~~Mean~~

Types of Data

- Grouped
  - Continuous
- Ungrouped
  - Individual
  - Discrete

Mean

→ **Individual**

Direct Method: $\overline{X} = \dfrac{x_1 + x_2 + x_3 + \dots + x_n}{N}$

$d_i = x_i - A$, A = assumed mean N = no. of observations

Short-cut Method:

$$\overline{X} = A + \frac{1}{N}\sum_{i=1}^{n} d_i$$

→ **Discrete**

Direct Method: $\overline{X} = \sum_{i=1}^{n} f_i x_i \cdot \frac{1}{N}$ N = sum of observations

$d_i = x_i - A$, A = assumed mean

Short-cut Method: $\overline{X} = A + \frac{1}{n}\sum_{i=1}^{n} f_i d_i$

~~Grouped~~
Continuous

Direct: $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} f_i m_i$ m_i is the mid-pt. of class interval

Short-cut Method: $\overline{x} = A + \frac{1}{n} \sum_{i=1}^{n} f_i d_i$, $n = \Sigma f$

$\cancel{d_i = x_i - A}$ $d_i = m_i - A$

Step Deviation Method

$$\overline{x} = A + \frac{\Sigma f_i u_i}{n} \times h$$

$\cancel{u_i = \frac{x - A}{h}}$ $u_i = \frac{m_i - A}{h}$, h is width of class interval

Median

Ungrouped Data:

- odd (n) observations

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

- even (n) observations

$$\text{Median} = \frac{(n/2)^{th} \text{term} + (n/2 + 1)^{th} \text{term}}{2}$$

Grouped Data

$$\text{Median} = l + \frac{(n/2) - c.f}{f} \times h$$

Mode

$$Mode = L + \frac{f_1 + f_0}{2f_1 - f_0 - f_2} \times h$$

$$2\,Mean = 3\,Median - Mode$$

Range

$$Range = highest\,(H) - lowest\,(L)$$

⇒ Defined as the difference b/w largest and lowest observed value in data set.

$$Co\text{-}efficient\ of\ Range = \frac{H-L}{H+L}$$

Quartile Deviation

Used to overcome limitations of range to minimise the influence of outliers.

IQR (Interquartile Range) is calculate to measure dispersion of value in data set b/w Q_1 and Q_3

$$IQR = Q_3 - Q_1$$

$$Quartile\ Deviation\ (QD) = \frac{Q_3 - Q_1}{2}$$

* Less value of QD indicate less variation among the middle 50% of observed values.

(discrete) $Q_1 = \frac{N+1}{4}$ ⇒ find adjacent cf of whose n is answer

Coefficient of QD = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

In Continuous series find Cf in ascending order and calculate for it

If even no. of observations then

eg: $Q_1 = 2.75$ as 2.75 b/w 2 and 3

$Q_1 = 2^{nd}$ value $+ 0.75(3^{rd} - 2^{nd})$

find Q_1 and find suitable cf class, $Q_1 = L + \left(\frac{N/4 - cf}{f}\right) h$ ⇒ cf of previous

Mean Deviation

Covers the limitation of Range and QD as they do not cover the centre most value for dispersion.

Mean Deviation can be from Mean / Median

$D = X - \overline{X}$ or $D = X - \text{Median}$

Individual: $M.D = \frac{\Sigma |D|}{N}$

Discrete: $MD = \frac{\Sigma f |D|}{N}$

Continuous: $M.D = \frac{\Sigma f |D|}{N}$

Coefficient of $MD_{\overline{X}} = \frac{M.D}{\overline{X}}$

Co-efficient of $M.D_M = \frac{M.D}{Median}$

If frequency distribution is symmetrical both $MAD_{\bar{X}}$ and$_M$ are same.

If signs are not ignored the sum of value will be zero in case of $\bar{x}$ and near zero for median

Standard Deviation

Variance $= \sigma^2$

Individual : $\sigma = \sqrt{\frac{\Sigma (X-\bar{X})^2}{n}}$

Discrete: $\sigma = \sqrt{\frac{\Sigma f(X-\bar{X})^2}{N}}$

Continuous: $\sigma = \sqrt{\frac{\Sigma f(X-\bar{X})^2}{N}}$

Combined S.D.

