
 considerable morbidity and mortality rate for specific disease

E.g. control program against COVID 19. TB, Malaria, Dengue fever ,etc.

ADVANTAGES OF SECONDARY DATA

- **Ease of Access**
- Low Cost to Acquire

- **Clarification of Research Question**
- **May Answer Research Question**

May Show Difficulties in Conducting Primary Researching

DISADVANTAGES OF SECONDARY DATA

There are some disadvantages to using secondary research.

- **Quality of Research**

Not Specific to Researcher's Needs

Incomplete Information

Not Timely

Difference between Primary and Secondary Data

Primary Data	Secondary Data
Real time data.	Past data.
Sure about sources of data.	Not sure about sources of data.
Help to give results/finding.	Refining the problem.
Costly and Time consuming process.	Cheap and No time consuming process.
Avoid biasness of response data.	Can not know in data biasness or not.
More flexible.	Less Flexible.

best of luck! in your exams