Central Tendency & Dispersion

CHEAT SHEET
(कचरा)

Central Tendency

meaning of central tendency
→ central value of all observation
→ Representative of entire series

Characters of Good central Tendency
→ Easy To calculate → mean & mode
→ Easy To understand → mean
→ Based on all observation → mean, GM, HM
→ Rigidly Defined → mean
→ least affected by extreme values → median
→ Have mathematical properties → mean, GM & HM

Note : → AM is Best central Tendency
→ median is best for open ended series

Arithmetic mean ($\bar{x}$)

$$AM = \frac{\text{Sum of all observation}}{\text{Total no. of observations}}$$

methods

Direct method

$$\bar{x} = \frac{\Sigma f_i x_i}{N}$$

(where $N = \Sigma f_i$)

Shortcut Method

→ Assume mean method

If $x_i - A = d_i$

Then $\bar{x} = A + \frac{\Sigma f_i d_i}{N}$

→ Step Deviation method

If $\frac{x_i - A}{h} = u_i$

Then $\bar{x} = A + \frac{\Sigma f_i u_i}{N} \times h$

Properties

1> If all observations are same (let K)
then mean also K

mean of 5, 5, 5, 5, 5 will be 5

2> Sum of Deviations from their arithmetic mean is always zero

i.e. $$\Sigma (x_i - \bar{x}) = 0$$

3> Sum of the squares of Deviation is minimum when Deviations are taken from their arithmetic mean

i.e. $\Sigma (x_i - K)^2$ is minimum
when $K = \bar{x}$

4> Combined mean

$$\bar{x}_{12} = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2}{N_1 + N_2}$$

5> Arithmetic mean changes with change of origin & change of scale

i.e. If $y_i = a x_i + b$
Then $\bar{y} = a\bar{x} + b$

$$1+2+3+\cdots+n = \frac{n(n+1)}{2}$$

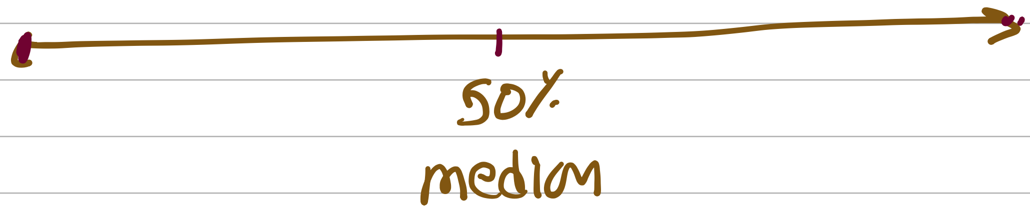
$$1^2+2^2+3^2+\cdots+n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$1^3+2^3+3^3+\cdots+n^3 = \left[\frac{n(n+1)}{2}\right]^2$$

A.M. of First 'n' natural numbers $= \frac{n+1}{2}$

Median

A number which divides entire Distribution in two Parts is known as median. It represents half (50%) of Total numbers.



→ It is not based on all observations so it can be used for open ended series.

methods

$\swarrow$ for individual & Descrete series

$\searrow$ For continous series

median = $\left(\frac{n+1}{2}\right)^{th}$ term

$\Downarrow$

→ make cf column

→ Locate $\frac{N}{2}$ in cf

→ select median class

→ use formula

$$\text{median} = l + \left[\frac{\frac{N}{2} - cf}{f}\right] \times h$$

→ median changes with change of origin & change in scale

i.e if $y = a + bx$

Then med. of $y = a + b(\text{med. of } x)$

→ Sum of absolute deviation is minimum when deviations are taken from the median

ie $\Sigma |x_i - K|$ is minimum when $K = \text{median}$

Partition values (Fractiles)

Quartiles (Q_1, Q_2 & Q_3)

Divides entire series in 4 parts)

Q_1	Q_2	Q_3
$=25\%$	$=50\%$	$=75\%$
$\left(\frac{N}{4}\right)$	$\left(\frac{N}{2}\right)$	$\left(\frac{3N}{4}\right)$

