Probability (40%) and statistics (60%)

17/1/23

• measure of central tendency Imp

Imp

① **mean** (scientifically known as arithmatic mean) / Avarage

To calculate mean 1st calculate sum of all of the no. then divide it by the no. of elements.

$$\boxed{mean = \frac{\sum_{i=1}^{n} x_i}{n}}$$

[i = indices]
[n = no. of elements]

Ex mean
[If $\{2, 3, 11, 15, 17, 19\}$
n = 6]
[Here, mean = $\frac{67}{6}$ = 11.16]

mean is the average of the given no. and is calculated by dividing the sum of given no. by the total no. of numbers.

x = Suppose, the data set is =
$\{15, 17, x, 31, 43\}$, given mean = 33

mean = 33

$\Rightarrow \frac{15+17+x+31+43}{5} = 33$

$\Rightarrow x = 59$

or mean = 33

$\Rightarrow \sum_{i=1}^{5} x_i = 33 \times 5$

$\Rightarrow x = 59$

② Imp **median**

It is a measure which is used to divide data sets into 2 parts two data sets. The median is the middle value of a set of numbers.

Suppose, $x = \{15, 2, 7, 9, 13, 18\}$
$= \{2, 7, 9, 13, 15, 18\}$
n = 6 (Even)

(e.g.)

Whenever the no of elements ~~and~~ is even, then

median = $\frac{9+13}{2}$ = 11 (without calculating loc)

Suppose = { 15, 2, 7, 9, 13, 18, 12 }, n = 7 (odd)

~~Whenever the~~

1st arrang the no in sorted order

→ { 2, 7, 9, 12, 13, 15, 18 }

Whenever the no of elements is odd then the median = cancle out from both first and end, continue this process and you can find only one no is left and it is the median.

now for median = 12

Calculation of median by loc of the element

Suppose n = 7 (odd)

$\boxed{\text{location of median} = \frac{n+1}{2}}$

Here = $\frac{7+1}{2}$ = 4 } ~~for on~~ for odd case

the location of the median = 4

the element present in that loc = median = 9

Suppose n = 6 (even)

loc = $\frac{6+1}{2}$ = 3.5

median = $\frac{x[3] + x[4]}{2}$ = $\frac{7+9}{2}$ = 8 } for even case

x[3] = the element present at index = 3

x[4] = the element present at index = 4

[In this case the median ~~for~~ may not be present at data set]

indicating floor: floor of $\lfloor 3.57 \rfloor$ = 3
ceiling of $\lceil 3.57 \rceil$ = 4 indicating ceiling

(3) ~~and~~ mode

The no of occurance of an element of a given data set, max that element is mode.

Suppose, x = { 2, 3, 4, 4, 5, 6, 6, 6, 7, 8, 8 }

	freq
2 →	1
3 →	1
4 →	1
5 →	1
6 →	3
7 →	1
8 →	3

In that case some says that there is no mode and some people say that the data is bimodal and the modes is (6, 8)

if

	freq
2 →	1
6 →	3
8 →	2

∴ mode = 6 (Signal mode)

• A data set can't ~~be~~ have more than 2 modes.

~~Part~~
• Measures of speed/dispersion/scatter/ Scatterness

It indicates how your data is scattered or spread around the mean of your data set.

i.e. ~~how~~ measures of speed summarise the data in a way that shows how scattered the values are and how much they differ from the mean value.

i) ~~Range~~
Stem - and - leaf plot

Stems → leafs, whatever that is after part

Stem	Leaf
3	0, 1, 1, 2, 2, 3, 4, 5, 5, 9
4	0, 1, 1, 2, 3, 3, 5, 8

(If the prefix is common then only used)

(leafs are always in sorted order as well as stem)

part → after part

before part

Legend
3|0 = 30

n = 18 (total no. of both 3 & 4)

~~Ans~~
i) Range = nothing but the difference between the highest and lowest no. of your data set. [the range of a set of a data is the difference between the largest and smallest values, the result of subtracting the sample maximum and minimum]

highest ele = 48 (Last stem and Last leaf)

lowest " = 30 (First " and first ")

∴ Range = 48 - 30 = 18

Range (X) = max(X) - min(X)

(ii) ~~Ans~~ Interquartile Range (IQR) ← Abbreviated as

from stem - and - leaf graph we get this sequence of no. = 30, 31, 31, 32, 32, 33, 34, 35, 35, 39, 40, 41, 41, 42, 43, 43, 45, 48

[If no of elements n = 18, loc = $\frac{18+1}{2}$ = 9.5, we have to take the mean of 9 & 10.]

