ELEMENTORY STATISTICAL CONCEPTS

Stat. 3202 – Statistical methods and Applications (2+1)

Dept. Of Ag. Stat.

STATISTICS

- Aggregates of facts using figures or numeric values, analyse the data and interpretation of the results.
- Central point of statistics – Data
- Statistics – Something used to verify a qualitative statement.
- Limitations;

✓ Doesn't study the qualitative phenomenon

✓ Statistical laws are true to on average

✓ Do not study of an individual

✓ Liable to misuse

DATA COLLECTION

Primary Data

- Conducting Surveys
- Conducting experiments

Secondary Data

- Not originally collected
- Obtained from published or unpublished sources

CLASSIFICATIONS OF DATA

- Process of arranging huge mass of heterogeneous data into groups or classes according to resemblance and similarities.
- Facilitates comparison and proper analysis of data.
- Four Types;

1) Geographical Classification – According to geographical location such as continents, countries, state, districts, panchayats, etc. – E.g. State wise production of rice.
2) Chronological Classification – According to time- E.g. Month wise distribution of rain fall.
3) Qualitative Classification – According to different qualitative characters – E.g. Sex, colour, etc.
4) Quantitative Classification – Can be expressed numerically – E.g. height, weight, yield, etc.

VARIABLE

Continuous Variable

- Assuming fractional values within a range of numbers
- Point to point estimation is possible

Discrete Variable

- Takes values obtained by counting
- Point to point estimation is not possible

TABULATION OF DATA

- Systematic representation of data in columns and rows.
- Used for – presentation of data for the purpose of analysis and statistical inferences.
- "Tabulation as the process of condensing classified data in the form of a table, so that it may be more easily understood and so that any comparison involved may be more really made" – D. Greory and Howard.
- Two types;

1) Simple Tabulation – gives single or one-way table.
2) Complex Tabulation – gives two-way table, information about two related characteristics of a particular phenomenon.

FORMULATION OF FREQUENCY DISTRIBUTION

- Arrangement of data in different classes along with the corresponding class frequency.

- Mainly two types;

1. Discrete Frequency Distribution (DFC) – Suitable for discrete variable, use different values of variable taken together with the frequency of each value.

2. Continues Frequency Distribution (CFC) – Suitable for continues variables, range of variation of continues variable is divided into different classes and the number of values in these classes is found out.

FORMATION OF CLASS INTERVAL

- Class limits (R) – The difference between upper and lower limits of a class.

  $R = L - S$, L – Larger value, S – Smaller value

- Sturge's Rules

  $K = 1 + 3.322 \log_{10} N$, N – Total number , K – No. of classes

- Class Interval (i)

  $i = R / K$

- Class Boundaries – The lower and upper limits of the class.
- Mid value – The average of the class limits or class boundaries.

CUMULATIVE FREQUENCY (CF)

Less than CF (LCF)

- No. of values less than a specified value
- Obtained by adding frequencies in the successive classes

Greater than CF (GCF)

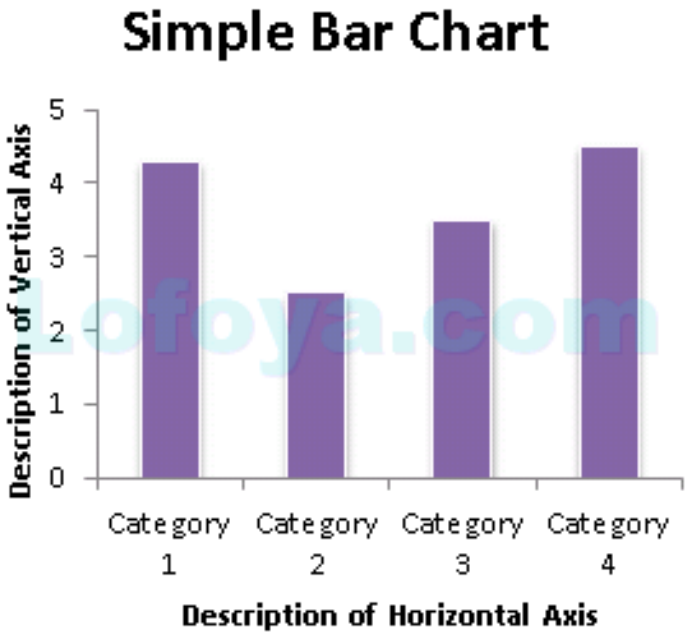
- No. of observation greater than a specified value
- Obtained by subtracting frequencies from the total frequency

Diagrammatical Representation of data

- Used to understand the nature of data diagrammatically or it will give visual impression of data.
- 1-D Diagrams – Height of bars is taken into consideration.

