

STATISTICS

Characteristics of Statistics as Data

- Relate to aggregate of facts
- Affected by multiple causes
- Numerically expressed
- Estimated by reasonable standard of accuracy
- Collected in systematic manner
- Collected for predetermined purpose
- Placed in relation to each other

Aggregate of facts

- Single observation cannot be called as Statistics
- Production of a food grain in a particular year, percentage marks of a single student can not be Statistics
- Yearly production of food grain for few years, percentage marks of all students in a class can constitute Statistical Data

Affected by multiple causes

- If we consider production of a rice in a Maharashtra for few years , obviously the figures will differ in every year.
- This is because the Statistical data are affected to a marked extent by multiple causes.
- The causes for change in the production of rice may be rainfall , fertility of soil , seeds & fertilizers used, methods of cultivations etc.
- The percentage marks scored may be affected by intelligence , guidance , capacity to do hard work etc.

Numerically expressed

- Statistical data should be expressed in terms of numbers.
- Production of sugar is excellent in Maharashtra, Performance of students at HSC Examination has improved over a period of time can not constitute Statistical Data.
- Production of sugar in Maharashtra , Pass percentage of students at HSC Examination should be recorded over a period of time to constitute Statistical Data.

Estimated by reasonable standard of accuracy

- Estimation is always crude expression without actual measurement.
- If a teacher casually says that 50 students attended the lecture , 200 students witnessed Annual Athletic Meet . It means the figures are estimated through just inspection without actual measurement. However , if these figures are to be included in any report then the teacher has to refer to the signature sheets to arrive at these figures.
- The accuracy standard should be pre decided

Collected in systematic manner & for pre determined purpose

- The purpose should be well defined and clear, otherwise some unnecessary information may be collected or necessary information may be ignored.
- Suitable plan for data collection to be chalked out initially. This may include what data to be collected , from whom to be collected , what method of collection to be adopted , what degree of accuracy to be aimed etc.

Placed in relation to each other

- Statistical data are always placed in relation to each other ie they are comparable.
- The comparison is either period wise or region wise.
- Eg. Production of food grain in a particular state should be observed over few years (period wise) or Production of food grain in different states in a particular year should be compared (region wise.)

Characteristics of Statistics as Method

- The large amount of numerical information give rise to need for systematic method which can be used to organise , present , analyse and interpret the information.
- Collection
- Organisation
- Presentation
- Analysis
- Interpretation

Collection

- It is the first fundamental step in the statistical investigation
- The data should be collected with utmost care as wrong data will lead to wrong results.
- Primary data – The collected for the first time by investigator himself
- Secondary data – Data collected by somebody else and it well organised and published.

Organisation

- The data collected from published sources (secondary data) are generally in organised form
- The data collected from actual surveys (primary data) needs to be organised
- It can be done in three stages
- Editing
- Classifying
- Tabulation

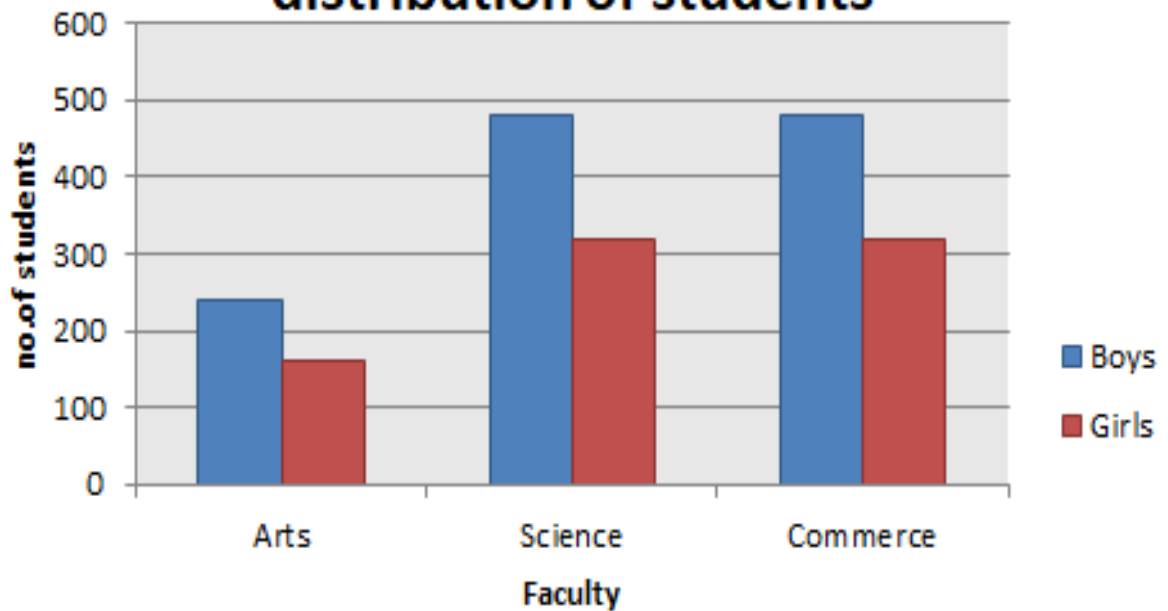
Presentation

- Presenting data in organised manner facilitate statistical analysis.
- The data can be presented in the form of
- Tables
- Graphs
- Diagrams