∴ median = $\frac{35+39}{2}$ = 37 (Q_2) [Q_2 = 2nd quarter]

• median of ~~1st quo~~ lower bound is Q_1 (1st quartile)

• 2nd quartile(Q_2) = median of the whole set = 37.

• median of upper bound is 3rd quartile (Q_3).

Now, we have to divide the data set into 2 halfs.

1st half = {30, ... 35}
2nd " = {39, 40 ... 48}

Let, Lower half (L) = {30, 31, 31, 32, 32, 33, 34, 35, 35, ~~39~~}
bound

Upper bound (U) = {39, 40, 41, 41, 42, 43, 43, 45, 48}

median of L = 32 [∴ median loc = $\frac{9+1}{2}$ = 5] (1st quartile = Q_1)

" " U = 42 [∴ loc = $\frac{9+1}{2}$ = 5] (3rd quartile = Q_3)

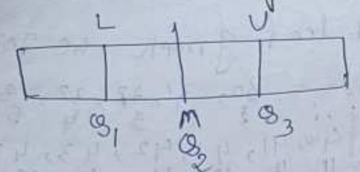
∴ $\boxed{IQR = Q_3 - Q_1}$ = 42 - 32 = 10

[For the case of Q_2 the median may not be present at data set]

20/1/23

• Inter quartile Range (IQR)

(1)



$IQR = Q_3 - Q_1$

Stem	leaf
5	0, 0, 1, 1, 3, 7
6	1, 3, 3, 4, 8

$N = 11$

50, 50, 51, 51, 53 | 57 | 61, 63, 63, 64, 68
1 2 3 4 5 6 7 8 9 10 11

$loc_{Q_2} = \frac{11+1}{2} = 6$

median $(Q_2) = 57$

L = 50, 50, 51, 51, 53 (bound)
1 2 3 4 5

$loc\ Q_1 = \frac{5+1}{2} = 3$; $\therefore Q_1 = 51$

U = 61, 63, 63, 64, 68
1 2 3 4 5

$Q_3 = 63$

$IQR = Q_3 - Q_1 = 63 - 51 = 12$

(2)

5	0, 0, 1, 1, 3, 7, 9
6	1, 3, 3, 4, 8, 9

$N = 13$

50, 50, 51, 51, 53, 57 | 59 | 61, 63, 63, 64, 68, 69
1 2 3 4 5 6 7 8 9 10 11 12 13

$Q_2 = 59$

L = 50, 50, 51, 51, 53, 57
1 2 3 4 5 6

$loc = \frac{6+1}{2} = \frac{7}{2} = 3$ to 4

~~loc median = $\frac{3+4}{2}$~~ 5

$Q_1 = \frac{51+51}{2} = \frac{102}{2} = 51$

U = 61, 63, 63, 64, 68, 69
1 2 3 4 5 6

loc = 3 to 4

$\therefore Q_3 = \frac{x[3] + x[4]}{2} = \frac{63+64}{2} = \frac{127}{2} = 63.5$

$IQR = 63.5 - 51 = 12.5$

• IQR = The IQR describes the middle 50% of values when ordered from lowest to heighest. To find interquartile range (IQR), first find the median of the lower(Q_1) and upper half(Q_3) of the data. Then IQR is the difference between Q3 and Q1

Population and sample

(Representative of entire population)

Population | Sample

i) No. of objects (observation) = Indicated by N | Indicated by n

ii) mean $= \mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$ | $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

iii) Variance -

$\sigma^2 = \dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$

(Indicates how your data is scattering)

$s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$

(unbiased variance formula for sample variance)

[For biased variance formula = $s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$]

[If not mentioned → biased / unbiased then do unbiased use formula]

Standard deviation - $\sigma = \sqrt{\sigma^2}$ $\quad s = \sqrt{s^2}$

Ex: Let the marks obtained by ~~all~~ 13 students of a class (Individual observation)