For Individual & Descrete series

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{ term}$$

$$Q_2 = \text{median} = \left(\frac{n+1}{2}\right)^{th}$$

$$Q_3 = \left[3\left(\frac{n+1}{4}\right)\right]^{th} \text{ term}$$

In continous series

for Q_1

→ Locate $\frac{N}{4}$ in cf

→ select 'Q_1' class

→ $Q_1 = l + \left[\frac{\frac{N}{4} - cf}{f}\right] \times h$

For Q_3

→ Locate $\frac{3N}{4}$ in cf

→ select Q_3 class

→ $Q_3 = l + \left[\frac{\frac{3N}{4} - cf}{f}\right] \times h$

Deciles ($D_1, D_2, \ldots, D_9$)

Divides entire series in 10 parts

D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9
↓	↓	↓	↓	↓	↓	↓	↓	↓
10%	20%	30%	40%	50%	60%	70%	80%	90%

50% ⇓ 5^{th} Decile

$= D_5 =$ median

For individual & Descrete series

$D_1 = \left(\frac{n+1}{10}\right)^{th}$ term

$D_2 = \left[2\left(\frac{n+1}{10}\right)\right]^{th}$

$D_K = \left[K\left(\frac{n+1}{10}\right)\right]^{th}$ term

For continous series

For D_1

→ locate $\frac{N}{10}$ in cf

→ select D_1 class

→ $D_1 = l + \left[\frac{\frac{N}{10} - cf}{f}\right] \times h$

For D_3

→ Locate $\frac{3N}{10}$ in cf

→ select D_3 class

→ $D_3 = l + \left[\frac{\frac{3N}{10} - cf}{f}\right] \times h$

Percentiles $(P_1, P_2, \ldots\ldots P_{99})$

⇓

Divides entire series in 100 parts

P_1 → 1%, P_2 → 2%, ... P_{49} → 49%, P_{50} ⇒ median → 50%, P_{51} → 51%, ... P_{99} → 99%

For Individual & Descrete series

$$P_1 = \left(\frac{n+1}{100}\right)^{th} \text{ term}$$

$$P_6 = \left[6\left(\frac{n+1}{100}\right)\right]^{th} \text{ term}$$

$$P_K = \left[K\left(\frac{n+1}{100}\right)\right]^{th}$$

For continous series

For P_1

→ Locate $\frac{N}{100}$ in cf

→ select P_1 class

→ $P_1 = l + \left[\frac{\frac{N}{100} - cf}{f}\right] \times h$

For P_7

→ Locate $\frac{7N}{100}$ in cf

→ select P_7 class

→ $P_7 = l + \left[\frac{\frac{7N}{100} - cf}{f}\right] \times h$

mode

⇓

An observation with highest frequency

Individual series

eg marks: 2, 1, 3, 1, 4, 3, 2, 3, 5, 2, 3, 1, 3, 1, 3, 5, 3, 4, 3

mode = 3

Descrete series

eg

x_i	f_i
2	6
3	12
4	3
5	5

→ mode = 3

For continous series

→ Check the class interval with highest frequency (It is modal class)

→ Use formula

$$\text{mode} = l + \left\{ \frac{f_1 - f_0}{2f_1 - f_2 - f_0} \right\} \times h$$

Note: mode also changes with change of origin or change in scale

ie. $y = a + bx$

Then mode of $y = a + b(\text{mode of } x)$

Geometric mean

n^{th} Root of the Product of n observations

used for
→ Average Dep?
→ Average of %

For individual series

$$GM = (x_1 \times x_2 \times \cdots\cdots x_n)^{1/n}$$

For Descrete & continous series

$$GM = \left[(x_1)^{f_1} \times (x_2)^{f_2} \times \cdots \times x_n^{f_n}\right]^{\frac{1}{N}}$$

