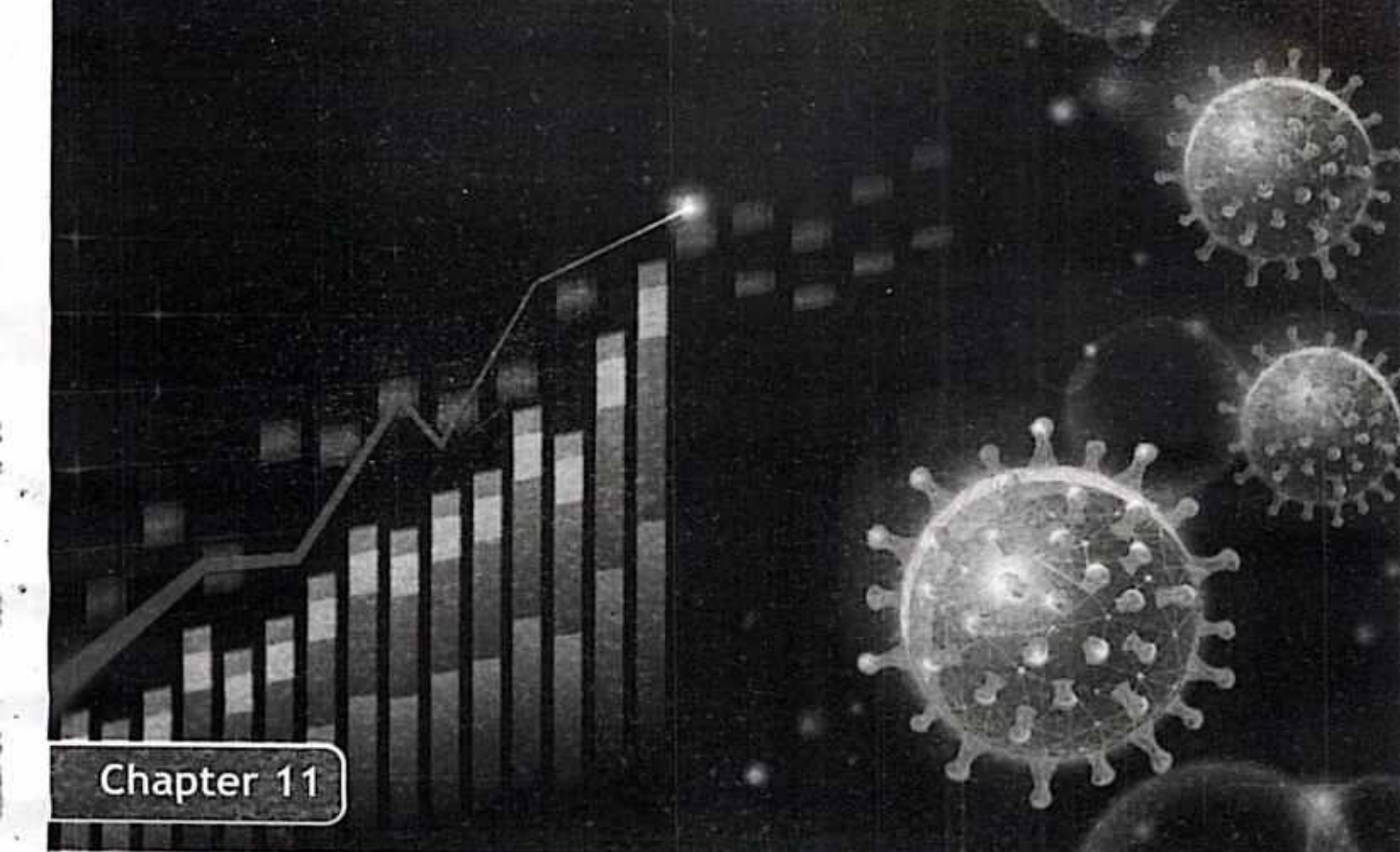




Chapter 11

BIostatISTICS AND DATA HANDLING

SLOs: After completing this lesson, the student will be able to:

1. [B-10-K-01] Define biostatistics and its uses.
2. [B-10-K-02] Define and calculate mean, median and mode,
3. [B-10-K-03] Sketch a bar chart for a given set of biological data.