5	0, 0, 1, 1, 3, 7, 9
6	1, 3, 3, 4, 8, 9

Case of Population,

$N = 13$

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$= \frac{50+50+51+51+53+57+59+61+63+63+64+68+69}{13}$$

$$= \frac{759}{13}$$

$$\simeq 58.38$$

variance,

$$\frac{(50-58.38)^2+(50-58.38)^2+(51-58.38)^2+(51-58.38)^2+(53-58.38)^2+(57-58.38)^2+(59-58.38)^2+(61-58.38)^2+(63-58.38)^2+(63-58.38)^2+(64-58.38)^2+(68-58.38)^2+(69-58.38)^2}{13}$$

$$= \frac{280.2264 + 174.0664 + 112.7844}{13}$$

$\sigma = \sigma^2 \simeq 43.62$

standard deviation

$\sigma = \sqrt{43.62}$

$= 6.604$

$\mu \simeq 58.38$

$\sigma^2 \simeq 43.62$

$\sigma \simeq 6.604$

Sample,

$\bar{x} \simeq 58.38$

$s^2 \simeq 47.25$

$s \simeq 6.87$

Frequency Distribution

They are portrayed as frequency tables or charts. Frequency distributions can show either the actual number of observations falling in each range or the percentage of observations. In the latter instance, the distribution is called a relative frequency distribution.

Ex- No. of car accidents in a particular city over 30 days

Say, No. of in day 1 = 4

2 = 5

3 = 3

4, 5, 3, 4, 5, 6, 7, 4, 5, 5, 4, 3, 5, 4, 5

6, 4, 4, 3, 5, 7, 3, 4, 4, 5, 3, 4, 5, 5, 6

Frequency Distribution Table

No of car accidents (x_i)	Frequency (No. of days) (f_i)
3	5
4	10
5	10
6	3
7	2

→ This table is also known as simple freq distribution table / ungrouped freq distribution table

total freq = $\sum f_i = 30$

Find the mean (?)

for this the mean will be

$$\boxed{\text{weighted mean (arithmatic mean)} = \frac{\sum f_i \cdot x_i}{\sum f_i}}$$

$$\text{mean} = \frac{\sum_i f_i x_i}{\sum f_i} \quad [\sum f_i = N \text{ (must be)}] \quad [\text{No. of observations}]$$

estimated mean

$$= \frac{5\times 3 + 4\times 10 + 5\times 10 + 3\times 6 + 7\times 2}{30}$$

$$= \frac{15 + 40 + 50 + 18 + 14}{30}$$

$$\simeq 4.56$$

- **Population** — In statistics, a population is a set of similar items or events which is of interest for some question or experiment. A statistical population can be a group of existing objects or a hypothetical and potentially infinite group of objects conceived as a generalization from experience.

- **Sample** — A sample refers to a smaller manageable version of a larger group. It is a subset containing the characteristics of a larger population. Sample are used in statistical testing when population sizes are too large for the test to include all possible members or observations.

In probability theory and statistics, **variance** — Variance is the expectation of the squared deviation of a random variable from its population mean or sample mean. Variance is a measure of dispersion, meaning it is a measure of how far a set of numbers is spread out from their average value.

[**Squared deviation** — It is the difference between 2 values, squared. The number that minimizes the sum of squared deviations for a variable is its mean]

mean estimation is a statistical inference problem in which a sample is used to produce a point estimate of the mean of an unknown distribution.

for ungrouped freq distribution —
median, cf - needed, x_i - given
mean, cf = not needed, x_i = given
Q.D, cf = needed, x_i = given
for grouped freq distribution,
median, cf = needed, x_i = not needed
mean, cf = not needed, x_i = needed
Q.D, cf = needed, x_i = not needed

24/1/23

Frequency Distribution

Prob ① Marks in a particular subject (out of 100)
20 students

35, 68, 71, 77, 81, 87, 91, 43, 29, 41,
15, 93, 82, 53, 63, 78, 47, 58, 54, 64

Grouped Frequency data → {15-25, 25-35}

Simple / ungrouped

$$\frac{\sum_{i}^{n} f_i \cdot x_i}{\sum f_i} \qquad [\sum f_i = N]$$

class width $= \frac{15}{20}$ ← No. of students

0 - 15 — 0
15 - 30 — 2
30 - 45 — 3
45 - 60 — 4
60 - 75 — 4
75 - 90 — 5
90 - 100 — 2

