F010104T-Business Statistics BBA Sem1 (Unit I)

Amit U. Dwivedi

August 2025

Lecture Notes - Unit I (Introduction to Statistics)

Lecture:1

Learning Objectives: Warmups, notion and notations

1. **What is statistics?**
   - Descriptive statistics, inferential statistics.
   - Distinguish between a sample and a population.
2. **Understand how data are collected.**

0.1 Statistics

Definition: *It is the art and science of learning from the data.* The minute we say learning from the data, it is concerned with..

- *Collecting the data*
- *Organizing the data*
- *Representing the data*
- *Interpreting the data*
- *Summarizing the data*
- *Analyzing the data*
- *How to draw conclusions from the data and making decisions.*

0.2 Major branches of statistics

1. **Description**

   Definition: *The part of statistics concerned with the description and summarization of data is called* ***descriptive statistics***

2. **Inference**

   Definition: *The part of statistics concerned with the drawing of conclusions from data is called* ***inferential statistics****.*

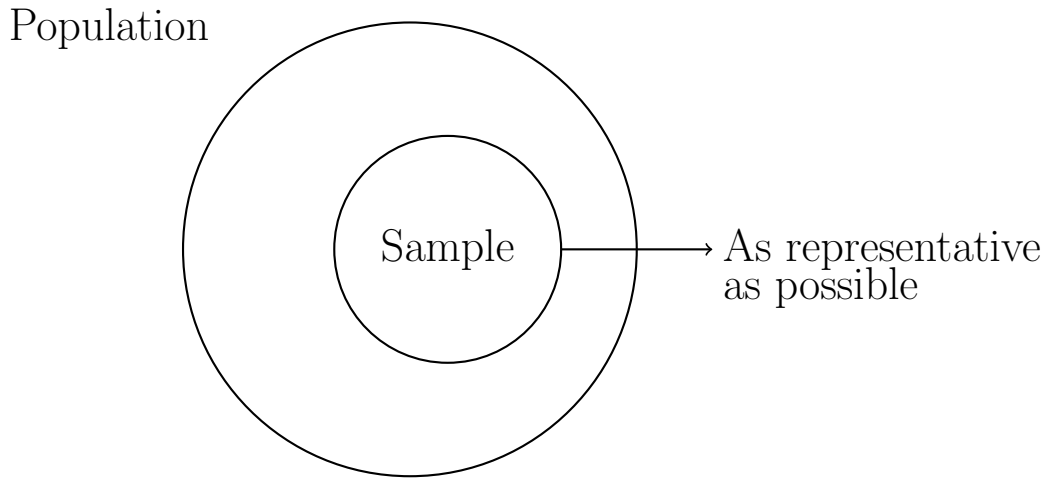
To draw a conclusion from the data, we must consider the possibility of chance, which introduces the concept of probability.

The minute we define the inferential statistics, we need to define Population and Sample

0.3 Population & Sample

Suppose we are interested in knowing

- The percentage of all students in India who have passed their Class 12 exams and study engineering.
- The prices of all houses in Tamil Nadu.
- The total sales of all cars in India in the year 2019.
- The age distribution of people who visit a city mall in a particular month.



Definition

The total collection of all the elements that we are interested in is called a **population**.

Definition

A subgroup of the population that will be studied in detail is called a **sample**.

0.4 Purpose of Statistical Analysis

- If the purpose of the analysis[1] is to examine and explore information for its own intrinsic interest only, the study is **descriptive**.
- If the information is obtained from a sample of a population and the purpose of the study is to use that information to draw conclusions about the population, the study is **inferential**.
- A descriptive study may be performed either on a sample or on a population.
- When an inference is made about the population, based on information obtained from the sample, does the study become **inferential**.

In the following data set in Tbale 1: Cricket Player's Statistics, if we ask following questions

- What is the **average (mean) number of runs** scored by batsmen in the dataset?
- What is the **median batting average** of all players?

