Basic Statistics: List of Formulas.

Mean, Median, Mode, S.D.

Mean: ($\bar{X}$)

Notations to remember

n = Total no. of observation or data points.

m = mid pt of interval (in continuous data)

d = deviation (difference betⁿ given series and origin)

A = Assumed mean or origin or middle value

i or h = Interval difference, N = Total of frequency series

or common value in d by which data is divided ie $d = \frac{x-A}{i}$ → common difference

Formulas

	Individual series	Discrete series	continuous series
Direct Method:	$\bar{X} = \frac{\Sigma x}{n}$	$\frac{\Sigma fx}{\Sigma f}$ or $\frac{\Sigma fx}{N}$	$\frac{\Sigma fm}{\Sigma f}$
short cut Method:	$A + \frac{\Sigma d}{n}$; $d = x - A$	$A + \frac{\Sigma fd}{N}$; $d = x - A$	$A + \frac{\Sigma fd}{N}$; $d = m - A$
Step deviation method:	$A + \frac{\Sigma d}{n} \times i$; $d = \frac{x-A}{i}$	$A + \frac{\Sigma fd}{N} \times i$; $d = \frac{x-A}{i}$	$A + \frac{\Sigma fd}{N} \times i$; $d = \frac{m-A}{i}$

Median: Middle value after rearranging data points in ascending or descending order.

ungrouped data: Observation as Median = size of $\frac{n+1}{2}$

if n = no of observation Odd

Median = observation corresponding to average of $\frac{n}{2}^{th}$ observation and $(\frac{n}{2}+1)^{th}$ observation

Median continue

Grouped data:

Steps

1) Arrange data in ↑ or ↓ order
2) Make column of cummulative frequency (c.f.)
3) Find c.f $= \frac{N}{2}$ (equal or just greater than $\frac{N}{2}$)
4) corresponding observation is median

Eg:

x	3	5	6	7	10	12
f	5	4	2	6	6	7
C.F	5	9	11	17	20	$\boxed{30}$ = N

$\frac{N}{2} = 15$

C.F just $\geq \frac{N}{2}$ is 17.

∴ corresponding value of $\boxed{x = 7 = \text{Median}}$

eg

x	10	20	30	40	50	
f	2	3	2	3	1	$N = \Sigma f = 11$
C.f	2	5	7	10	11	

Total values are odd & $\frac{N+1}{2} = \frac{11+1}{2} = 6^{th}$ item

C.F. just $\geq \frac{N+1}{2}^{th}$ ~~item~~ observation

$= \frac{11+1}{2}^{th}$ "

$= 6^{th}$ "

$= 7$

corresponding value of $\boxed{x = 30 = \text{Median}}$

For continuous data:

$$\text{Median} = l + \frac{\frac{N}{2} - C.f.}{f} \times h$$

where l = lower limit of median class

$N = \Sigma f$

C.f. = c.f. of preceeding of median class

f = frequency of median class

h = difference of interval

Mode: observation repeat maximum times.

ungrouped data: eg. 11 12 24 12 13 12 15 18 20 12

Mode = 12 { repeat maximum times)

grouped data: (for continuous data)

$$\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$$

$$\text{Mode} = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$$

where l = lower limit of modal class

h = size of class interval

f_m = frequency of modal class

f_1 = frequency of the class preceding the modal class

f_2 = " " the class succeding the modal class

Steps:

1. Find the maximum class frequency
2. Find class corresponding to this frequency = modal class
3. class interval = upper limit of modal class − lower limit of modal class
4. Identify. f_m, f_1, f_2 & l, h
5. substitute in above formula

Harmonic Mean: H.M.

$$= \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\Sigma\left(\frac{1}{x}\right)}$$

for frequency data H.M. $= \dfrac{f_1 + f_2 + \cdots + f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_n}{x_n}} = \dfrac{\Sigma f}{\Sigma\left(\frac{f}{n}\right)}$

Geometric Mean: (G.M.) $= \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n}$ or Antilog $\dfrac{\Sigma \log x_i}{n}$

or $\Sigma f \log x_i$

Measure of Dispersion: Quartile deviation, Mean deviation, Standard deviation, Moments, Skewness, Kurtosis

Mean deviation $= \dfrac{\sum |x_i - A|}{n}$ or $\dfrac{\sum f_i |x_i - A|}{N}$

(deviation using arbitrary pt)

or

$\dfrac{\sum |d_i|}{n}$ & $\dfrac{\sum f |d|}{N}$ where $d = x - A$

Note: To keep deviation always +ve take absolute value. (Important to remember)