Combined Mean $(\bar{X}) = \frac{n_1 \bar{x}_1 + n_2 x^2}{n_1+n_2}$

$d_1 = \bar{X}_1 - \bar{X}$, $d_2 = \bar{X}_2 - \bar{X}$

Combined S.D $= \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1+n_2}}$

Mathematical Properties : Independent of change of origin but not scale
i.e. No effect on addition / subtraction but change visible for multiplication and division.

Moments

→ Used with reference to frequency distribution

→ describes two characteristics of ~~skewness~~ frequency distribution i.e. Skewness and Kurtosis.

① Moments about Mean (Central Moments)

μ

Individual Series

$$\mu_r = \frac{\Sigma (x-\bar{x})^r}{N}$$

Continuous / Discrete

$$\mu_r = \frac{\Sigma f (x-\bar{x})^r}{N}$$

\# $\mu_1 = 0$ (always)

$\mu_2 =$ variance

if odd moments about mean are 0 then distribution is symmetrical

② Moments about Arbitary Value (Raw Moments)

μ'

Individual

$$\mu'_r = \frac{\Sigma (X-A)^r}{N}$$, A is always given

Continuous / Descrete

$$\mu'_r = \frac{\Sigma f (X-A)^r}{N}$$

~~③ Moments about Origin~~

Relation b/w Central and Raw Moments

as $\mu_1 = 0$ and moments are asked till μ_4 then there will be 3 eq^n

$\mu_1 = 0$

$\leftarrow$ increasing μ' value and power increasing of last term

$$\mu_2 = \mu'_2 - (\mu'_1)^2$$

$$\mu_3 = \mu'_3 - 3\mu'_1\mu'_2 + 2(\mu'_1)^3$$

$$\cancel{\mu_4 = \mu'_4 - 4\mu'_1\mu'_3 + 6(\mu'_2)(\mu'_1)^2 + 4(\mu'_1)^4}$$

$$\mu_4 = \mu'_4 - 4\mu'_1\mu'_3 + 6\mu'_2(\mu'_1)^2 - 3(\mu'_1)^4$$

(3) **Moments about Origin**

ν, when $A = 0$

same as Raw moments

Relation of Moments about origin in moments about mean

$$\nu_1 = A + \mu_1$$

$$\nu_2 = \mu_2 + \nu_1^2$$

$$\nu_3 = \mu_3 + 3\nu_1\nu_2 - 2(\nu_1)^3.$$

$$\nu_4 = \mu_4 + 4\nu_1\nu_3 - 6\nu_2(\nu_1)^2 - 3(\nu_1)^4$$

$\Rightarrow$ Mean $\Rightarrow \nu_1$

$\Rightarrow$ Variance $\Rightarrow \mu_2$

$\Rightarrow$ Skewness $\Rightarrow \mu_3$

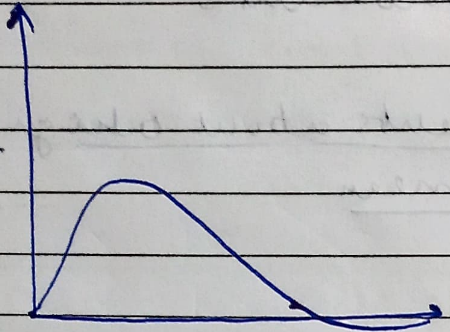
$\Rightarrow$ Kurtosis $\Rightarrow \mu_4$

$\beta_1 = \frac{(\mu_3)^2}{(\mu_2)^3}$ (skewness) $\beta_2 = \frac{\mu_4}{\mu_2^2}$ (Kurtosis)

if $\beta_1 = 0$, symmetrical distribution

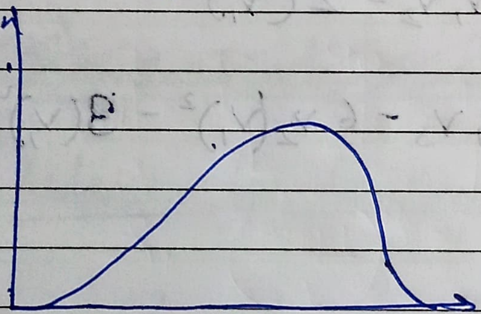
Skewness

⇒ It means lack of symmetry from symmetry of distribution.