(Measures of Central Tendency)

- What is the **range** of batting averages in the dataset?
- How **consistent** are bowlers in terms of their bowling average? (Standard Deviation)

(Measures of Dispersion)

- What percentage of batsmen have a **batting average above 40**?
- How many bowlers have taken **more than 100 wickets**?

(Frequency Distribution & Percentiles)

- How does the **batting average** vary across different types of players (batsmen, all-rounders, wicketkeepers)?
- What is the **distribution of runs scored** among all players?

(Data Visualization & Summary)

- Who are the **top 5 batsmen** based on total runs scored?
- Who are the **top 3 bowlers** based on bowling average?

(Ranking & Comparison)

If these are the question of Interest then all these question of ineterest can directly be answered from the data by describing and summarizing the data set. Hence the study is ***descriptive.***

[1] Weiss, Neil A. *Introductory Statistics: Pearson New International Edition.* Pearson Education Limited, 2014.

S.No	Player Name	Jersey No	Matches Played	Role	Runs	Batting Avg	Highest Score	Wickets	Bowling Avg	Best Bowling
1	Sachin Tendulkar	10	463	Batsman	18426	44.83	200	154	44.48	5/32
2	Virat Kohli	18	275	Batsman	12700	58.85	183	5	162.25	1/12
3	MS Dhoni	7	350	Batsman, WK	10773	50.58	183	1	31	1/14
4	Rohit Sharma	45	270	Batsman	11420	50.27	267	9	64.38	2/27
5	Sehwag	46	251	Batsman	8273	35.04	219	96	40.14	4/6
6	Gautam Gambhir	5	147	Batsman	5238	39.68	150	0	0	0/13
7	Yuvraj Singh	12	304	All-rounder	8701	36.56	150	111	45.48	5/31
8	R Jadeja	8	185	All-rounder	2425	33.55	100	230	37.12	6/25
9	Zaheer Khan	34	200	Bowler	792	12	34	282	29.4	5/42
10	Harbhajan Singh	3	236	Bowler	1237	13.3	49	269	33.36	5/31
11	Jasprit Bumrah	93	90	Bowler	56	4.2	11	124	22.43	5/25
12	Mohammed Shami	11	100	Bowler	234	9.3	29	150	26.21	5/55
13	R Ashwin	99	120	Bowler	875	17.45	74	170	38.98	5/26
14	Kuldeep Yadav	23	78	Bowler	150	14.31	22	130	27.56	6/25
15	Y Chahal	6	55	Bowler	15	3.5	7	65	24.35	6/27
16	Hardik Pandya	33	75	All-rounder	1583	35.41	91	74	38.47	4/25
17	Kedar Jadhav	81	95	All-rounder	1525	41.09	128	36	29.68	3/23
18	KL Rahul	1	95	Batsman, WK	2701	45.87	121	0	–	–
19	Shikhar Dhawan	25	145	Batsman	6103	40.22	143	0	–	–
20	Ambati Rayudu	5	55	Batsman	1694	47.05	124	3	41.33	1/5
21	Robin Uthappa	6	46	Batsman	934	25.94	86	0	–	–
22	Dinesh Karthik	21	94	Batsman, WK	1752	30.2	79	0	–	–
23	Suresh Raina	48	226	All-rounder	5615	35.31	116	36	45.2	3/34
24	Ishant Sharma	97	105	Bowler	92	4.8	17	140	31.5	5/29
25	Bhuvneshwar Kumar	15	125	Bowler	575	10.12	55	150	28.4	5/22
26	Rishabh Pant	17	45	Batsman, WK	1521	42.8	128	0	–	–
27	Prithvi Shaw	100	20	Batsman	567	38.9	99	0	–	–
28	Shubman Gill	77	50	Batsman	2150	47.7	120	0	–	–
29	Shivam Dube	70	18	All-rounder	402	29.4	88	19	38.4	2/22

Table 1: Cricket Player's Statistics

If the questions of interest are as follows.

- Does the batting average of bowlers significantly differ from that of batsmen?
- Is there a significant difference in the number of matches played by all-rounders and batsmen?

(Hypothesis Testing Questions)

- What is the estimated average batting average of all players based on this dataset?
- What is the likely range of total wickets taken by bowlers in international matches based on this sample?