Mean deviation (Using actual Mean ($\bar{x}$)) $= \dfrac{\sum |(x - \bar{x})|}{n}$ & $\dfrac{\sum f |(x - \bar{x})|}{N}$

or $\dfrac{\sum |d|}{n}$ & $\dfrac{\sum f |d|}{N}$ where $d = x - \bar{x}$

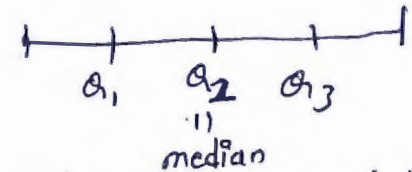
Note: above formula we are using to discuss deviation in the data w.r.t. mean. If we may required to discuss deviation in the data w.r.t median as measure of Central value then formula is as below,

Mean deviation $= \dfrac{\sum |x - \text{median}|}{n}$ for $\dfrac{\sum f |x - \text{median}|}{N}$

or $\dfrac{\sum |d|}{n}$. & $\dfrac{\sum f |d|}{N}$ where $d = x - \text{median}$

Quartile deviation: (Q.D.) It is a one type of measure of dispersion

$$\boxed{Q.D. = \frac{Q_3 - Q_1}{2}}$$



Q_3 = Third quartile (upper quartile) = median of the upper half of the data

Q_1 = ~~lower~~ first quartile (Lower quartile) = median of the lower half of the data

Q_2 = 2^{nd} quartile (median)

Q.D. for ungrouped data:

First arranged data in ↑ order

$Q_1 = \frac{n+1}{4}$ th item (observation)

$Q_2 = \frac{n+1}{2}$ th "

$Q_3 = \frac{3}{4}(n+1)$ th "

Q.D. for Grouped data:

First arranged data in ↑ order

$$Q_r = l_1 + \frac{r\left(\frac{N}{4}\right) - c.f.}{f}(l_2 - l_1)$$

where Q_r = the r^{th} quartile

l_1 = lower limit of quartile class

l_2 = upper " "

f = frequency of "

c.f. = Cummulative freq. of the class preceeding the quartile class eg Find the quartiles of 17, 2, 7, 27, 15, 5, 14, 8, 10, 24, 48, 10, 9, 7, 18, 28

Arrange data in ↑ order:

upper half

lower half [2, 5, 7, 7, 8, 9, 10 | 10], [14, 15, 17, 18, 24, 27, 28, 48]

$n = 16$

$$Q_2 = \frac{n+1}{2} = \frac{16+1}{2} = \frac{17}{2} = 8.5 \approx 9$$

As even data

$$Q_2 = \text{median of data}$$

$$= \frac{10+14}{2} = 12$$

$$Q_1 = \text{Median of lower half of the data}$$

$$= \frac{1}{2}\left[4^{th}\text{ observation} + 5^{th}\text{ observation}\right]$$

$$= \frac{1}{2}\left[7+8\right] = 7.5$$

$$Q_3 = \text{Median of upper half of the data}$$

$$= \frac{1}{2}\left[18+24\right] = 21$$

$$\therefore Q.D = (Q_3 - Q_1)/2 = (21 - 7.5)/2 = 6.75$$

eg

class	f	c.f
0-10	5	5
10-20	3	8
20-30	4	12
30-40	3	15
40-50	3	18
50-60	4	22
60-70	7	29
70-80	9	38
80-90	7	45
90-100	8	$\boxed{53}$ = N

$$Q_r = l_1 + \frac{r\left(\frac{N}{4}\right) - c.f.}{f}(l_2 - l_1)$$

$$\therefore Q_1 = l_1 + \frac{\left(\frac{N}{4}\right) - c.f.}{f}(l_2 - l_1)$$

Here $N = 53$ $\quad \frac{N}{4} = \frac{53}{4} = 13.25$

$\therefore Q_1$ lies in the interval $\underset{l_1}{30}$-40

(substitute all values in above formula & $Q_1 = 34.167$

To find Q_3, $\frac{3N}{4} = 39.75$ $\therefore Q_3$ lies in the interval 80-90

$\boxed{Q_3 = 82.5}$ Follow same steps as above

Standard deviation:

$\sigma = \sqrt{\text{variance}}$

Ungrouped data

Defⁿ:

$$S.d. = \sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{n}} \quad \text{or} \quad \sqrt{\frac{\sum d^2}{n}}$$