Freq Distribution table

Class-Interval	Freq (f_i)	Classmark (x_i) $\boxed{\frac{l+u}{2}}$	$f_i \cdot x_i$
0 - 15	0	7.5	0
15 - 30	2	22.5	45
30 - 45	3	37.5	112.5
45 - 60	4	52.5	210
60 - 75	4	67.5	270
75 - 90	5	82.5	412.5
90 - 100	2	95	190
	$\sum f_i = 20$		$\sum f_i x_i = 1240$

$$\text{mean} = \frac{\sum f_i \cdot x_i}{\sum f_i} = 62$$

② Same problem (1)

class interval	Freq (f_i)	$f_i x_i$	cumulative freq	$f_i x_i$
0 - 20	1	10	10	10
20 - 40	2	60	30	60
40 - 60	6	300	50	300
60 - 80	6	420	70	420
80 - 100	5	450	90	450
	$\sum f_i = 20$	$\sum f_i x_i = 1240$		$\sum f_i x_i =$

$$\therefore \text{mean} = \frac{\sum f_i x_i}{\sum f_i} = \frac{1240}{20} = 62 \text{ (Same ans as Prob)}$$

Cumulative freq (fi)

1

3 (How many no. of students got marks less than 40)

9 [if fi isn't given the fi = diff of cumulative freq i.e., 1, 3-1, 9-3, 15-9, 20-15]

15

20

class marks (x_i)	$f_i x_i$
10	10
30	90
50	450
70	1050
90	1800
	$\Sigma f_i x_i = 3400$

why do we use cumulative frequency here to find out the mean? no use of cum freq, this math is incorrect, we have to use fi to find out mean.

$\therefore$ mean $= \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{3400}{20} = ~~122.75~~$ 170

[when N (even)

median = mean $\left(x_{\left[\frac{n}{2}\right]}, x_{\left[\frac{n}{2}+1\right]}\right)$

when N (odd)

median $= x_{\left[\frac{n+1}{2}\right]}$]

① Car accidents (20) (simply ungrouped freq dist table)

No. of car accidents	No. of days
less than 3	4
4	6
5	7
6	2
7	1
	20

n = 20 (even)

$\frac{n}{2}, \frac{n}{2}+1$

in this case median is located between 10th position, 11th position.

index = 10 observation

not for median for calculating mean for s.g.f.d. x_i is needed

Cumulative freq

4

10

17 (u is under 17)

19

20

from the 10th observation the no. of car acci = 4

" " = 11th = 5

median $= \frac{4+5}{2}$

$\therefore$ median $= \frac{4+5}{2} = 4.5$

Median for grouped Frequency Data

class interval	Freq	C.F (cu. Freq)
0-20	10	10
20-40	15	25
40-60	11	36
60-80	13	49

28.5 falls between 25 & 36

N = 57, N/2 = 57/2 = 28.5

[40-60 interval's value = 36 is just above 28.5]

80-100 8 57

57 manas

[40-60 is the median class]

method 1

$$\therefore \boxed{median = l + \frac{(N/2 - c.f.)}{f} * h}$$

l → Lower limit

cf → Cumulative frequency of the preceding class of the median class.

f → Frequency of the median class

h → class size of the median class

$$\therefore median = 40 + \frac{(28.5 - 25)}{11} * 20$$

= 46.36

31/1/23

method 2

$$\boxed{\frac{median - l_1}{l_2 - l_1} = \frac{N/2 - f_1}{f_2 - f_1}}$$

N → no. of observation

l_1 → lower limit of the median class

l_2 → upper limit of the median class

~~f → freq of the median class~~

cummulative

f_1 → freq ~~of the~~ just below N/2

f_2 → Cummulative freq just above N/2

Ex-①

class Interval	Freq (f_i)	Cummulative freq
10-20	3	3
20-30	4	7
30-40	2	9
40-50	3	12
50-60	15	27 (just above N/2)
60-70	17	44
70-80	16	60

$N = \Sigma f_i = 60$

(no of observation)