(Estimation & Confidence Interval Questions).

- Can we predict a player's highest score based on their total runs?
- How does the number of matches played affect a player's batting average?

(Regression & Prediction Questions)

- Is there a significant relationship between wickets taken and bowling average?
- Does a player's role (batsman, bowler, all-rounder) correlate with their highest score?

(Correlation & Association Questions)

- Do batsmen and all-rounders have significantly different bowling averages?
- Is the number of centuries significantly different between batsmen and all-rounders?

(Population Comparison Questions)

then all these questions of interest can be answered not only summarizing the data but further more techniques are used. Hence the study is called as ***inferential statistics.***

Lecture1 Summary

1. **Descriptive Statistics** – Summarizes and presents data, providing insights into its key characteristics. It can be applied to both **samples and populations** without making predictions.
2. **Inferential Statistics** – Uses **sample data** to make generalizations or draw conclusions about a **larger population**, helping in decision-making and predictions.

Lecture2

Learning Objectives: Data

1. Understanding Data
2. Data Collection
3. Structured v/s Unstructured data
4. Variable and Cases
5. Types of data
    - Catagorical v/s Numerical
    - Cross Sectional v/s Time Series
6. Measurement Scales
7. Comparison of Data and Measurement Scales

1 Understating the data

Centuries ago, data was primarily classified into just a few types, such as categorical (qualitative) and numerical (quantitative) data. However, in today's world, discussions extend to social media analytics, multimedia analytics, and even a simple comment on social media or a photo is considered data. The definition of data evolves over time, adapting to technological advancements and societal changes. In our daily lives, we constantly contribute to data generation—whether by clicking a keyboard or mouse, commenting on a photo, or posting on social media.
Statistics relies on data, drawing insights from the vast information surrounding us

1.1 Definition

Data *refers to the collection of facts and figures that are gathered, analyzed, and summarized for presentation and interpretation.*

1.2 Why Do We Collect Data?

We collect data to understand the characteristics of a specific group of people, places, things, or events.
Example: To analyze the temperature in Prayagraj during a particular month.
Example: To examine the marks obtained by students in their Class 12 examinations.
Exampe: To know about how many people like new song/ product/ video- collected through comments.

1.3 Where do we get the data? (Data Collection)

- **Data Available: Published Data**

  Published data refers to information that has already been collected, processed, and made publicly available by government agencies, research organizations, or institutions. These datasets are useful for research, analysis, and decision-making.

  Some well-known sources of published data include:

  data.gov.in – The open government data platform of India, offering datasets from various government departments, including demographics, health, education, and economy.

  data.gov – The U.S. government's open data portal, providing datasets across various sectors such as agriculture, business, climate, and public safety.

  data.europa.eu – The European Union's open data platform, hosting datasets from EU institutions and member states.

  World Bank Open Data – A global repository of economic and development data.

  UNData – A platform that provides statistical data from the United Nations on global issues like population, trade, and development.

  Kaggle Datasets – A collection of datasets used for machine learning, statistical modeling, and research.
  These type of Data are called as Secondary Data

- **Data Not Available: need to collect or generate the data.**

  When the needed data is not available from published sources, it has to be collected or created in different ways, such as:

  Surveys and Questionnaires – Asking people or organizations for information.
  Experiments and Observations – Watching and recording events in a controlled or natural setting.
  IoT and Sensors – Using devices to collect real-time data.
  Web Scraping – Gathering data from websites where allowed.

  Collecting data properly helps ensure it is accurate, reliable, and used ethically.
  These type of data are called as Primary Data

1.4 Structured and Unstructured Data

- For information of data to be useful, we must know the context of the numbers and text it holds.

- When they are scattered around with no structure, the information is of very little use.

- Hence, we need to organize the data.

We are going to understand what is the meaning of structured collection of data? When we need structured collection of data to form a data set, we first need to understand what are variables and cases.