$d = x - \bar{x}$
$\bar{x}$ = actual mean

simplified formula : (actual data - No deviation)

$$S.d. = \sqrt{\frac{\sum (x^2 - 2x\bar{x} + \bar{x})^2}{n}}$$

$$= \sqrt{\frac{\sum x^2 - 2\bar{x}\sum x + \bar{x}^2 \sum 1}{n}} \quad \because \bar{x} = \frac{\sum x}{n}$$

$$= \sqrt{\frac{\sum x^2}{n} - 2\frac{\sum x}{n} \cdot \frac{\sum x}{n} + \left(\frac{\sum x}{n}\right)^2 \frac{n}{n}}$$

$$= \sqrt{\frac{\sum x^2}{n} - 2\left(\frac{\sum x}{n}\right)^2 + \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$\therefore \boxed{S.d = \sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}}$$

Used without any deviation:

For grouped data:

$$S.d = \sigma = \sqrt{\frac{\sum f(x-\bar{x})^2}{N}} \quad \text{or} \quad \sqrt{\frac{\sum fd^2}{N}}$$

where
$d = x - \bar{x}$
$\bar{x}$ = actual mea
$N = \sum f$

$$\boxed{S.d = \sigma = \sqrt{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2}}$$

Used without any deviation (actual data)

Deviation
Using assumed mean formula of S.D.

~~step devi~~

$$S.d = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \quad \text{where } d = x - A$$

A = Assumed mean

Step deviation method to find S.D using Assumed mean

$$S.D = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i \qquad d = \frac{x-A}{i}$$

i = difference in deviation

for continuous data: - Same formula
only $d = \frac{m-A}{i}$

Moments: moments gives data analysis about the center value.

Four moments:
1. Expected value or mean (center value)
2. Variance and S.D. (measure spread of values)
3. Skewness (measures asymmetric).
4. Kurtosis (measures the amount in the tails and outliers)

Type of moments
- Raw moment (moments about origin) Here A = 0
- Centered moment
- Moments about Arbitrary point (A)

Important Point → when actual mean is in fraction, moments are first calculated about an arbitrary point and then converted to moments about the actual mean using relⁿ betⁿ moments of actual mean & moments of arbitrary pt.

Moments about Arbitrary pt : (A)

Ungrouped data: $\boxed{d = x - A}$ Grouped data

1^{st} order moment: $M_1 = \frac{\sum_{i=1}^{n}(x_i - A)}{n}$ or $\frac{\sum d}{n}$ $M_1 = \frac{\sum f(x-A)}{\sum f}$ or $\frac{\sum fd}{N}$

2^{nd} order moment: $M_2 = \frac{\sum (x-A)^2}{n}$ $M_2 = \frac{\sum f(x-A)^2}{N}$ or $\frac{\sum fd^2}{N}$

3^{rd} " : $M_3 = \frac{\sum (x-A)^3}{n}$ $M_3 = \frac{\sum f(x-A)^3}{N} = \frac{\sum fd^3}{N}$

4^{th} " : $M_4 = \frac{\sum (x-A)^4}{n}$ $M_4 = \frac{\sum f(x-A)^4}{N} = \frac{\sum fd^4}{N}$

If $d = \frac{x-A}{h}$ (step deviation Method) then

1^{st} order moment $M_1 = \frac{\sum fd}{N} \times h$	3^{rd} order Moment : $M_3 = \frac{\sum fd^3}{N} \times h^3$
2^{nd} order " $M_2 = \frac{\sum fd^2}{N} \times h^2$	4^{th} " $M_4 = \frac{\sum fd^4}{N} \times h^4$

Moments about Origin: Take $A = 0$ in above formula & find all moments

for eg. 2^{nd} order Moment for grouped data

$$M_2 = \frac{\sum f(x-0)^2}{N} = \frac{\sum fx^2}{N}$$

ungroup data $M_2 = \frac{\sum x^2}{n}$

Take note of the symbol.

Moments about Mean.