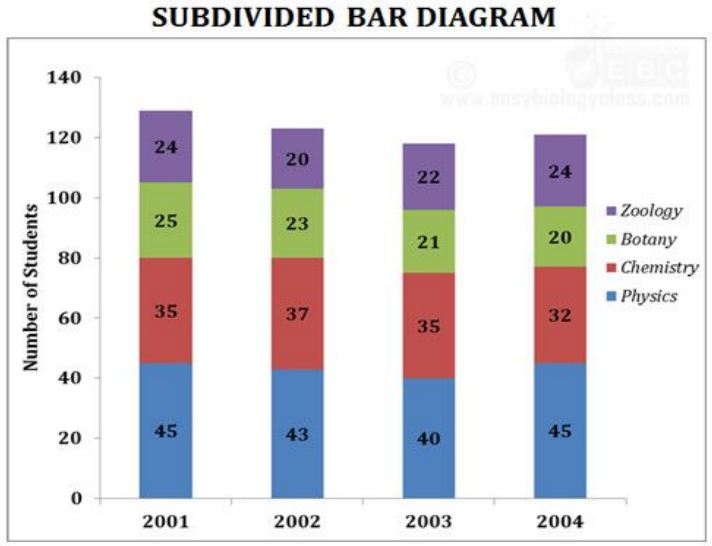
1. Simple Bar Diagram;

Number of bars drawn over a horizontal line at equal distance with equal width, whose height being proportional to the magnitude of the variable.

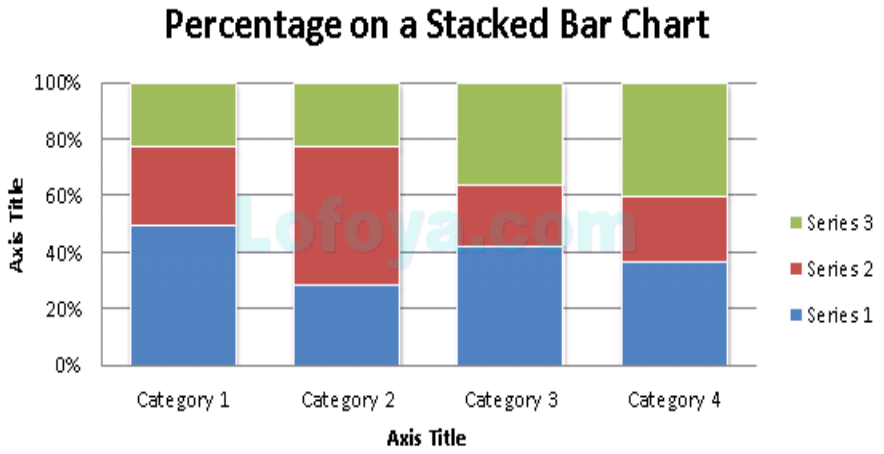


2. Sub divided bar diagram;

Exhibit the division of the total of a variable into different component parts. The magnitude of component is proportional to components part of the variable.



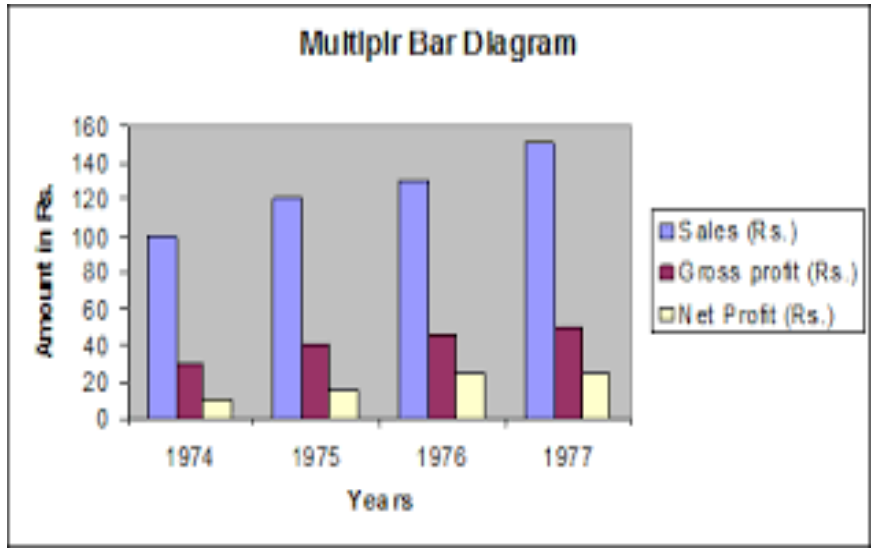
3. Percentage bar diagram;



The length of each bar is the same, taken as 100 in general. Divide each bar into small rectangles to represent the magnitude of the component part.