1.5 Variable and Cases

Names	Fees paid	Marks Obtained	Board
Anjali	30,000	75	CBSE
Aparna	30,000	87	ICSE
Rohan	30,000	92	CBSE
Neha	30,000	81	State Board
Saurabh	30,000	78	State Board
Priya	30,000	85	CBSE
Varun	30,000	88	ICSE
Sneha	30,000	90	ICSE
Aditya	30,000	83	State Board

Table 2: Student Data

Case(Observation)

A unit from which the data are collected.

Variable

Intuitive: A variable is that *"Varies"*
Formally: A characteristic or attribute that varies across all the units.

Example: In above Student data set in table.2

- Case - Each student.
- Variable- Name, Marks Obtained, Board etc
- Row represents Cases, foe each case same attribute is recorded.
- Columns represents variables, For each variables, same type of value for each case is recorded.
- Each variable must have its own column.
- Each observation(cases) must have its own row.
- Same type of attribute have same unit.

Types of Data

1.6 Categorical v/s Numerical

Data can be broadly categorized into two main types:

Categorical (Qualitative) Data

- Represents characteristics or labels.
- Cannot be measured numerically but can be classified into groups.

Examples:

- Colors (Red, Blue, Green)
- Gender (Male, Female)
- Types of cuisine (Indian, Italian, Chinese)

Subtypes of Categorical Data:

- **Nominal Data:** No inherent order (e.g., Blood Type: A, B, O, AB).
- **Ordinal Data:** Ordered but differences between values are not meaningful (e.g., Education Level: Primary, Secondary, Higher Secondary).

Numerical (Quantitative) Data

- Represents measurable quantities.
- Can be counted or measured numerically.

Examples:

- Height of students (in cm)
- Marks obtained in an exam
- Temperature readings

Subtypes of Numerical Data:

- **Discrete Data:** Can take only certain fixed values (e.g., Number of students in a class: 30, 31, 32).
- **Continuous Data:** Can take any value within a given range (e.g., Temperature: 36.5°C, 37.2°C).

1.7 Cross-Sectional vs. Time-Series Data

- **Cross-Sectional Data:** Data collected at a single point in time.
  - Example: Population data of different states in a particular year.
- **Time-Series Data:** Data collected over time at regular intervals.
  - Example: Stock prices recorded daily for a month.

1.8 Measurement Scales

Measurement scales define how data values can be interpreted mathematically.

Nominal Scale (Categorical)

- Represents categories with no inherent order.
- **Example:** Types of vehicles (Car, Bike, Bus).

Ordinal Scale (Categorical)

- Represents categories with a meaningful order, but the differences between them are not measurable.
- **Example:** Performance Rating (Excellent, Good, Average, Poor).

Interval Scale (Numerical)

- Represents ordered values where the difference is meaningful, but there is no absolute zero.
- **Example:** Temperature in Celsius or Fahrenheit (0°C is not the absence of temperature).

Ratio Scale (Numerical)

- Represents values with an absolute zero, allowing meaningful ratio comparisons.
- **Example:** Height, Weight, Income (0 kg means no weight).

1.9 Comparison of Types of Data

Basis of Comparison	**Categorical Data**	**Numerical Data**
Nature	Labels, categories	Measurable numbers
Subtypes	Nominal, Ordinal	Discrete, Continuous
Example	Gender, Blood Type	Marks, Height

Table 3: Comparison of Categorical and Numerical Data

1.10 Comparison of Measurement Scales

Basis of Comparison	Nominal Scale	Ordinal Scale	Interval Scale	Ratio Sca
Type of Data	Categorical	Categorical	Numerical	Numerical
Order Exists?	No	Yes	Yes	Yes
Meaningful Difference?	No	No	Yes	Yes
True Zero Exists?	No	No	No	Yes
Example	Eye Color	Movie Ratings	Temperature (°C, °F)	Height, Age, W