⇒ 

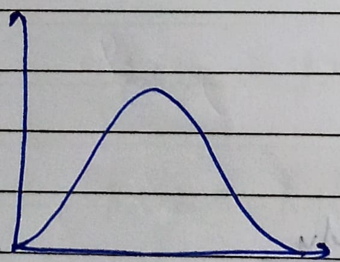
Mean > Median > Mode

Positive symmetry as tail is towards right

⇒ 

Mean < Median < Mode

Negative symmetry as tail is towards left.

⇒ 

Symmetric Mean = Median = Mode

Measures of Skewness

Absolute

* $\boxed{\text{Skewness} = \text{Mean} - \text{Mode}}$

⇒ if +ve then +ve skewness otherwise −ve

$\boxed{\text{Skewness} = Q_3 + Q_1 - 2\,\text{Median}}$

Not used due to 'diff.' units and not comparable

Relative

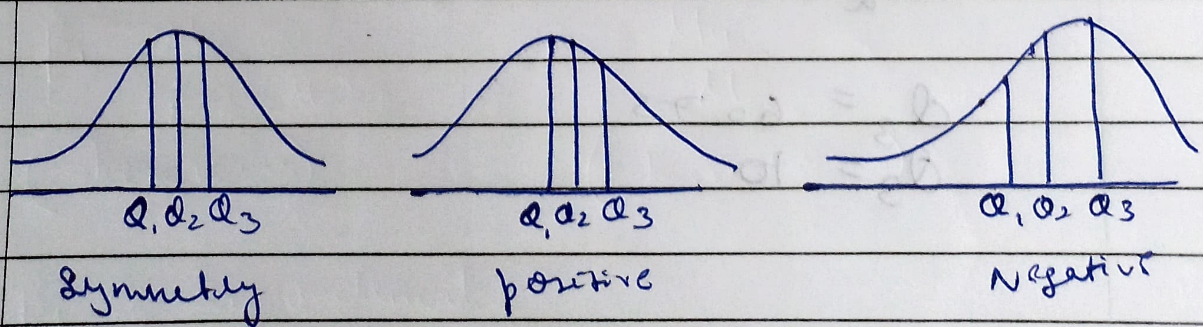
Karl Pearson's Coefficient of Skewness

$$SK_p = \frac{\text{Mean} - \text{Mode}}{S.D}$$

or

$$= \frac{3(\text{Mean} - \text{Median})}{S.D.}$$

Bowley's Coefficient of Skewness



$$SK_B = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

or

$$\frac{Q_3 + Q_1 - 2\,\text{Median}}{Q_3 - Q_1}$$

Discrete Series

eg:

x	10	5	7	11	8
f	15	20	15	18	12

~~x~~	~~f~~	~~cf~~
~~10~~	~~15~~	~~15~~
~~8~~	~~20~~	~~35~~
~~7~~	~~15~~	~~40~~
~~11~~	~~18~~	~~58~~
~~8~~	~~12~~	

x	f	cf
5	20	20
7	15	35
8	12	47
10	15	62
11	18	80

$Q_1 = \left(\frac{80+1}{4}\right)^{th} = \left(\frac{81}{4}\right)^{th} = 20.25$ → check cf for next highest