4. Multiple bar diagram;

To show two or more inter related values of the variable. Consists of a group of bars drawn side by side to represent different related variables.



- 2 –D Diagrams;
- Here height and width are taken into consideration
- Also called Area diagrams or surface diagrams.
- Area = length * breadth

1. Rectangle graph;

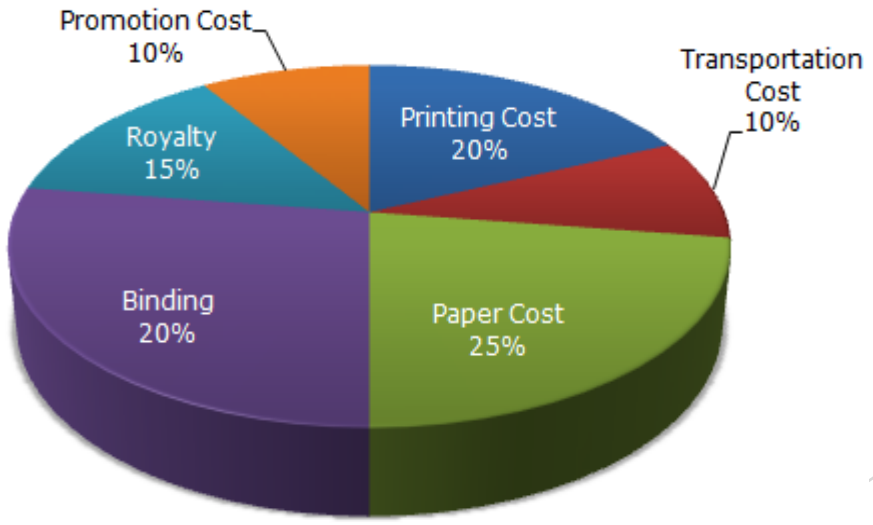
- Commonly used, represent two or more than two set of data
- Two methods used; i). Width equal and height varies with proportional and ii). Height equal, width varies with proportional.

2. Square graph;

- Represent the data when the size of the item to be compared different greatly in magnitude.
- The length of the square is proportional to the size of the different items.

- 3. Circles / pie diagrams;
- Consists of circle, whose area being proportional to the total of the magnitude and the circle is subdivided into different sectors to represent component parts.
- The radius of a circle is generally taken as the square root of the total of the magnitude.

The component parts are represented by sectors of the circle, whose angle is proportional to the magnitude of the component part.



- 3 –D Diagrams;
  - Cylinders, Spheres and cubes
  - Consider length, breadth and thickness

- Pictograms – Relative values of the items are represented by pictures.

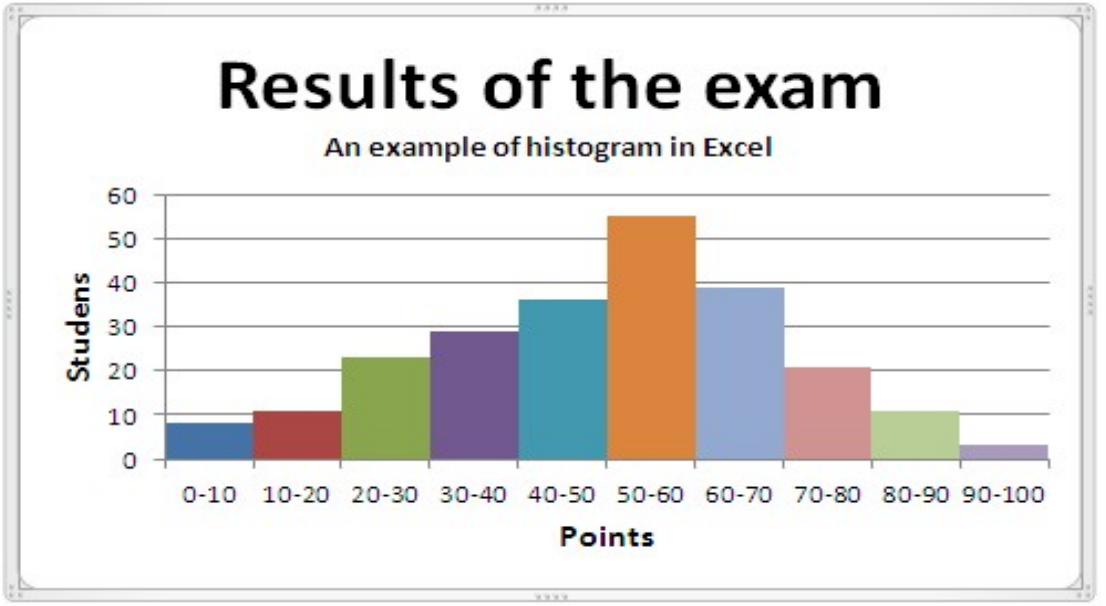
No. of pictures drawn or size of the different pictures is in proportional to the values of various items.