Table 4: Comparison of Measurement Scales

Lecture2 Summary

- **Understanding Data**
  - Data has evolved from simple categories (qualitative, quantitative) to complex forms like social media and multimedia analytics.
  - It is collected, analyzed, and used for decision-making in various fields.
- **Data Collection**
  - **Published Data:** Available from sources like government databases, research institutions, and online platforms.
  - **Collected Data:** Obtained through surveys, experiments, observations, IoT devices, and web scraping.
- **Structured vs. Unstructured Data**
  - Structured data is organized in datasets with variables and cases.
  - Variables (attributes) vary across cases (individual observations).
- **Types of Data**
  - **Categorical (Qualitative):** Labels or groups (Nominal, Ordinal).
  - **Numerical (Quantitative):** Measurable values (Discrete, Continuous).
- **Cross-Sectional vs. Time-Series Data**
  - **Cross-Sectional:** Snapshot at a single time point (e.g., population in 2024).
  - **Time-Series:** Data recorded over time (e.g., daily stock prices).
- **Measurement Scales**
  - **Nominal:** Categories without order (e.g., Blood Type).
  - **Ordinal:** Ordered categories (e.g., Education Level).
  - **Interval:** Ordered, measurable differences but no true zero (e.g., Temperature).

  - **Ratio:** Ordered, measurable, and has a true zero (e.g., Height, Weight).

- **Comparison of Data & Measurement Scales**
  - Categorical vs. Numerical data differ in structure and usability.
  - Measurement scales define how data can be interpreted and analyzed.

Lecture 3: Data Classification

1. Definition

Data classification is the process of arranging data into different categories or groups based on their common characteristics or attributes. It makes the data organized, systematic, and easy to analyze for statistical purposes.

Example: Suppose we collect data of students in a college — names, marks, hometowns, and age. We can classify them by location, marks range, gender, or year of admission to make it more meaningful.

2. Basis of Classification

Classification can be done on different bases depending on the nature of the data. The main types are:

a) Geographical Classification

Meaning: Data is classified based on geographical or location-wise differences such as country, state, city, region, district, etc.

Purpose: Useful in comparing data between different areas.

Examples:

- Population of India by states.
- Average rainfall in different districts of Uttar Pradesh.
- Literacy rate by country.

b) Chronological Classification

Meaning: Data is classified according to time (chronological order).

Purpose: Useful for studying trends and changes over time.

Examples:

- Annual production of wheat from 2015 to 2025.
- Monthly sales of a company.
- Year-wise GDP growth.

c) Qualitative Classification

Meaning: Data is classified based on qualities or attributes that cannot be measured in numerical terms but can be described.

Attributes: Gender, religion, marital status, educational qualification, etc.

Examples:

- Students classified by gender: male, female, others.
- Classification of employees by education level: graduate, postgraduate, diploma holder.

d) Quantitative Classification

Meaning: Data is classified based on measurable numerical values.

Purpose: Helpful for statistical analysis like averages, percentages, etc.

Examples:

- Classification of employees based on salary ranges.
- Students grouped according to marks obtained.
- Cities classified by population size.

Note: Once classified quantitatively, statistical measures like Mean, Median, and Mode can be calculated. For example:

$$\text{Arithmetic Mean: } \bar{X} = \frac{\sum X}{N}$$

$$\text{Median (for ordered data): } M = \left(\frac{n+1}{2}\right)^{th} \text{ observation}$$

$$\text{Mode (for grouped data): } Z = L + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$$

Where:

- L = lower boundary of the modal class
- f_1 = frequency of the modal class
- f_0 = frequency of class before modal class
- f_2 = frequency of class after modal class
- h = class interval size

Lecture 3 Summary Table

ccc

Type of Classification	Basis	Example
Geographical	Place / Location	Literacy rate by states
Chronological	Time	GDP growth from 2010–2020
Qualitative	Attributes / Qualities	Students by gender
Quantitative	Measurable numerical data	Employees by salary ranges

2 Lecture 4: Frequency Distributions

A frequency distribution is a way to organize and present data to show how often each value (or range of values) occurs in a dataset. It helps in summarizing large data sets and makes it easier to identify patterns, trends, and outliers.

2.1 Types of Frequency Distributions

1. **Ungrouped Frequency Distribution:** It lists each individual data point along with its frequency (number of occurrences). This is useful for small datasets where each value is distinct and meaningful.