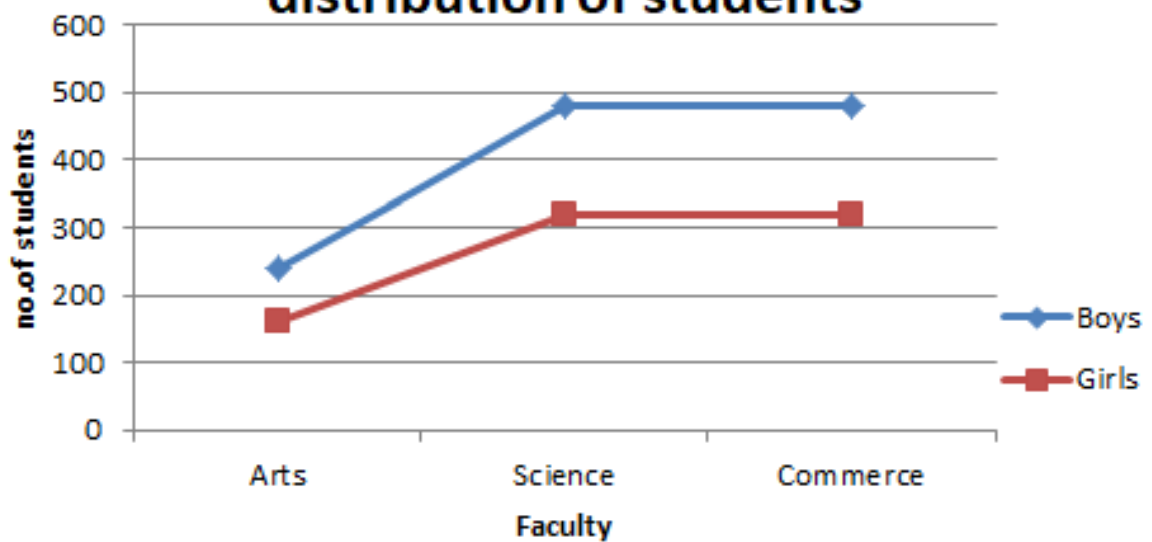
Gender wise distribution of students in a college

Faculty \ Gender	Arts	Science	Commerce	Total
Boys	240	480	480	1200
Girls	160	320	320	800
Total	400	800	800	2000

Genderwise & Facultywise distribution of students



Genderwise & Facultywise distribution of students



Analysis

- The most commonly used form of data for analysis is Tabular Form.
- Numerous methods are available of analysis of the data ranging from simple inspection to complicated techniques.
- Measures of Central tendency , Measures of variation , correlation , regression etc

Interpretation

- Interpretation is the final stage in the Statistical Investigation
- It means drawing conclusion from the data collected and analysed
- It requires great skills and experience

Basic concepts

- Classification of data means separating data according to similar characteristics and grouping them
- **Geographical Classification (Classification on the basis of regions)**

City	Temperature
New Delhi	43
Mumbai	32
Chennai	36
Kolkata	37
Surat	42
Pune	31

- **Chronological Classification (Classification on the basis of time)**

Year	Production in CR
1995	27
1996	29
1997	32
1998	35

- **Qualitative Classification (Classification on the basis of qualities)**

-

Students in F.Y.B.COM. A

Male 65

Female 55

- **Quantitative Classification**

Income	No. of workers
Up to 10000	200
10001-20000	400
20001- 50000	500
50001 onwards	100

The way in which items are distributed into various classes is known as Frequency distribution

Frequency distribution

- **Class interval** is a range into which data is grouped.
- **Lower limit** – Lowest value in the class interval
- **Upper limit** – Highest value in the class interval
- **Class frequency** – Total number of items belonging to same interval
- **Class width** - Upper limit – Lower limit
- **Class mark**- $(\text{Upper limit} + \text{Lower limit}) / 2$
- **Open Ended Intervals** – One of the class limit is not rigidly defined eg up to 20 , 70 or more
- **Closed Ended Intervals** – Both the class limits are rigidly defined eg 10 to 20 , 90 to 100
- **Inclusive Intervals** - Both the class limits are included in the intervals eg. 1-10 , 11-20 etc
- **Exclusive Intervals** – Upper class limit is excluded from the intervals eg. 0-10 , 10-20 , 20-30 etc
- **Inclusive Intervals**

Income	No. of workers
1- 10000	200
10001-20000	400
20001- 50000	500
50001 - 70000	100

- **Exclusive Intervals**

Income	No. of workers
0- 10000	200
10000-20000	400
20000- 50000	500
50000 - 70000	100

Cumulative Frequencies

- Less than type cumulative frequencies – it denotes number of observations less than upper limit
- Greater than type cumulative frequencies – it denotes number of observations greater than or equal to the lower limit

Less than type Cumulative Frequencies

Marks	No.of students	Cumulative Freq.(<)
0- 20	2	2
20-40	10	$10+2 =12$
40-60	28	$28 + 12 =40$
60-80	25	$40 + 25 =65$
80-100	15	$15 + 65 =80$

Income	No. of workers	Cum Freq	less than
0- 10000	200	200	
10000-20000	400	600	
20000- 50000	500	1100	
50000 - 70000	100	1200	

Greater than type Cumulative Frequencies

Marks	No.of students	Cumulative Freq.(\geq)
0- 20	2	$78+2 =80$
20-40	10	$68 +10 =78$
40-60	28	$40 +28 =68$
60-80	25	$15 +25 =40$
80-100	15	15

Income	No. of workers	Cum Freq	greater than
0- 10000	200	1200	
10000-20000	400	1000	
20000- 50000	500	600	
50000 - 70000	100	100	