11.1 BIOSTATISTICS AND ITS USES

11.1.1 Biostatistics

Biostatistics is the application of statistical methods and principles to analyse and interpret the findings of biological, medical and health sciences. Biostatistics mainly deals with the collection, analysis and interpretation of data associated to living organisms and then presentation of findings for possible plans or solutions. Biostatistics is applicable to multiple fields of life sciences particularly health, medicine and agriculture.

Biostatistics help in estimation and calculation of productivity of agriculture, poultry and dairy farming. It is essential part of modern health science. It is used in multiple areas of public health including medicine, epidemiology and research. It plays a crucial role in understanding the disease outbreaks, health trend, possible preventive measure and effectiveness of specific treatment.

Components of Biostatistics studies: In biostatistics studies, following steps or components are followed:

Identification of problem: it is first ever parameter to observe any specific issue affecting the living organisms especially human

Designing experiments: for the solution of any specific issue, one must plan, how to conduct biological experiments

Collecting and analysing data: data collected from the experiments is analyzed thoroughly

Interpreting results: observed reasons and possible future effects are calculated on the basis of data collected from the experiments

Developing new tools: interpretation of experimental results lead to creating new tools or plan for dealing the problem e.g. health related issues

It is highly recommended for every biologist to have at least some basic understanding or learning of biostatistics.

11.1.2 Uses of Biostatistics

1. Agriculture and cattle farming needs for growing population: Biostatistics is involved in the assessment of demands of agriculture and dairy farming according to the rate of population growth. Furthermore it also help the government to assess the need for import or export of food items as per growing population.

2. Medical and Pharmacological Research: Biostatistics helps to design clinical trials and controlled studies to assess the efficacy and safety of newly designed drugs, treatment plan or medical instrument. Findings of medical research guide whether treatment resulted to improvements or side effects. Assessment of the effectiveness and safety of new drugs and deciding optimal doses for any new pharmacological drug.

3. Epidemiological studies and Policy Development: Biostatistics help in monitoring and analysis of data collected from population about the spread of any epidemic disease. Researchers can identify the risk factors, pattern and rate of disease. Epidemiological studies by using biostatic tools help to control and prevent future outbreaks (e.g., prevalence of Hepatitis and Polio, spread of COVID-19 etc.). Biostatistics evidence helps governments and public organizations to make decisions about population management, healthcare strategies and funding sources.

4. Management of Public Health: Each country conduct biostatics studies to estimate health trends within its population. Routine studies include data about birth rates, death rates and prevalence of different disease. It guides the government in planning health initiatives and allocation of resources for public. Biostatistics is useful to monitors and optimize the hospital

performance by calculating patient number, availability of doctors and medicine and effectiveness of treatment.

5. Genetic diseases: Biostatistics is also useful to analyse the data about inheritance patterns of genetic diseases in any population. For example, number of thalassemia and muscular dystrophy patients. It helps in guessing risk factors and behaviour of genetic disorders

6. Pollution indicators and Environmental protection: Biostatistics is used to analyse the pollution level, its causes and impact on population health. Different biostatic tools are designed to monitor the pollution level and potential health risks. It helps in designing policies to reduce health risks from environmental hazards. On the basis of findings, multiple steps are taken to protect the environment. For example, causes of smog, identification of affected areas, possible solution like plantation derives in Punjab especially in Lahore.

7. Survival Analysis: Biostatistics is used to predict and then note the survival rate of patients after a particular treatment. It helps to estimate life expectancy and the chances of success for medical treatments. For example, five years survival after the treatment of cancer is considered a successful treatment.