cf $= 35 \Rightarrow f = 7.$

so, $Q_1 = 7$

$Q_2 = 2 \times 20.25 = 40.50$

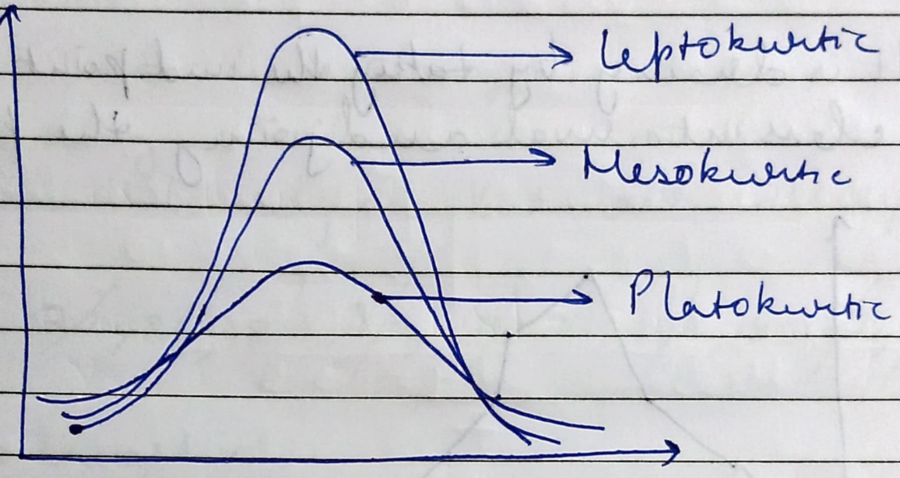
$Q_2 = 8$

$Q_3 = 60.75$

$Q_3 = 10$

Kurtosis

⇒ Tells about peakness of the curve



⇒ Measured using moment

$\beta_2 = \frac{\mu_4}{\mu_2^2}$

$\gamma_2 = \beta_2 - 3$

if $\beta_2 = 3$ or $\gamma_2 = 0$, mesokurtic

if $\beta_2 < 3$, or $\gamma_2 < 0$, platykurtic

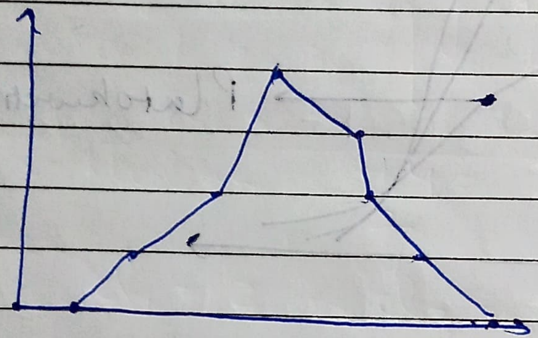
if $\beta_2 > 3$, or $\gamma > 0$, leptokurtic

if not mentioned find β_2

Histogram

Frequency Curve

It is drawing by taking the mid point of class interval and joining the points



Ogive

Plot both the less than and more than ogive where they intersect is the mode on x-axis.

If plotted only one take $N/2$ and draw a line from y-axis intersecting ogive giving mode

Stem and Leaf Plots

⇒ Divide numbers in the range of 10 like 0 to 10 or 100 to 110

⇒ then take the first digit as stem and remaining as leaf

⇒ stem will be only one but leaf will repeat

eg: ~~72~~, ~~85~~, ~~89~~, ~~93~~, ~~88~~, ~~76~~, ~~108~~, ~~115~~, ~~97~~, ~~102~~, ~~113~~

now lowest is 72

so, 70-80 : 72, 76

80-90 : 85, 88, 89

90-100 : 93, 97

100-110 : 102, 108

110-120 : 113, 115

72 76

S	L
7	2, 6
8	5, 8, 9
9	3, 7
10	2, 8
11	3, 5

5-Number Summary

⇒ It consists min, Q_1, Q_2 (median), Q_3, max which is used in Box plot

⇒ arrange in ascending order

eg: 0, 1, 2, 3, 3 | 4, 4, 5, 6, 10

$$\text{Median} = \frac{3+4}{2} = 3.5$$

$Q_1 = 2$, $Q_3 = 5$

Min = 0 , Max = 10

Check for Outlier

$Q_1 - 1.5\,IQR$ & $Q_3 + 1.5\,IQR$
if min and max values lie outside move towards the median and check

Box Plot