$\frac{N}{2} = 30$

60-70 = median class

by method 1

$$median = 60 + \frac{(30 - 27)}{17} * 10$$

$$= 60 + \frac{3 \times 10}{17}$$

61.72

method - ① & method ④ are just same as,

$$\frac{\text{median} - l_1}{l_2 - l_1} = \frac{n/2 - f_1}{f_2 - f_1}$$

f_1 is nothing but the c.f of preceding class

$f_2 - f_1$ is nothing but the freq of the median class $= f$

$l_2 - l_1$ is nothing but the ~~freq~~ size of the class (median clas) $= h$

$$\therefore \frac{\text{median} - l_1}{h} = \frac{n/2 - c.f}{f}$$

$$\Rightarrow \text{median} = l_1 + \left(\frac{n/2 - c.f}{f}\right) * h$$

$$\frac{\text{median} - 40}{60 - 40} = \frac{28.5 - 25}{36 - 25}$$

$\Rightarrow$ median $= 46.36$

By method 2,

$$\frac{\text{median} - 60}{\cancel{60}\ 70-60} = \frac{30-27}{44-27}$$

$$\Rightarrow \frac{\text{median} - 60}{10} = \frac{3}{17}$$

$\Rightarrow$ median = ~~61.64~~ 61.72

Mean for grouped freq distribution data

① $$\boxed{\text{mean} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum f_i}}$$

$f_i \to$ freq of the i class intervals

$x_i \to$ class marks (mid-points) of class intervals.

class mark = mean (L, U)

[L = Lower lim of class interval

U = upper - - -]

(If f_i and x_i are long data then this formula will consume time)

a [arbitary value (generally median of the class-marks)]

then, we calculate $d_i = x_i - a$

② $$\boxed{\text{mean} = a + \frac{\sum_{i=1}^{n} f_i d_i}{\sum_{i=1}^{n} f_i}}$$

[For decimal this formula is more useful]

If d_i is divisable by class size or any common divisor, then

$$u_i = \frac{d_i}{h} = \frac{x_i - a}{h} \quad [h = \ ]$$

③ $$\boxed{\text{mean} = a + h \cdot \left(\frac{\sum f_i u_i}{\sum f_i}\right)}$$

[for decimal this formula is more useful]

Ans ②

Class Interval	Freq (f_i)	class mark (x_i)	$x_i - a = d_i$	$u_i = \frac{d_i}{h}$
10 – 20	13	15	−30	−3
20 – 30	7	25	−20	−2
30 – 40	9	35	−10	−1
40 – 50	23	45	0	0
50 – 60	25	55	10	1
60 – 70	27	65	20	2
70 – 80	13	75	30	3
	$\sum f_i = 117$			

Here, a = 45

Let us consider h = 10 (class interval)

By using formula ③,

$$\text{mean} = 45 + 10\left(\frac{27}{117}\right)$$

$$\approx 49.78$$

Here a is arbitary median, so no need to cal. median by formula to verify arbitrary for grouped f. d. d.

Applicable for simple / ungrouped freq dis. d.

$$\text{mean} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}, \quad \frac{\sum_{i=1}^{n} f_i d_i}{\sum_{i=1}^{n} f_i} + a, \quad \frac{\sum_{i=1}^{n} f_i u_i}{\sum_{i=1}^{n} f_i} \times h + a$$

$$\delta^2 = \frac{1}{n} \sum_{i=1}^{n} x_i - \left(\frac{1}{n} \sum x_i\right)^2$$

[There is no need of freq distribution (f_i) in case of individual d.d]

$$\delta^2 = \frac{1}{n} \sum_{i=1}^{n} f_i - \left(\frac{1}{n} \sum f_i\right)^2$$

[Individual & simple freq isn't same]

$$\delta^2 = \frac{1}{n} \sum_{i=1}^{n} u_i - \left(\frac{1}{n} \sum u_i\right)^2$$

this formulas are applicable for both of ~~individual / gr~~ ungrouped / simple & grouped distribution data / observations by ①,

mean $= \frac{27}{1.17}$

$= 10.23 = \cancel{49.78}$ 49.78

by ②, mean $= 45 + \frac{56}{1.17}$

$= 49.78$

7/2/23

Mode for frequency distribution data

Individual Observations

6, 3, 4, 5, 4, 7, 9   mode = 4

simple / ungrouped freq distribution data

Ex ①

No. of matches	3	2	4	2	1	1
No. of wickets	1	2	3	4	5	6

mode = 3 (wicket)

1, 1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 5, 6

∴ The bowler achiev 3 wickets in 4 matches (total)

mode = 3

Grouped frequency data

Ex ①

marks range	20-40	40-60	60-80	80-100
No. students	5	10	13	6

modal class - It is the class for which the freq is highest

∴ modal class : 60 - 80

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

where, $l \rightarrow$ lower limit of the modal class

f_0, f_1, f_2 $\rightarrow$ frequency of the preceding class of the modal class, modal class, succeeding class of the modal class

h = equal class size accross all the classes.

