

# STATISTICS



# Meaning and scope of Statistics

Statistics can be defined as the collection, classification, presentation analysis and interpretation of data

- statistics in planning-planning for economic development
- Statistics in economics-to solve economic problems like demand analysis, price etc.
- Statistics is also applied in fields like business, industry, mathematics, modern science, psychology and education as well

# Population and Sample

- Population- the population in statistics includes all members of a defined group, that we are studying or collecting information on for data driven decisions
- Sample-a part of the population is called sample. It is a proportion of the population, a part of it with all its characteristics

# Census and Sampling

- Census- information collected from every member of population. The collection of data from each and every unit of the population.
- Sampling- information collected from just representative part of a population.

# Advantages of sampling over census

## Census

- Expensive
- Time consuming
- If population is too large census is not applicable
- It is not possible for destructive type of experiments

## Sampling

- Less expensive
- Less time consuming
- Applicable for large population too
- Applicable for destructive experiments also
- Less error occurrence
- Easy error analysis

# What is collection of data?

- The process by which the researcher collects the information needed to answer the research problem
- The process of collecting data from surveys or from independent or networked locations through data capture ,data entry, or data logging

# PRIMARY DATA

- Data collected by the investigator himself/herself for a specific purpose (or data collected by the investigator for his own purpose)
- example-:  
Data collected by a student for his/her research project

# SECONDARY DATA

- Data collected by someone else for some other purpose. (the investigator uses the data collected by someone else)
- Example:-
  1. Government publications like census reports
  2. Research journals
  3. Publications of institutions like banks, companies. Etc.



# Questionnaire

- A questionnaire is a list of questions used for the collection of information in an investigation.
- Questionnaire are necessary for both census and sample studies.
- A questionnaire is a research instrument consisting of a series of questions and other prompts for the purpose of gathering information from respondents.

# Questionnaire

- List of a research or survey questions asked to respondents and designed to extract information.
- It serves 4 basic purposes:
  - 1) To collect the appropriate data
  - 2) To make data comparable and amenable to analysis
  - 3) To minimize bias in formulating and asking question
  - 4) To make questions engaging and varied

# Types of data

- Data is a collection of information from which conclusions may be drawn
- There are 2 types of data
  - i. Quantitative data
  - ii. Qualitative data

# Quantitative

- Quantitative data is information about quantities that is information that can be measured and written down with numbers
- Example:-
  - a. scores on achievement test
  - b. Number of hours of study

# Qualitative

- Qualitative data is information about qualities information that can't be actually be measured

- example-:-

Softness of your skin

Color of the sky

# CLASSIFICATION

➤ IT IS THE ARRANGEMENT OF DATA.

➤ FOUR TYPES :

1. GEOGRAPHICAL

2. CHRONOLOGICAL

3. QUANTITATIVE

4. QUALITATIVE

# TABULATION

- ❖ Arrange data in the form of rows and columns.
- ❖ Heading.
- ❖ Finite number of rows and column.
- ❖ columns heading and rows heading.
- ❖ Scale measurement.
- ❖ Foot step (if needed).

# FREQUENCY DISTRIBUTION

A representation, either in a graphical or tabular format, which displays the number of observations within a given interval.

The intervals must be mutually exclusive and exhaustive.

Frequency distributions are usually used within a statistical context.



## Relative Frequency Histogram



# DISCRETE

- **This data set shows a group of discrete data.**
- **Discrete data can only take particular values.**
- **Discrete data can be numeric -- like numbers of apples -- but it can also be categorical -- like red or blue, or male or female, or good or bad.**
- **This is called discrete data because the units of measurement (for example, CDs) cannot be split up; there is nothing between 1 CD and 2 CDs.**

| <b>Music format</b> | <b>Number sold</b> |
|---------------------|--------------------|
| <b>CD albums</b>    | <b>140</b>         |
| <b>CD singles</b>   | <b>70</b>          |
| <b>Downloads</b>    | <b>55</b>          |
| <b>Vinyl</b>        | <b>5</b>           |
| <b>Total sales</b>  | <b>270</b>         |

