Studybeesofficial

Where Minds Pollinate

Class 12

Biology 🧬

Chapter 11 Notes

Follow Our Socials

studybeesofficial

studybeesofficialpak

studybeesofficialpak

Chapter: 11 Biostatistics ♥

Biostatistics:
The application of statistical principles and techniques to analyze and interpret research findings in biological, medical and health sciences.

Primary Focus:
Dealing specifically with data associated with living organisms.

Goal:
Presenting findings in an appropriate way to support effective plans or solutions for biological problems.

Data:
Facts, measurements or observations related to living organisms (e.g height, weight, enzyme activity or disease rates).

The Data Set:
A systematic collection of related data values, usually organized in a table.

-Structure:
Rows: Represent a subject or experimental unit.
Columns: Represent a measured variable
Example: A table recording the height and weight of 100 plants under different light conditions.

Components of a Biostatistic Study:

To solve biological problems effectively, researchers follow these five steps:

Identification of Problem:
Identifying issues affecting living organisms, especially humans.

Data Collection and Analysis:
Gathering pre-experiment data and analyzing outcomes thoroughly.

Experimental Design:
Planning specific requirements and durations for biological experiments.

Interpretation of Results:
Using outcomes to estimate reasons for the results and future effects.

Tool Development:
Creating predictive tools and plans to manage health related issues.

Uses of Biostatistics:

Agriculture and Cattle:
Assessing agricultural and dairy farming demands relative to population growth and determining food import/export needs.

Pharmacology:
Designing clinical trials to assess the safety and efficacy of new drugs and medical instruments.

2x10=20

Dosage and Side Effects:
Determining optimal drug doses and identifying potential side effects.

Epidemology:
Monitoring the spread of diseases (e.g covid, polio, hepatitis) to identify risk factors and patterns.

Public Health Management:
Tracking birth/death rate and hospital performance to optimize resource allocation.

Genetics:
Analyzing inheritance patterns of disorders like thalassemia and muscular dystrophy

Environment:
Analyzing pollution levels, their causes (like smog), and designing protective analysing policies.

Survival Analysis:
Predicting life expectancy and treatment success rates (e.g 5-year cancer survival rate)

Average:
A single value calculated to represent the central idea of a whole data set, such as a sample of patients.

Measures of Central Tendancy:
While many types exist (including Geometric and Harmonic mean) the three primary types are:
1) Arithmetic mean
2) Median
3) Mode

Arthimetic Mean:
The sum of all the values in a data set divided by the total number of values.

Denoted by: read as "x-bar" or "x-snake"

Ungrouped Data:
$$\bar{x} = \frac{\sum x}{n}$$
- $\sum x$ - value of the data set (values not organized in categories)
- n - Total no. of values

Grouped Data:
$$\bar{x} = \frac{\sum fx}{\sum f}$$
- x - mid-point (values organized in categories)
- f - Frequency distribution like classes or intervals

To calculate Mid-point (x):
$$x = \frac{\text{lower limit} + \text{upper limit}}{2}$$

Frequency Distribution: A way of organizing and summarizing data in form of table or graph to display the number of occurences (frequencies) of different outcomes or data values in a dataset.

Example: (Ungroup Data)

Sr. no.	Month of sampling	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**

Mean of Polio patients per month in 2022:

$\bar{x} = \frac{\Sigma x}{n} = \frac{192}{12} = 16$ (an average of 16 patients were affected by polio per month)

Mean of Polio patients per month in 2023:

$\bar{x} = \frac{\Sigma x}{n} = \frac{144}{12} = 12$ (an average of 12 patients were affected by polio per month)

Difference of Polio patients per month in 2022 and 2023:

16 - 12 = 04

Prevalence of polio diseased decreased in 2023 compared to 2022 by 04 patients per month

Example:

United Nation..... table given below.