2. **Grouped Frequency Distribution:** It organizes data into class intervals (or bins) instead of listing each individual value. This is helpful when dealing with large datasets.

3. **Cumulative Frequency Distribution:** It shows the total number of observations that fall below or up to a certain value. There are two types:

   - *Less than cumulative frequency:* Sum of frequencies up to a given class interval.
   - *More than cumulative frequency:* Sum of frequencies from a given class interval onward.

4. **Relative Frequency Distribution:** Represents the frequency of each class as a fraction or percentage of the total observations.

$$\text{Relative Frequency} = \frac{\text{Frequency}}{\text{Total Observations}}$$

5. **Percentage Frequency Distribution:** Converts relative frequency into percentage form by multiplying by 100.

$$\text{Percentage Frequency} = \left(\frac{\text{Frequency}}{\text{Total Observations}} \right) \times 100$$

2.2 Converstion of Raw Data into frequency distributions

Given a set of test scores from 30 students:

$$4, 6, 7, 5, 8, 10, 12, 14, 7, 9, 10, 15, 18, 20, 6, 13, 11, 19, 16, 10, 9, 7, 8, 12, 14, 13, 17, 16, 11, 5$$

Ungrouped Frequency Distribution

Search the smallest value of data set and write its frequency(no. of times it occurs in the data set), the find the next smallest value and do the same, repeat the process until all the data covered.

Score (X)	Frequency (f)
4	1
5	2
6	2
7	3
8	2
9	2
10	3
11	2
12	2
13	2
14	2
15	1
16	2
17	1
18	1
19	1
20	1

Table 5: Ungrouped Frequency Distribution

2.3 Grouped Frequency Distribution

Step1: Find the Smallest value (S=4) and Largest Value(L=20) in the data set. Then find Range(R)=L-S=20-4=16.
Step2: Find the Total number of obervation(N=30) to decide the number of classes(k) using the formula smallest k such that $2^k \geq N$. here we can take k=5 since $2^5 = 32 \geq 30 = N$
Step3: FInd width of the class (h) using the formula, $h = \frac{Range}{\text{Number of classes}} = \frac{R}{h}$
Here $h = \frac{16}{5} = 3.2 \approx 3$
classes will start from the smallest value and having width h=3, i.e. 4-7, 7-10, 10-13, 13-17, 17-20. We shall be taking the frequency as the number of element greater than or equal to lower limit of the class and less than upper limit.
Hence Grouped frequency distribution is as follows

Class Interval	Frequency (f)
4 - 7	5
7 - 10	7
10 - 13	7
13 - 16	5
16 and above	6

Table 6: Grouped Frequency Distribution

2.4 Cumulative Frequency Distribution(Less than Type)

It counts the total number of observations which are less than or equals to lower limit of the class, or less than or equals to perticular value of the data.

Class Interval	Frequency (f)	Cumulative Frequency(Less than type)
4 - 7	5	5
7 - 10	7	12
10 - 13	7	19
13 - 16	5	24
16 and above	6	30

Table 7: Cumulative Frequency Distribution

2.5 Relative Frequency Distribution & Percentage Relative Distribution

Relative frequency is calculated as:

$$\text{Relative Frequency} = \frac{\text{Frequency}}{\text{Total Observations}}$$

Since total observations = 30, we get:

Class Interval	Frequency (f)	Relative Frequency	Percentage Relative Frequency (%)
4 - 7	8	0.2667	26.67%
8 - 11	9	0.3000	30%
12 - 15	7	0.2333	23.33%
16 - 20	6	0.2000	20%

Table 8: Relative Frequency Distribution

Example

Convert the following in Continuous Frequency Distributions
23, 45, 12, 67, 89, 34, 56, 78, 90, 21,
43, 65, 87, 32, 54, 76, 98, 20, 42, 64,
86, 31, 53, 75, 97, 19, 41, 63, 85, 30,
52, 74, 96, 18, 40, 62, 84, 29, 51, 73,
95, 17, 39, 61, 83, 28, 50, 72, 94, 16