# CONTINUOUS

**The data set shows a group of continuous data.**

**This data is called continuous because the scale of measurement - distance - has meaning at all points between the numbers given, eg :we can travel a distance of 1.2 and 1.85 and even 1.632 miles.**

**Continuous data are not restricted to defined separate values, but can occupy any value over a continuous range. Between any two continuous data values there may be an infinite number of others. Continuous data are always essentially numeric.**

**Continuous data can be shown on a number line, and all points on the line have meaning and are different, but with discrete data only certain values have meaning.**

**Length of journey to work**

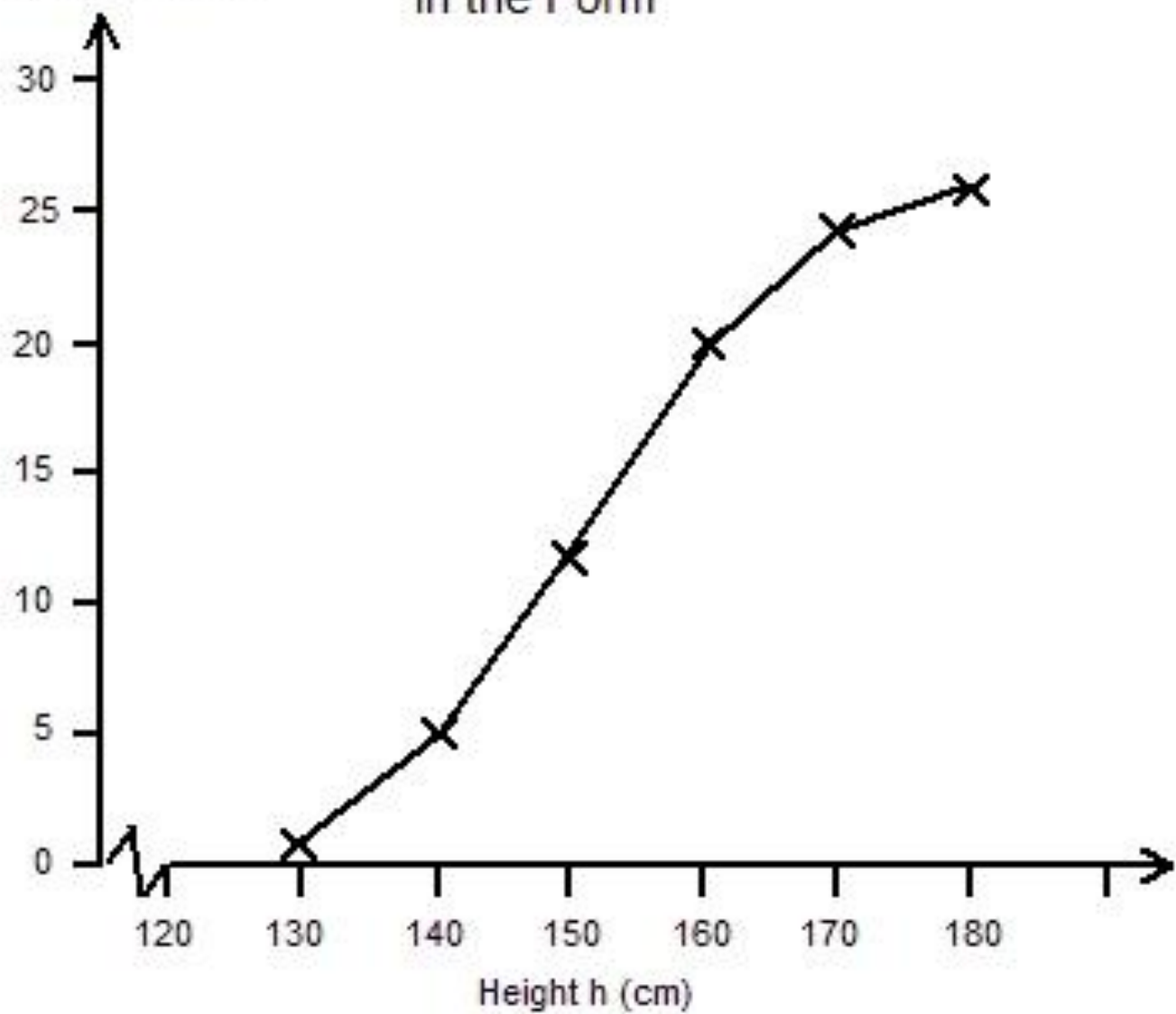
**Distance in miles :: 0.1 , 0.2 , 0.6 , 1.1 , 1.2 , 1.8 , 2.0, 2.7, 3.4 ,4.6 , 6.2 , 8.0 ,12.1 , 14.2**

