Statistics

Ch-1 → Sir Ronald A Fisher → father of statistics.

Keywords

Population	Sample
Parameter → eg avg.	Sample data

Descriptive Stats → describing property of data.
Inferential Stats → drawing conclusion from info

Data
- → Qualitative / Categorical.
- → Quantitative

Roles of Statistics in day-to-day life :-

1.] Economy
2.] Tourism
3.] Business
4.] Quality control [Industry → production]
5.] Insurance [premium of insurance]
6.] Research
7.] Medical
8.] Education
9.] War
10.] Meteorological department

• Limitations of Stats :-

1.] not suited to the study of qualitative phenomenon.
2.] not a study of individuals.
3.] Statistical laws are not exact.
4.] Statistics is liable to be misused.

Scales of Measurements

→ Way in which variables are defined.

4 Scales → Nominal
→ Ordinal
→ Interval
→ Ratio

Properties of Measurement :-

1.] Identity
2.] Magnitude
3.] Equal Interval
4.] Minimum value of zero.

• Nominal measurement

→ identity property of data.
→ non-numeric values.
→ order is not compulsory.

eg- eye colour, country, etc.

Nominal with order :- warm, cold, hot & very hot.
Nominal without order :- male & female.
Dichotomous :- Only two categories yes or no.

- Ordinal Scale -

→ used to depict non-mathematical ideas.
eg - frequency, satisfaction, etc.

very satisfied	①
satisfied	②
Neutral	③
bad	④
very bad	⑤

zero means data doesn't exist.

Limitation → we cannot measure it.

- Interval Scale.

→ 'Zero' Exists

→ Scales are numeric scales with order & proper interval of time / difference.

- Ratio Scale Measurement :-

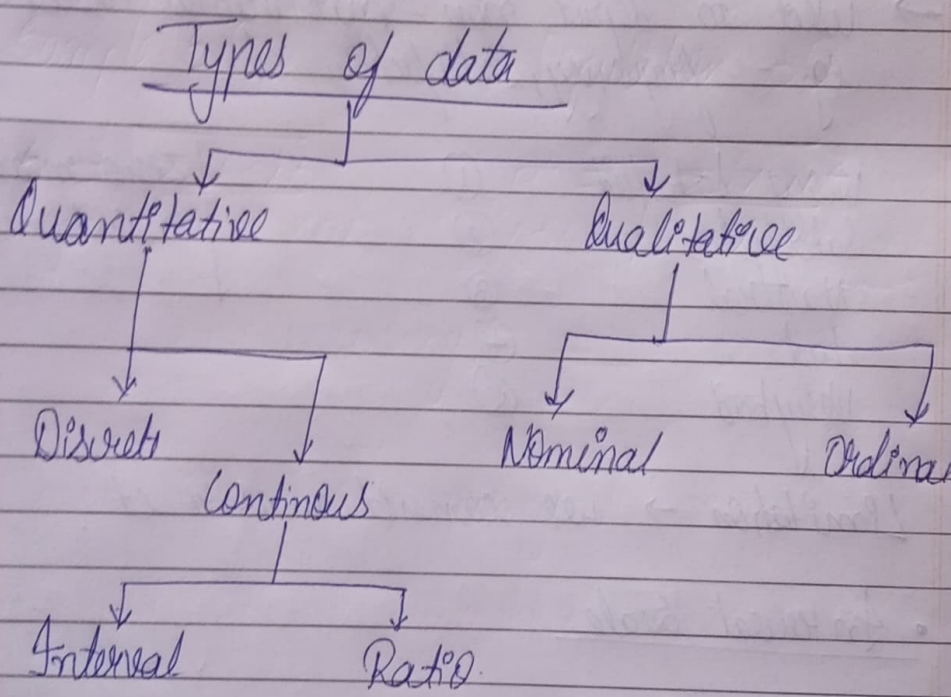
→ data is nominal & defined by an identity, classified in an order, contains intervals & can be broken down into exact values.

eg - weight, height, distance.

→ Can be added, '−', '÷' & 'x'.

→ 'True zero'. Zero means data has no value point.

Eg- Shares & Sales.



Ch-2 Measures of Central Tendency

→ Mean ($\bar{x}$)

$\bar{x} = \frac{\Sigma x}{n}$ sample mean.

$\mu = \frac{\Sigma x}{N}$ population mean.

→ Median (M)

divides data in two exact halfs.

even ~~N/2 + N+1/2~~ $N/2$, $\frac{N}{2}+1$.

odd. $\left[\frac{N}{2}+1\right]$

→ Mode

most frequently occuring data.

• Measures of Variability or dispersion.

3 measures → 1) Range
2) Variance
3) Standar deviation

$R = x_{max} - x_{min}$

Variance & standard deviation.

Sample Variance : Population Variance

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n-1} \quad ; \quad \sigma^2 = \frac{\Sigma (x - \mu)^2}{N}$$

Sample SD : Population SD.

$$s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n-1}} \quad ; \quad \sigma = \sqrt{\frac{\Sigma (x - \mu)^2}{N}}$$

ch-3 Data Analysis

- Steam & leaf diagram
- Frequency histogram
- Relative frequency histogram
- Dot plots
- Bar charts
- Pie charts
- Symmetry

1] Steam & leaf diagram

→ unit place measurement — leaf
tens place measurement — stem

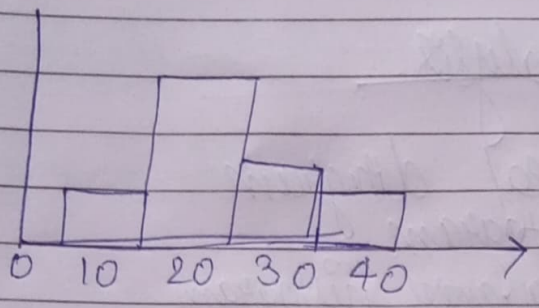
tens place ↓	unit place →

if is in order than ordered steam & leaf dia.

→ small data sets.
→ original data can be recovered.

2] Frequency histogram

→ large data sets.
→ length of vertical bar = no. of observation of group.



3] Relative frequency histogram.

- Shows proportion

eg- frequency → (1)/(30) ← total outcome.

Relative frequency = $\frac{1}{30}$ = 0.03.

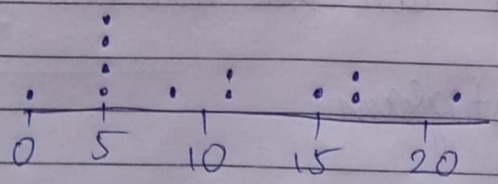
plot: Relative frequency Vs Score.

4] Dotplots of Numerical Data

→ given data set is extremly small.

→ if value repeats, stack the dot vertically.

eg.

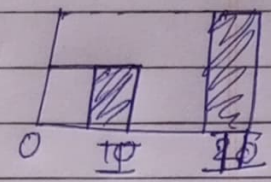


⇒ For categorial data set, use frequency distribution table.

Relative frequency = $\frac{\text{frequency}}{\text{no. of observation in data set}}$

5.] Bar charts:-

→ width of each bar similar



6.] Pie chart

Circle → whole data
slices → possible categories

less categories most effective.

Slice size :- 360 × (category relative frequency)

• Symmetry

- close to being symmetric are called approximately symmetric
- data set is symmetric about value x_0 if the frequency of values $x_0 - c$ & $x_0 + c$ are same for all c.