Sr. No.	Mass (Lbs)	Frequency (f)	$X = \frac{\text{Lower limit} - \text{Upper limit}}{2}$	Frequency x mean of limits (fx)
1	35-39	13	$\frac{35+39}{2} - 37$	$13 \times 37 - 481$
2	40-44	15	$\frac{40+44}{2} - 42$	$15 \times 42 = 630$
3	45-49	28	$\frac{45+49}{2} - 47$	$28 \times 47 = 1316$
4	50-54	17	$\frac{50+54}{2} - 52$	$17 \times 52 = 884$
5	55-59	12	$\frac{55+59}{2} - 57$	$12 \times 57 = 684$
6	60-64	10	$\frac{60+64}{2} = 62$	$10 \times 62 = 620$
7	65-69	05	$\frac{65+69}{2} = 67$	$05 \times 67 = 335$
Total		100		4950

To Find Mean:

$$\bar{x} = \frac{\sum fx}{f} = \frac{4950}{100} = 49.5 \text{ Lbs}$$

49.5Lbs is the mean mass of 100 students of different groups before the start of experiment.

Median:

The median is the middle value in a dataset that has been arranged in ascending or descending order of magnitude. It divides the data set into two equal parts: one part with values less than the middle value and the other with values greater than the middle item.

Denoted by: read as "$\tilde{x}$-tilde" or "$\tilde{x}$-snalce" $\tilde{x}$

Median for Ungrouped Data:

For odd Number:

$$\tilde{x} = \left(\frac{n+1}{2}\right)^{th} \text{ value}$$

For Even Number:

$$\tilde{x} = \frac{\left(\frac{n}{2}\right)^{th} \text{ value} + \left(\frac{n}{2}+1\right)^{th} \text{ value}}{2}$$

Activity:
Calculate the median of the heights of 22 plants.
Data: 8, 9, 10, 18, 10, 8, 9, 7, 8, 11, 23, 14, 17, 15, 7, 12, 14, 12, 11, 14, 15, 13

Arranging data:
7, 7, 8, 8, 8, 9, 9, 10, 10, 11, 11, 12, 12, 13, 14, 14, 15, 15, 17, 18, 19, 23

As the data is of Even Number:

$$\tilde{x} = \frac{\left(\frac{n}{2}^{th} \text{ value}\right) + \left(\frac{n}{2}+1\right)^{th} \text{ value}}{2}$$

$$= \frac{(22/2)^{th} + (22/2 + 1)^{th}}{2}$$

$$= \frac{11^{th} + 12^{th}}{2} = \frac{23}{2} = 11.5 \text{ Fruits per second}$$

Median for Grouped Data:

Median for dataset with odd number of values will be $\tilde{x}$ = value of data against $\left(\frac{n+1}{2}\text{th}\right)$ cumulative frequency

Median for dataset with even number of values will be $\tilde{x}$ = Class values against $\frac{\frac{n}{2}\text{th} + \frac{n}{2}+1\text{th}}{2}$ cumulative frequency

Class (no. of flowers/plant)	1	2	3	4	5
Frequency (no. of plants)	3	5	6	4	3

Class (no. of flowers/plant)	Frequency (no. of plants = n)	Cumulative frequency
1	3	3
2	5	8
3	6	14
4	4	18
5	3	21

Median = $\frac{n+1}{2}$ th cumulative frequency

Median = $\frac{21+1}{2} = \frac{22}{2} = 11$

Since cumulative frequency 11 is included in 14 which represents the class value 3, therefore for cumulative frequency 11, the class value will be 3 which is median

Example For Even number data set:

Calculate the median for following data recorded for heights (in cm) of 80.

Class (plant height in cm)	119	120	121	122	123	124	125
Frequency (no. of plants)	4	9	14	18	15	13	7

Class (plant height in cm)	Frequency (no. of plants = n)	Cumulative frequency
119	4	4
120	9	13
121	14	27
122	18	45
123	15	60
124	13	73
125	7	80

$\tilde{x}$ = Class values against $\frac{n}{2}$th + $\frac{n}{2}$ + 1th cumulative frequency / 2

$\tilde{x}$ = $\frac{80}{2}$th + 80/2 + 1th cumulative frequency / 2

$\tilde{x}$ = Class values against cumulative frequencies 40 and 41 / 2

The class values for cumulative frequencies 40 and 41 are included in the class value of cumulative frequency 45 which is 122. Median = $\frac{122 + 122}{2} = 122$

Mode:

The mode is the value in a dataset that appears the maximum number of times.