# CUMULATIVE FREQUENCY

- ❖ *Cumulative frequency* is used to determine the number of observations that lie above (or below) a particular value in a data set.
- ❖ The cumulative frequency is calculated using a frequency distribution table, which can be constructed from stem and leaf plots or directly from the data.
- ❖ The cumulative frequency is calculated by adding each frequency from a frequency distribution table to the sum of its predecessors.
- ❖ The last value will always be equal to the total for all observations, since all frequencies will already have been added to the previous total.

Cumulative Frequency Polygon Showing Heights of Pupils in the Form

Cumulative Freq.



# REPRESENTATION OF STATISTICAL DATA

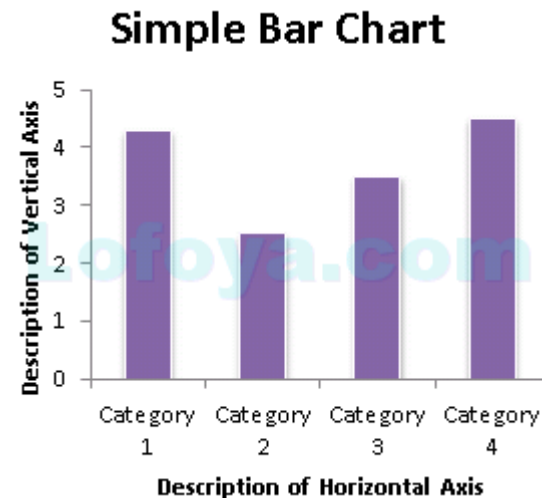


Presentation of collected information in an attractive and informative format

# Bar Diagrams(1-D)

- Information represented as horizontal or vertical bars
- Length of the bars corresponds to magnitude
- Used when data is of different categories
- Eg:-

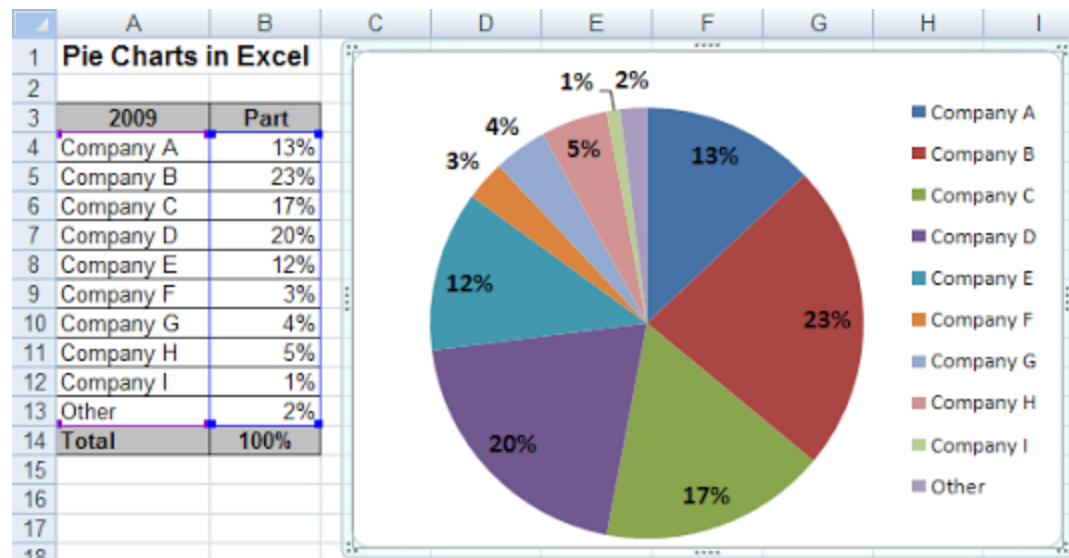
| Categories | Frequency |
|------------|-----------|
| Category 1 | 4.3       |
| Category 2 | 2.7       |
| Category 3 | 3.6       |
| Category 4 | 4.5       |