Ungrouped

$m_1 = \frac{\sum (x - \bar{x})}{n}$ where $\bar{x}$ = actual mean

$m_2 = \frac{\sum (x-\bar{x})^2}{n}$ & $m_4 = \frac{\sum (x-\bar{x})^4}{n}$

$m_3 = \frac{\sum (x-\bar{x})^3}{n}$

Grouped $\boxed{d = x - \bar{x}}$ $\bar{x}$ = actual mean

$m_1 = \frac{\sum f(x-\bar{x})}{N}$ $N = \sum f$

$m_2 = \frac{\sum f(x-\bar{x})^2}{N}$ $m_4 = \frac{\sum f(x-\bar{x})^4}{N} = \frac{\sum fd^4}{N}$

$m_3 = \frac{\sum f(x-\bar{x})^3}{N} = \frac{\sum fd^3}{N}$

Relation betⁿ moments about Mean and moment about arbitrary point:

$m_1 = 0$ (always)

$m_2 = M_2' - M_1'^2$

$m_3 = M_3' - 3M_2'M_1' + 2M_1'^3$

$m_4 = M_4' - 4M_3'M_1' + 6M_2'M_1'^2 - 3M_1'^4$

$m_1 = 0$ (always)

$m_2 = M_2 - M_1^2$

$m_3 = M_3 - 3M_2M_1 + 2M_1^3$

$m_4 = M_4 - 4M_3M_1 + 6M_2M_1^2 - 3M_1^4$

(Just observe pattern and you will be able to remember it)

skewness using moments $\beta_1 = \dfrac{m_3^2}{m_2^3}$ (note we need to calculate moment about actual mean to find skewness)

Remember this as well.

Karl pearson's coefficient of skewness

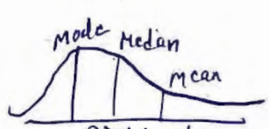
$= \dfrac{\text{Mean} - \text{Mode}}{S.D}$ or $\dfrac{3 * (\text{Mean} - \text{Median})}{S.D}$

In symmetrical distribution: mean = median = mode

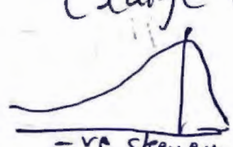
+ve skewed distribution : mode < median < mean
(large tail of high values)

-ve " " : mean < median < mode
(large tail of small values)

Normal (Symmetric) Skewness = 0



Right skew or +ve "



-ve skewed

Kurtosis using moments:

$$\beta_2 = \frac{m_4}{m_2^2}$$ (Here also need to calculate moments using actual mean)

For symmetric distribution: [Normal distribution]

Kurtosis $= 3$

For Lighter tails: Negative kurtosis (flat distribution)

Kurtosis < 3

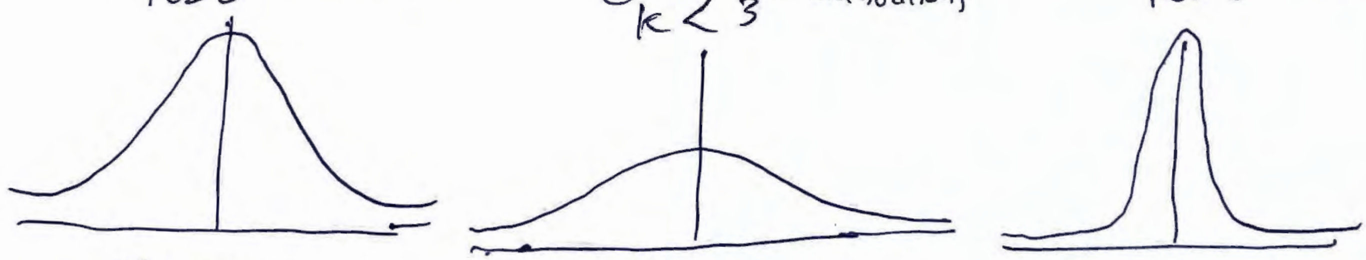
For Heavier tails: +ve kurtosis: (pointed distribution)

Kurtosis > 3

(Normal distribution) $k = 3$

(Uniform distribution) $k < 3$

(Laplace distribution) $k > 3$



Mesokurtic (Symmetric distribution) (medium tailed) (medium outlier frequency)

Platykurtic (Thin tailed) (Low outlier)

Leptokurtic (fat tailed) (High outliers)

$$\beta_2 = \frac{m_4}{m_2^2}$$

$m_4 = 4^{th}$ central moment

$m_2 = 2^{nd}$ central moment (which defines variance)

Also $S.D = \sqrt{vari}$

or $(S.D)^2 = vari = m_2$

$\therefore m_2^2 = (S.D)^4 = \sigma^4$

OR

$$\therefore \boxed{\beta_2 = \frac{m_4}{\sigma^4}}$$