∴ $l = 60$

$f_1 = 13$

$f_0 = 10$

$f_2 = 6$

$h = 20$

mode $= 60 + \frac{13 - 10}{2 \times 13 - 10 - 6} \times 20$

$= 60 + \frac{3}{10} \times 20$

mode = 66

If class size isn't equal, then we can estimate mode that already aviable in mean & median, using another formula –

$$\boxed{\text{mean} - \text{mode} = 3(\text{mean} - \text{median})}$$

[h is not equal]

Ex – 1

mean = 26.8, median = 27.9

estimate mode?

$26.8 - \text{mode} = 3(27.9 - 26.8)$

$\Rightarrow 26.8 + 3.3 = \text{mode}$

$\therefore \text{mode} = 30.1$

Variance of Individual observes

For the shake of simplicity we are using the population's formula

$$\boxed{\text{variance } (\delta^2) = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}} \leftarrow \text{formula (1)}$$

$$\boxed{\mu = \frac{\sum x_i}{n}}$$

Proof

$\delta^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$

$= \frac{1}{n} \sum_{i=1}^{n} (x_i^2 - 2 \cdot x_i \mu + \mu^2)$

$= \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \sum_{i=1}^{n} 2 \cdot x_i \cdot \mu + \frac{1}{n} \sum_{i=1}^{n} \mu^2$

$= \frac{1}{n} \sum_{i=1}^{n} x_i^2 - 2\mu \cdot \frac{1}{n} \sum x_i + \mu^2$ $\left[\frac{1}{n} \sum_{i=1}^{n} \mu^2 = \frac{1}{n} \cdot n\mu^2 = \mu^2\right]$

$= \frac{1}{n} \sum_{i=1}^{n} x_i^2 - 2\mu \cdot \mu + \mu^2$

$= \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \mu^2$

$= \frac{1}{n} \left[\sum_{i=1}^{n} x_i^2\right] - \left(\frac{1}{n} \sum x_i\right)^2$

$$\therefore \boxed{\delta^2 = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2\right) - \left(\frac{1}{n} \sum x_i\right)^2} \leftarrow \text{formula (2)}$$

(i) $d_i = x_i - a$

a → arbitary assumed median

(ii) $u_i = \frac{x_i - a}{h} = \frac{d_i}{h}$

$$\boxed{\delta^2 = \frac{1}{n} \sum_{i=1}^{n} d_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} d_i\right)^2}$$ [using formula (1)] ← formula (3)

$$\boxed{\delta^2 = \frac{1}{n} \sum_{i=1}^{n} u_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} u_i\right)^2}$$ [using formula (ii)] ← formula (4)

$\frac{1}{n}\sum x_i - 2\mu\cdot\mu + \mu^2$ $= \frac{1}{n}\cdot n\mu^2$

$= \mu^2$]

Prove

$$\delta^2 = \frac{1}{n}\sum_{i=1}^{n}(d_i-\mu_d)^2 = \frac{1}{n}\sum_{i=1}^{n} d_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)^2$$

$\therefore$

$$\delta^2 = \frac{1}{n}\sum_{i=1}^{n}(d_i-\mu_d)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(d_i^2 - 2\mu_d\cdot d_i + \mu_d^2)$$

(i) $$= \frac{1}{n}\sum_{i=1}^{n} d_i^2 - \frac{2\mu_d}{n}\sum_{i=1}^{n} d_i + \frac{1}{n}\sum_{i=1}^{n}\mu_d^2$$

(ii) $$= \frac{1}{n}\sum_{i=1}^{n} d_i - \frac{2\mu_d}{\cancel{n}}\cdot \mu_d\cdot n + \frac{1}{\cancel{n}}\cdot\cancel{n}\mu_d^2 \quad \left[\because \text{In this case, } \mu_d = \frac{1}{n}\sum_{i=1}^{n} d_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} d_i - \mu_d^2$$

δ

$$= \frac{1}{n}\sum_{i=1}^{n} d_i - \left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)^2 \text{ (proved)}$$

Prove

$$\delta^2 = \frac{1}{n}\sum_{i=1}^{n}(u_i-\mu_u)^2 = \frac{1}{n}\sum_{i=1}^{n}u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}u_i\right)^2$$