# Pie Diagrams(2-D)

- Pie diagrams are formed by circles
- Circles are divided into sectors with respect to magnitude represented by its area
- Used when comparison of components are important

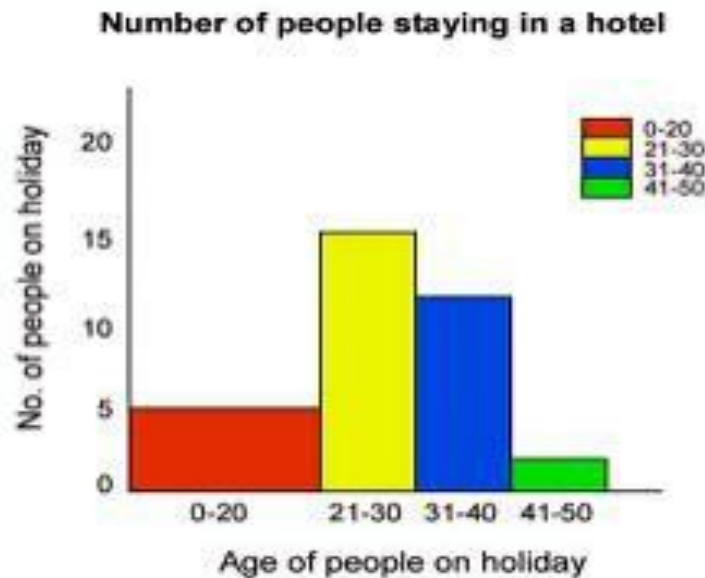
• Eg:-





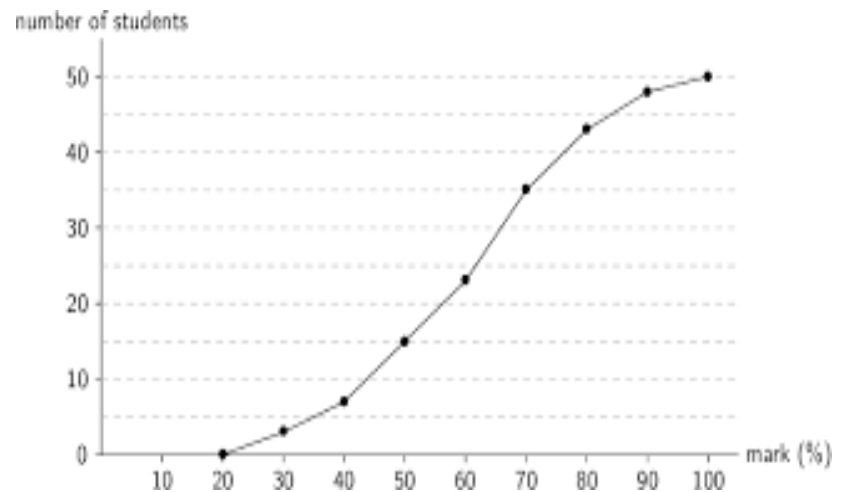
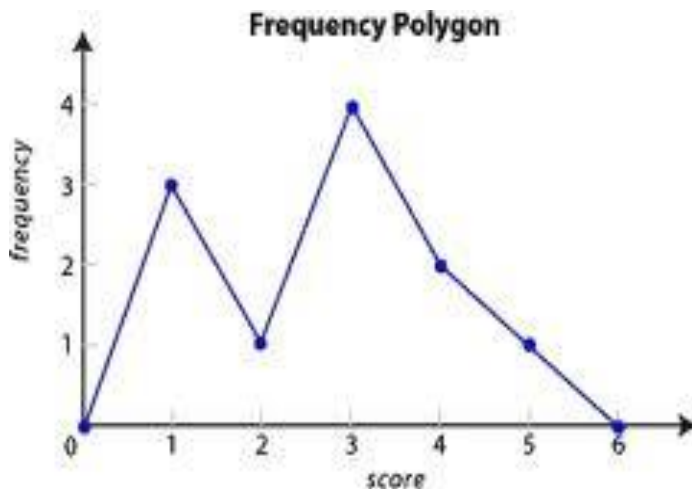
# Histograms(2-D)