- Pyramids – Data relating the distribution of various age groups.

- Cartograms – Regional distribution of data are shown by the use of maps.

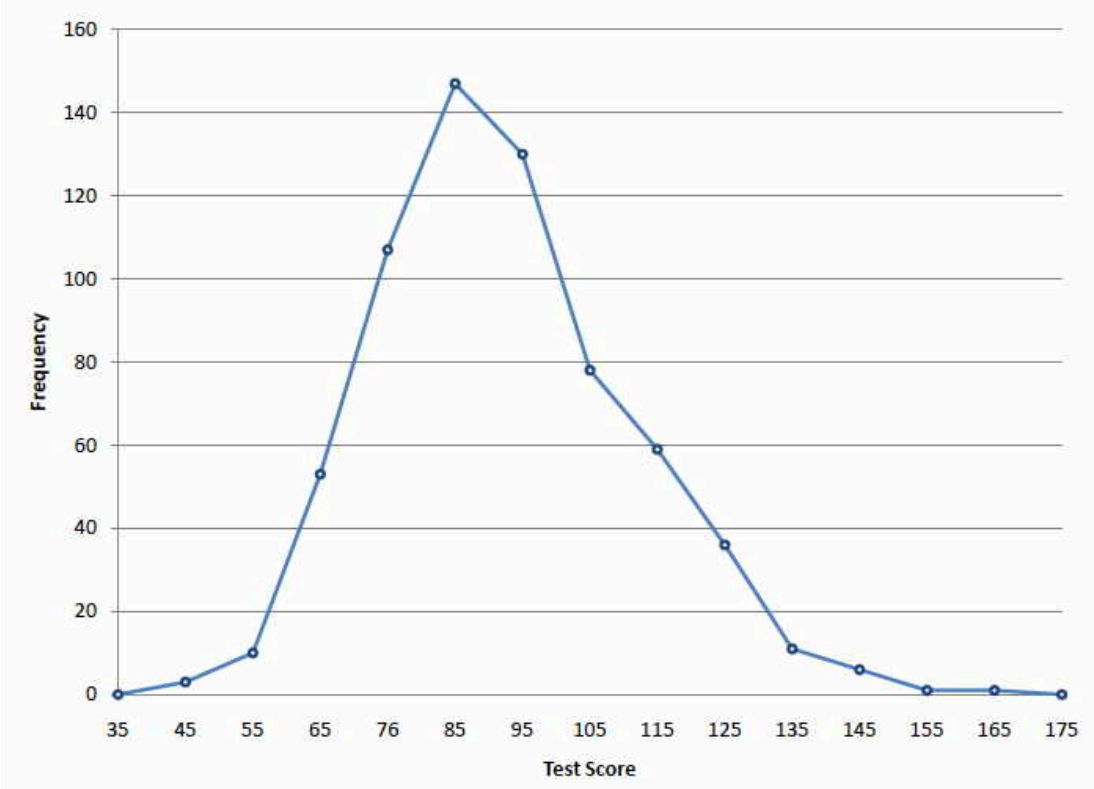
Graphical Representation of data

- Histograms;
  - Consists of adjacent rectangles erected over true class limits.
  - Whose height being proportional to the magnitude of the corresponding class frequencies.
  - Width – class interval, Height – frequency, Area – height * width
  - For unequal class interval. Frequency density = class frequency / width