11.2 DEFINITION AND CALCULATION OF MEAN, MEDIAN AND MODE

In our daily life routine, we commonly hear the somewhat similar to following statements.

- i. Heartbeat of human is 72 beats per minute
- ii. Death rate due to cancer in Pakistan is 25 per thousand patients.
- iii. Production of wheat in Punjab is 2000 Kg per acer.
- iv. The price of the bananas in the market is Rs. 120 per dozen.
- v. Food consumption of a human is 1kg per day.
- vi. The rain fall in Islamabad is 1500mm per year.

If we think on about above mentioned statements, none of them is found exactly correct. In statement no. "i", our heart may not be same 72 in each minute of the day. During running it may be 90 per minute and during sleep it may be 60 per minute. Our heartbeat is approximately 72 in each minute. As given in statement "ii" It is quite possible that in one hospital only 5 cancer patient died out of 1000 visited for treatment and in other hospital 60 patients died out of 1000 patients gone through treatment. The death rate of cancer patients is about 25 per 1000, may not be same all the year and in every hospital. The production of wheat in one district of Punjab may be 1500 Kg from an acer and it may be 2400 Kg from one acer in another district. Although above statements are not exactly true but still they are very important. Actually, these are approximate statements in specific situations. In terms of statistics, we call such statements as average statements. In our daily conversation, we make many statements which have some meaning only on average basis. In different fields of life like agriculture, health, poultry, the idea of average is very important. Many experts at national and international level discuss the findings of studies in averages. The average is also called measure of central tendency.

Calculation of Average: Average is a single value which is calculated to represent the whole data. It may be calculated for a data sample of patients or a population of migrating birds etc. The average is a value which expresses the central idea of the observations. There are different ways to represent average of a data in different situations. For representation of a specific data, proper type of average is used by the expert who is calculating the average.

Types of averages: The following types of averages are commonly used:

- (i) Arithmetic Mean (ii) Median (iii) Mode (iv) Geometric mean (v) Harmonic mean
- Here in this chapter we will study only first three types of averages in detail.

11.2.1 Arithmetic Mean or simply Mean

Definition: Mean is the sum of all the values of data set divided by the total number of values in the data set. It is the single value which is calculated to represent the whole set of data. The symbol “ $\bar{x}$ ” (read as “x bar”) represent the sample mean. The bar above the letter x represents the mean of a set of values:

$$\text{Mean} = \frac{\text{Sum of values}}{\text{Total number of values}}$$

Antibiotics	Source
$\bar{x} = \frac{\sum x}{n}$	$\bar{x} = \frac{\sum fx}{\sum f}$
	(where, $x = \frac{\text{lower limit} + \text{upper limit}}{2}$ and $f = \text{frequency distribution}$)

Example 1(un-group data)

Dataset: Team of world health organization (WHO) planned to assess the prevalence of polio disease in Pakistan. Study to record the polio cases in Pakistan continued for consecutive two years. Number of confirmed Polio patients detected each month of 2022 and 2023 are given in table below. Calculate the mean polio patients per month in each year. Also compare the prevalence of Polio disease in both years.

Sr. no.		Number of Polio patients in year 2022	Number of Polio patients in year 2023
1	January	9	7
2	February	13	9
3	March	15	13
4	April	19	17
5	May	22	18
6	June	25	17
7	July	20	15
8	August	22	14
9	September	18	11
10	October	13	9
11	November	10	8
12	December	6	6
Total	12 months	192	144

$$\text{Mean number of Polio patients per month in 2022} = \frac{\text{Sum of Polio cases in complete year}}{\text{Total number of Months}} = \frac{192}{12} = 16$$

$$\text{Mean number of Polio patients per month in 2023} = \frac{\text{Sum of Polio cases in complete year}}{\text{Total number of Months}} = \frac{144}{12} = 12$$

Difference of Polio patients per month detected in 2022 and 2023 = $16 - 12 = 04$

Prevalence of Polio disease decreased in 2023 compared to 2022 by 04 patients per month.