- Represents frequency distribution with class boundaries
- Heights of rectangles corresponds to frequency densities and width to class intervals
- Eg:-



# Frequency Polygons and curves

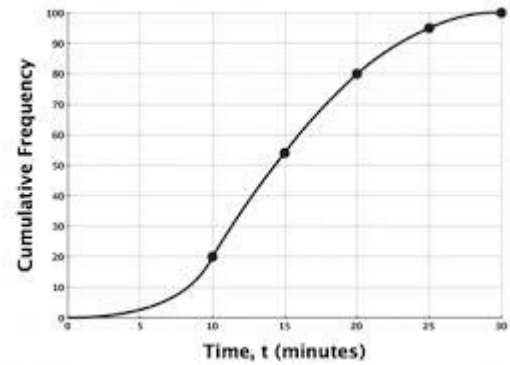
- Useful for comparing distributions
- Frequencies of classes plotted against mid-points of class
- In frequency polygon points are joined by straight line but the same using free hand smooth curves in frequency curve



# Ogives(Cumulative Frequency Curve)

- Two types:-Greater than ogive and Less than ogive
- Less than ogive-Cumulative frequencies are plotted against upper boundary of respective class interval
- Greater than ogive- Cumulative frequencies are plotted against lower boundary of respective class interval

| Time(in mins) | C.f |
|---------------|-----|
| 0             | 0   |
| 10            | 20  |
| 15            | 50  |
| 20            | 80  |
| 25            | 90  |
| 30            | 100 |



# Stem and Leaf Chart (Stemplot)

- Device used to represent quantitative data
- Stem and leaf method retains original data and put it in an order
- It contains 2 columns separated by vertical line with *stems* at left and *leaves* at right

• Eg:-

| Male          |   | Female     |
|---------------|---|------------|
| 5, 2, 0       | 1 | 5, 8       |
| 5, 1          | 2 | 1, 6, 9, 9 |
| 5, 5, 5, 3, 1 | 3 |            |
| 5, 2          | 4 | 1, 2, 6, 8 |
| 9, 8, 6, 1, 1 | 5 | 5          |
| 6, 5, 5, 0    | 6 | 0, 1       |
| 2, 1, 1, 0, 0 | 7 | 2          |

| Stems | Leaves |
|-------|--------|
| 0     | 2 6    |
| 1     | 0 3    |
| 2     | 4 5 8  |
| 3     | 4 4 9  |
| 4     |        |
| 5     | 5 9    |
| 6     | 1 5 7  |

Means 25

Means 55

THANK YOU

A decorative horizontal line consisting of a solid teal bar on top, followed by a white bar, and then three thin teal lines below it, extending across the width of the slide.