Step 1: Identify Smallest and Largest Values
Given data:
23, 45, 12, 67, 89, 34, 56, 78, 90, 21,
43, 65, 87, 32, 54, 76, 98, 20, 42, 64,
86, 31, 53, 75, 97, 19, 41, 63, 85, 30,
52, 74, 96, 18, 40, 62, 84, 29, 51, 73,
95, 17, 39, 61, 83, 28, 50, 72, 94, 16

- Smallest value $S = 4$
- Largest value $L = 20$

- Range $R = L - S = 20 - 4 = 16$

Step 2: Determine the Number of Classes

- Total number of observations $N = 30$
- We find the smallest k such that $2^k \geq N$.

$$2^5 = 32 \geq 30$$

- Thus, we take $k = 5$ (5 classes).

Step 3: Find the Class Width

$$h = \frac{\text{Range}}{\text{Number of classes}} = \frac{16}{5} = 3.2 \approx 3$$

Each class interval will have a width of $h = 3$.

Step 4: Create Class Intervals

The class intervals are as follows:

- $4 - 7$
- $7 - 10$
- $10 - 13$
- $13 - 16$
- 16 and above

Step 5: Count the frequency

We count the number of elements within each class range:

- $4 - 7$: 4 values
- $7 - 10$: 5 values
- $10 - 13$: 5 values
- $13 - 16$: 8 values
- 16 and above: 8 values

Step 6: Construct the Frequency Table

Class Interval	**Frequency (f)**
4 - 7	4
7 - 10	5
10 - 13	5
13 - 16	8
16 and above	8
Total (N)	**30**

Lecture 5: Bi-variate Frequency

1. Definition

A **Bivariate Frequency Distribution** is a statistical table that shows the frequency of occurrence of data corresponding to **two variables** taken together.

- It is an extension of **univariate frequency distribution** (one variable).
- Here we study the relationship between two characteristics simultaneously.
- Such data is often represented in the form of a **two-way table (contingency table)**.

Mathematically, bivariate data can be represented as:

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

2. Key Terms

1. **Variables:** Two attributes or characteristics under study. Example: Marks in Mathematics (X) and Marks in Statistics (Y).
2. **Class Intervals:** Each variable is divided into class intervals. Example: $0-10, 10-20, 20-30, \ldots$
3. **Cell Frequency:** The number of observations falling in a particular pair of class intervals.
4. **Marginal Frequencies:**
   - Row totals (distribution of X).
   - Column totals (distribution of Y).
5. **Grand Total (N):** The total number of observations, shown at the bottom-right corner.

3. Steps in Construction

1. Decide the two variables to be studied together. Example: Age of employees and Monthly Salary.
2. Determine the class intervals for each variable.
   - Variable X: Age groups $\rightarrow 20-30, 30-40, 40-50, \ldots$
   - Variable Y: Salary groups $\rightarrow 10,000-15,000, 15,000-20,000, \ldots$
3. Prepare a **two-way table** with:
   - Rows $\rightarrow$ class intervals of X.

- Columns → class intervals of Y.

4. Tally the observations into the appropriate cells.
5. Write down the cell frequencies, then calculate row totals, column totals, and finally the grand total.

4. Example

Suppose we collect data on **Marks in Mathematics** (X) and **Marks in Statistics** (Y) for 50 students.

Math (X) \ Stats (Y)	0–10	10–20	20–30	30–40	Row Total
0–10	2	3	1	0	6
10–20	1	4	5	2	12
20–30	0	3	8	4	15
30–40	0	1	5	11	17
Column Total	3	11	19	17	**50**

Explanation:

- Each **cell frequency** shows the number of students falling in that pair of intervals.
- Example: "8" means 8 students scored 20–30 in Math and 20–30 in Statistics.
- Row totals and column totals give the **marginal distributions**.
- Grand total $N = 50$.

5. Uses

- Helps to study the **relationship between two variables**.
- Forms the basis for:
  - Correlation analysis
  - Regression analysis
  - Chi-square test of independence