Properties

1> If all items are same (let K)
then $GM = K$

2> $$\log(GM) = \frac{\Sigma \log x_i}{N}$$

$$GM = AL\left[\frac{\Sigma \log x_i}{N}\right]$$

3) $GM(xy) = GM(x) \times GM(y)$

4) $GM\left(\frac{x}{y}\right) = \frac{GM(x)}{GM(y)}$

5) combined Geometric mean

$$G = \left[(G_1)^{N_1} \times (G_2)^{N_2}\right]^{\frac{1}{N_1+N_2}}$$

Harmonic mean

used for
→ Avg speed
→ Avg of Rates

Reciprocal of Average of Reciprocal of all 'N' observations

→ find reciprocal of all items
→ find Average of these reciprocals
→ find Reciprocal of Average

For individual series

$$HM = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{N}{\Sigma\left(\frac{1}{x_i}\right)}$$

for Descrete & Continous series

$$HM = \frac{N}{\Sigma\left(\frac{f_i}{x_i}\right)}$$

Properties

1> If all items are same (let K) then $HM = K$

2> Combined $HM = \dfrac{N_1 + N_2}{\frac{N_1}{H_1} + \frac{N_2}{H_2}}$

Relation b/w AM GM & HM

$$AM \geq GM \geq HM$$

If all items are different

$$AM > GM > HM$$

For any two items a & b

$$AM \times HM = GM^2$$

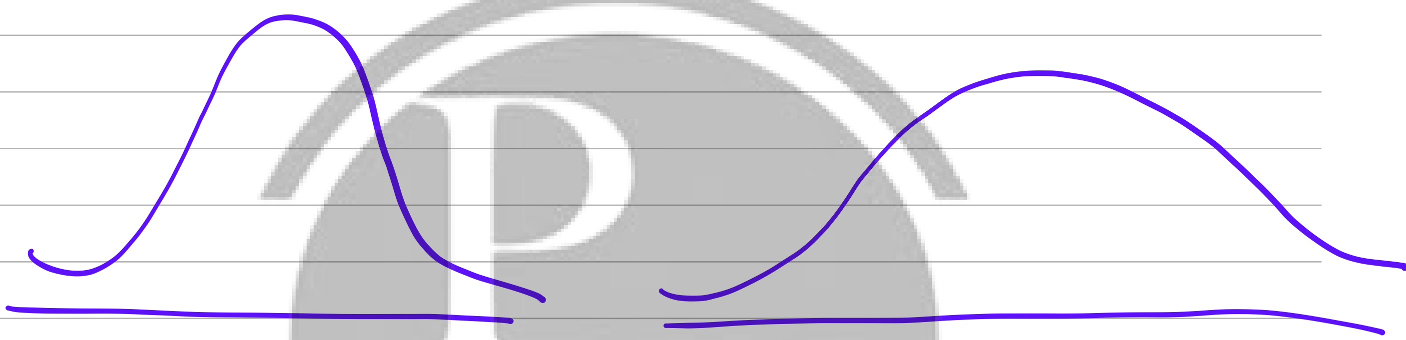
$$\text{weighted AM} = \frac{\Sigma w_i x_i}{\Sigma w_i}$$

$$\text{weighted HM} = \frac{\Sigma w_i}{\Sigma \left(\frac{w_i}{x_i} \right)}$$

$$\text{weighted GM} = AL \left[\frac{\Sigma w_i \log x_i}{\Sigma w_i} \right]$$

Dispersion

Statistical Technique to find Degree consistency (or variability) in all the observation.



→ low Dispersion
→ concentrated Data
→ less variability
→ more consistency

→ High Dispersion
→ scattered Data
→ more variability
→ less consistency

measures

Absolute
→ Range
→ Q.D
→ M.D
→ S.D.