Formula For Grouped Data:

The formula for calculating the mode for grouped data is:

$$\hat{x} = l + \left[\frac{(f_m - f_1) \times h}{(f_m - f_1) + (f_m - f_2)}\right]$$

Denoted by:

It is read as $\hat{x}$ (x cap or x-hat)

Where:

l = lowercase boundary of the modal group
f_m = Maximum frequency of a group
f_1 = previous frequency of f_m
f_2 = Next frequency of f_m
h = class interval of the modal group

Groups	Frequency (f)	Class Boundary (CB)	Notes
10-19	5	9.5-19.5	
20-29	25 (f_1)	19.5-29.5	
30-39	**40 (f_m)**	**29.5-39.5**	**(Modal Group)**
40-49	20 (f_2)	39.5-49.5	
50-59	10	49.5-59.5	

Solution:

$$\bar{x} = l + \left[\frac{(f_m - f_1) \times h}{(f_m - f_1) + (f_m - f_2)}\right]$$

$l = 29.5$
$f_m = 40$
$f_1 = 25$
$f_2 = 20$
$h = 10$

$$\bar{x} = 29.5 + \frac{(40-25) \times 10}{(40-25) + (40-20)}$$

Mode = 33.786

Standard Deviation:

The extent to which data values in a dataset deviate from the mean.

- It measures how much a set of data varies from its mean value around a normal curve.

Symbols: Represented by the lowercase Greek letter sigma (σ) for a population and the Latin letter (s) for a sampl

Population Formula:

$$\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{N}}$$

X = value in distribution
μ = population mean
N = Total numbers of observation

Sample Formula:

$$s = \sqrt{\frac{\Sigma(X-\bar{x})^2}{n-1}}$$

X = value in distributi
$\bar{x}$ = Sample mean
n = Total no. of observatio

Calculation steps:

→ Mean = 60

→ Differences: Karachi = 30, Lahore = 20, Peshawar = 20, Quetta = -30

→ Squared Differences = Karachi = 900, Lahore = 400, Peshawar = 400, Quetta = 900

→ Sum of Squares = 900 + 400 + 400 + 900 = 2600

→ Divide (Unbiased) = 2600/3 = 866.7

→ Square root = $\sqrt{866.7}$ = 29.45

Range:

The difference blw the highest and lowest value in a set of data:

Calculation:

Range = Highest / Maximum value - Lowest / Minimum value

Cities	Number of confirmed COVID cases	Mean of COVID cases
Karachi	90	
Lahore	80	60
Peshawar	40	
Quetta	30	

Percentile:

A value that indicates the percentage of data in a set that is below a certain value."

Steps to calculate Percentile:

Arrange Data: Sort the data values in ascending order

Calculate Position: Use the formula for the position of the pth percentile.

$$\text{Position} = \frac{P}{100} \times (N + 1)$$

→ P is the desired percentile (e.g 25 for the 25th Percentile)

→ N is the number of observation

Percentile Formula:

Percentile can also be calculated by using the data by formula.

$$P = \frac{n}{N} \times 100\%$$

where:

P is the percentile

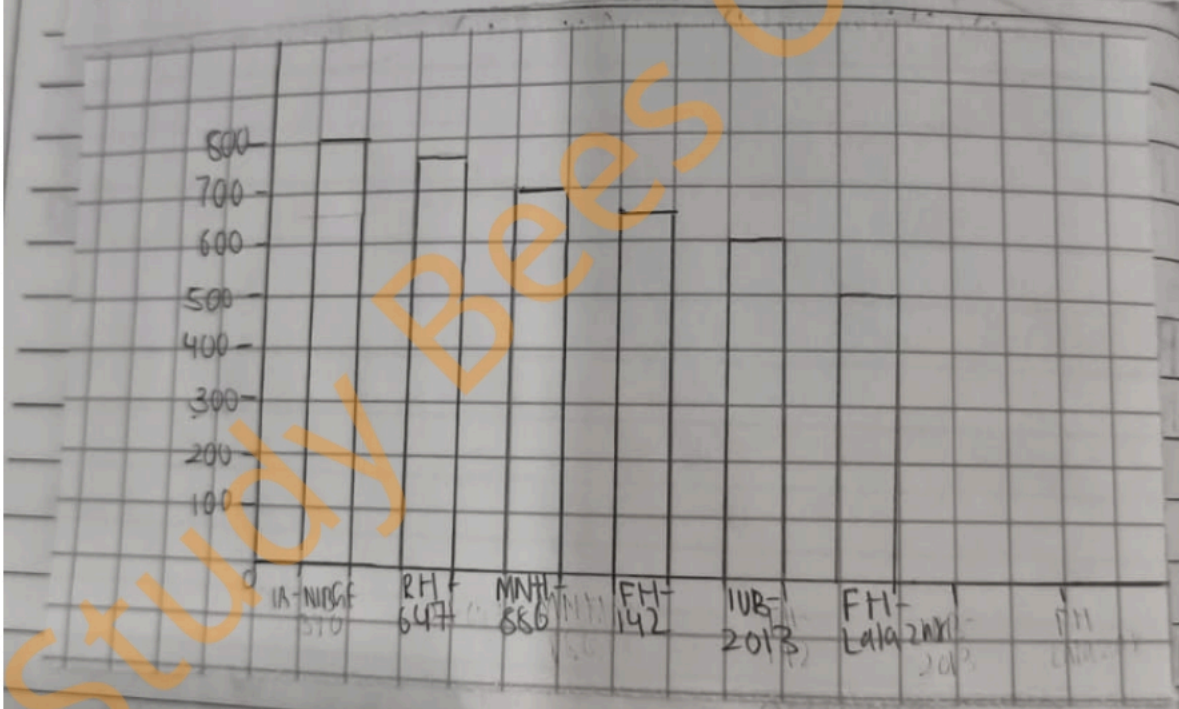
n is the number of data points below the data point of interest

N is the total number of data points in the data set

Sketching or Construction of Bar and Pie Charts:

Example 1: Bt Cotton Production Data (2023)

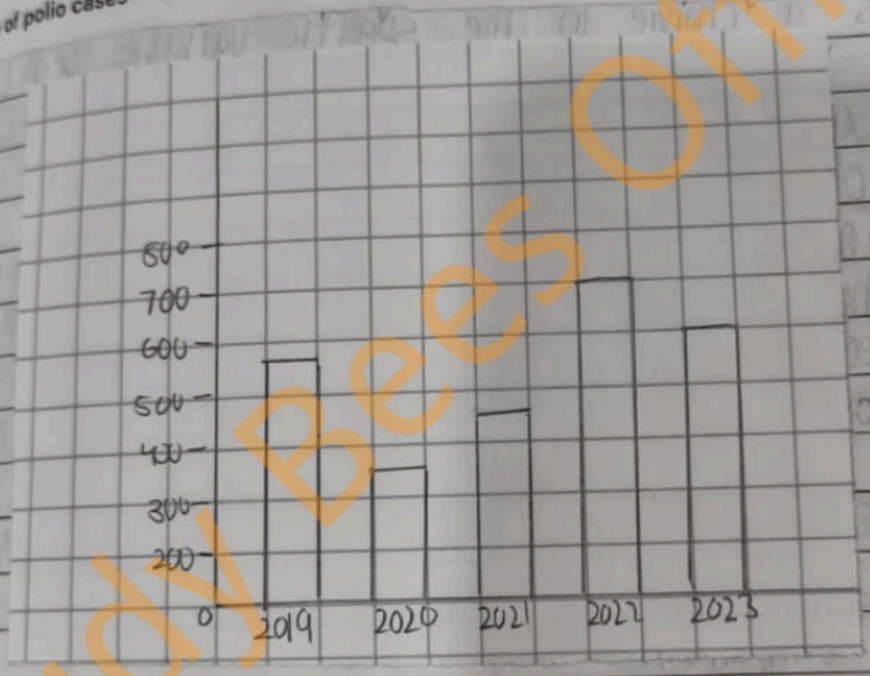
Cotton Varieties	IR-NIBGE-370	RH-647	MNH-886	FH-142	IUB-2013	FH-Lalazar
Production per acre (Kgs)	800	750	700	650	600	500
Production per acre (%age)	20%	18.75%	17.5%	16.25%	15%	12.5%



Example 2: Prevalence of Polio in Pakistan (2019–2023)

This table tracks the number of confirmed polio cases and their yearly percentage over a five-year period.