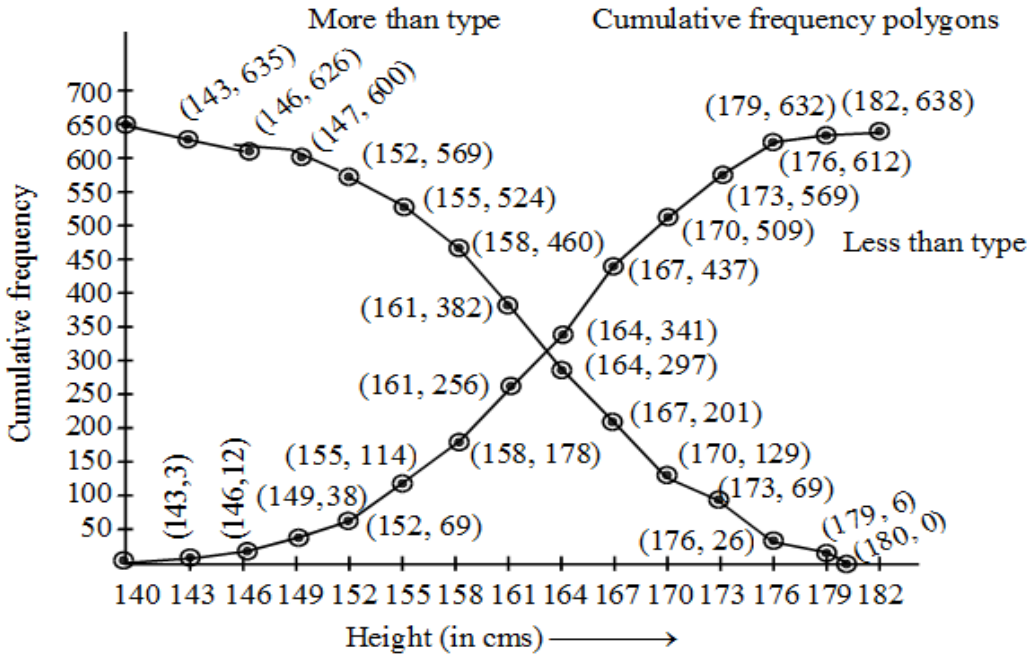
Frequency Polygon;

- Mark the mid value of the classes on X- axis, Frequencies on the Y-axis and join these points by a straight line.



If the points in frequency polygon are joined by a smooth curve – Frequency curve

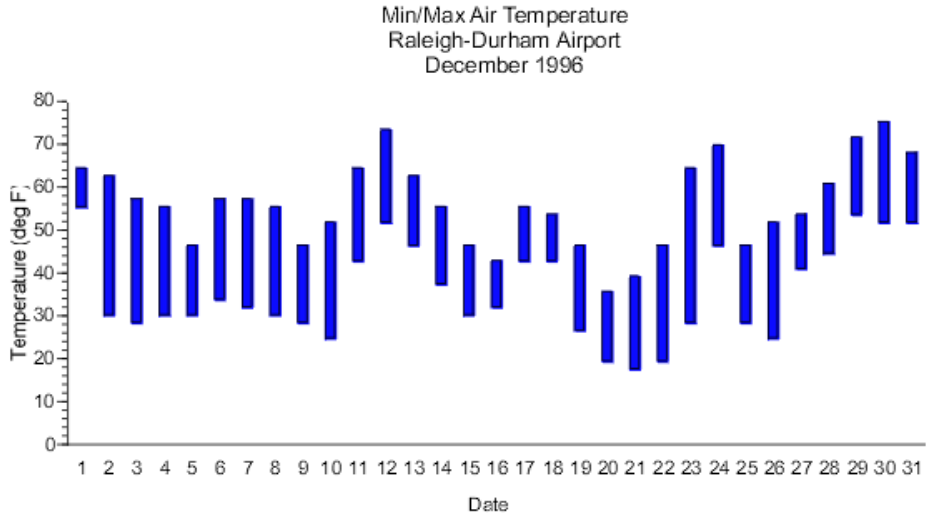
- Ogives;
  - Less than ogives / LCF Ogives – Upper limits of the classes on X – axis and LCF on Y – axis.
  - Greater than ogives / GCF Ogives – Lower limits of the classes on X – axis and GCF on Y – axis.



Cumulative Frequency polygon of less than and more than type

Range graph

Used to represent the minimum, maximum and average values of a variable



Band graph

To represent the total of various items at different time points