Relative
→ coefficient of Range
→ coefficient of QD.
→ coefficient of MD
→ coefficient of variance

Range

Difference b/w Largest & smallest observation

$$\boxed{R_x = L - S}$$

$$\boxed{\text{coefficient of Range} = \frac{L-S}{L+S} \times 100}$$

→ Range Does not Change with origin

→ Range changes with scale

$$\boxed{\begin{aligned} &\text{If} \quad y_i = a + b\,x \\ &\text{then} \quad R_y = |b| \times R_x \end{aligned}}$$

Quartile Deviations

⇒ Interquartile Range $= Q_3 - Q_1$

⇒ Semi quartile Range (Quartile Deviation) $= \dfrac{Q_3 - Q_1}{2}$

⇒ coefficient of Q.D $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$

or

$\dfrac{QD}{\text{median}} \times 100$ ⇒ only for symmetrical Distribution

→ QD does not change with origin

→ QD changes with scale

$$\boxed{\begin{aligned} y_i &= a + bx \\ QD \text{ of } y &= |b| \times QD \text{ of } x \end{aligned}}$$

Mean Deviation (MD)

Average of Absolute Deviations taken from mean, median or mode

MD

About mean

$$MD_{\bar{x}} = \frac{\sum f_i |x_i - \bar{x}|}{N}$$

About median

$$MD_m = \frac{\sum f_i |x_i - m|}{N}$$

$$\text{coefficient of M.D.} = \frac{MD}{\text{mean}} \times 100$$

or

$$\frac{MD}{\text{median}} \times 100$$

=> MD does not change with origin

=> MD changes with scale

If $y = a + bx$

then $MD_y = |b| \times MD_x$

Standard Deviation & variance

$$\text{variance} = \sigma^2$$

$$S.D. = \sigma$$

$$S.D. = \sqrt{\text{variance}}$$

$$\text{variance} = (S.D.)^2$$

$\underset{(\sigma)}{S.D.} = \sqrt{\dfrac{\Sigma f_i (x_i - \bar{x})^2}{N}}$

$S.D. = \sqrt{\dfrac{\Sigma f_i x_i^2}{N} - \left(\dfrac{\Sigma f_i x_i}{N}\right)^2}$

$S.D. = \sqrt{\dfrac{\Sigma f_i d_i^2}{N} - \left(\dfrac{\Sigma f_i d_i}{N}\right)^2}$

where $d_i = x_i - A$

$S.D = \sqrt{\dfrac{\Sigma f_i u_i^2}{N} - \left(\dfrac{\Sigma f_i u_i}{N}\right)^2} \times h$

where $\dfrac{x_i - A}{h} = u_i$

⇒ SD Does not change with origin

⇒ S.D. change with change of scale

If $y_i = a + bx$

S.D of y_i = $|b| \times$ S.D of x

variance of $y_i = b^2 \times$ variance of x_i

coefficient of variance (CV) $= \frac{\sigma}{\bar{X}} \times 100$

⇓

will be used for consistency & variability

Higher CV ⇒ Higher variablity

lesser CV ⇒ Higher consistency

S.D. of first 'n' natural number $= \sqrt{\dfrac{n^2-1}{12}}$

S.D of first 'n' even natural no. $= \sqrt{\dfrac{n^2-1}{3}}$

S.D of first 'n' odd natural no. $= \sqrt{\dfrac{n^2-1}{3}}$

S.D. of two numbers a & b $= \dfrac{|a-b|}{2}$

Q.D : MD : SD = 10 : 12 : 15

$\dfrac{QD}{MD} = \dfrac{10}{12}$ | $\dfrac{MD}{SD} = \dfrac{12}{15}$ | $\dfrac{QD}{SD} = \dfrac{10}{15}$

Combined S.D.

$$= \sqrt{\frac{N_1(\sigma_1^2 + d_1^2) + N_2(\sigma_2^2 + d_2^2)}{N_1 + N_2}}$$

where $d_1 = \overline{X}_{12} - \overline{X}_1$

$d_2 = \overline{X}_{12} - \overline{X}_2$

& $\overline{X}_{12} = \dfrac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$