Years	2019	2020	2021	2022	2023
No. of polio cases	550	350	450	700	600
Percentage of polio cases	20.76%	13.21%	16.98%	26.42%	22.64%



Experimental Design with Control Group and Dependent, Independent and Control Variables:

Scientific experiments are designed using two main types of groups:

Experimental groups and Control groups:

-Experimental group (treatment group):
→experiences a change in the specific variable being tested

Purpose: Scientists apply a specific treatment to observe its effects

Example: In a study testing fertilizer on plant growth, the experimental group consists of plant that recieve the fertilizer.

Control Group:

kept under normal unchanged conditions without the specific treatment applied to the experimental group

Purpose: It serves as a standard or reference for comparisons.

Example: In the fertilizer experiment the control group would be plants grown without any added fertilizer.

There are three primary types of variables used in specific investigations:

1) **Independent Variables:**
Factor that the researcher intentionally changes or manipulates.

2) **Dependent Variables:**
Factor that is measured or observed to determine the outcome of the experiment.

3) **Controlled Variables:**
Factors that must be kept the same for both the control and experimental groups.

Experimental Design:

Experiment 1:
Effect of Light Intensity on Plant Growth:

Objective: Investigate how different light intensities affect the growth of bean plants

Independent Variable: Light intensity (100%, 75%, 50%, 25% and 10%)

Dependent Variable: Plant growth (measured by height in cm and number of leaves)

Control Group: Plants grow under normall (100%) light intensity

Experiment 2: Effect of Water Amounts on Plant Growth

Objective: Investigate how different amounts of watering affect the growth of aloe plants.

Independent Variable: Amount of water per week (1000ml, 800ml, 600ml, 400ml, and 200ml)

Dependent Variable: Plant growth (measured by height and number of leaves)

Control Group: Plants receiving 1000ml of water per week

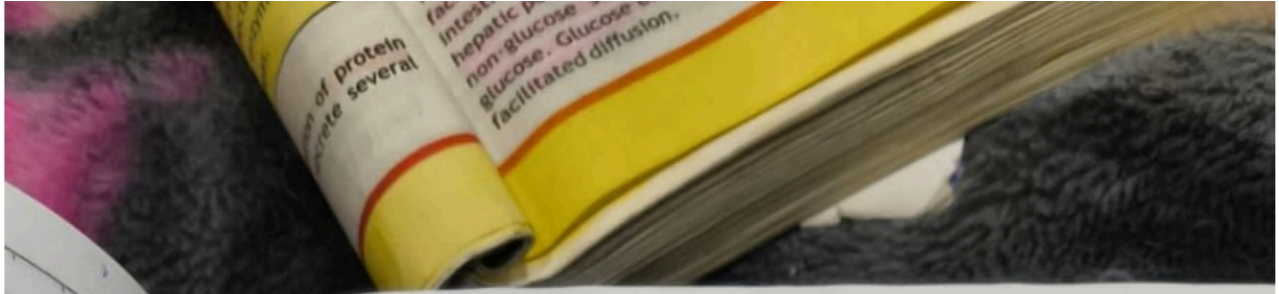
Experiment 3: Effect of pH levels on Enzyme Activity:

Objective: Examine how pH levels affect the activity of the enzyme catalase in breaking down hydrogen peroxide

Independent Variable: pH level (e.g. pH 4, pH 7, and pH 10).

Dependent Variable: Enzyme activity (measured by gas production or reaction time).

Control Group: Reaction occurring at a neutral pH (pH 7).



Exercise Questions

(Q12)

$$S.D = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

$$\bar{x} = \frac{968}{15} = 64.5$$

x	$x - \bar{x}$	$(x-\bar{x})^2$
33	-31.5	992.25
55	-9.5	90.25
66	1.5	2.25
84	19.5	380.25
91	26.5	702.25
72	7.5	56.25
68	3.5	12.25
87	22.5	506.25
78	13.5	182.25
66	1.5	2.25
82	17.5	306.25
46	-18.5	342.25
71	6.5	42.25
31	-33.5	1122.25
38	-26.5	702.25
$\Sigma = 968$		$\Sigma = 5441.75$

$$S.D = \sqrt{\frac{5441.75}{14}}$$

$$S.D = \sqrt{388.7}$$

$$\boxed{S.D = 19.72}$$

(013)

Median:

31, 33, 38, 46, 55, 66, 66, (68), 71, 72, 78, 84, 87, 91

Median

(x - x)

x - x