δ^2

$$\delta^2 = \frac{1}{n}\sum_{i=1}^{n}(u_i^2 + u_u^2 - 2u_i\mu_u)$$

$$= \frac{1}{n}\sum_{i=1}^{n}u_i^2 + \frac{1}{n}\sum_{i=1}^{n}\mu_u^2 - \frac{2\mu_u}{n}\sum_{i=1}^{n}u_i$$

$$= \frac{1}{n}\sum_{i=1}^{n}u_i^2 + \frac{1}{n}\cdot n\mu_u^2 - \frac{2\mu_u}{\cancel{n}}\cdot\cancel{n}\cdot\mu_u \quad \left[\because \text{In this case, } \mu_u = \frac{1}{n}\right.$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\mu\cdot\frac{n}{\ } \quad \left[= \frac{1}{n}\cdot n\mu^2 = \mu^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{h} x_i^2 - 2\mu\cdot\mu + \mu^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 \ \mu^2$$

$$=$$

$$\left[\delta^2\right.$$

(i) $d_i =$

(ii) u

$[C$

$$ib = \frac{1}{n}\sum_{i=1}^{n} u_i^2 - \mu_u^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} u_i\right)^2 \quad \left[\because \mu_u = \frac{1}{n}\sum_{i=1}^{n} u_i\right]$$

(Proved)

how $\boxed{\sum_{i=1}^{n} = n}$?

~~when~~ we know $\sum_{i=1}^{n} f_i = n$, $\rightarrow$ $f_1 + f_2 + f_3 + \cdots + f_n = n$

(1st obs's (2nd ...) ... f (nth
freq)
that is (...
(a part of ...
observation'n')))

part of 'n' + part of 'n' + part of 'n' = n

$$\left[\sum_{i=1}^{n} = \text{nothing means normal count} \right]$$

Similarly, $\sum_{i=1}^{n}$ = 1st obs + 2nd obs + --- + nth obs

= 1 val + 1 val + --- + n val

= n

Ex, = 1, 2, 3, 5, 7, 9
i = 1 2 3 4 5 6

$\therefore$ n = 6 (total no of obs)

$\sum_{i=1}$ = 1 val + 1 val + 1 val + 1 val + 1 val + 1 val = 6 = n variance for single/ungrouped frequency data

$$\delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \mu)^2 \qquad [f_i = \text{frequency}]$$

HW assignment = Prove → $\delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i x_i\right)^2$ [next day]

Prove that $\delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i x_i - \left(\frac{f}{n}\sum_{i=1}^{n} f_i x_i\right)^2$

//

$$\delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \mu)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i (x_i^2 - 2x_i\mu + \mu^2)$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - \frac{1}{n}\sum_{i=1}^{n} f_i \cdot 2x_i \cdot \mu + \frac{1}{n}\sum_{i=1}^{n} f_i\mu^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - 2f_i\mu \cdot \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} f_i\mu^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - 2f_i\mu \cdot \left(\frac{1}{n}\sum x_i\right) + \frac{1}{n}\cdot n f_i \mu^2$$

$$\left[\because \frac{1}{n}\sum_{i=1}^{n} f_i\mu^2 = \frac{1}{n}\cdot n \cdot f_i\mu^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - 2f_i\mu\cdot\mu + f_i\mu^2 \qquad \left[\because \mu = \frac{\sum x_i}{n}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - 2f_i\mu^2 + f_i\mu^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - f_i\mu^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - f_i\left(\frac{1}{n}\sum x_i\right)^2 \qquad \left[\because \mu = \frac{\sum x_i}{n}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - f_i\left(\frac{1}{n}\sum x_i\right)^2$$

$$\boxed{\delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - f_i\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2}$$ (Proved)

(or)

[We know that,

$$\delta^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2 \quad \text{--- (1)}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 \quad (1)$$

If we will multiply f_i with expression (1),

$$\therefore \frac{1}{n}\cdot f_i \sum_{i=1}^{n} (x_i - \mu)^2$$

$$= f_i\left[\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - f_i\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$

$$\therefore \delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - f_i\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2]$$

no need

In similar way,

$$\boxed{\begin{aligned}\delta^2 &= \frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i d_i\right)^2 \\ \delta^2 &= h^2\left[\frac{1}{n}\sum_{i=1}^{n} f_i u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i u_i\right)^2\right]\end{aligned}}$$

$\therefore$ prove $\frac{1}{n}\sum_{i=1}^{n} h^2 \cdot (u_i - \mu_u)^2 \cdot f_i = \frac{1}{n}\sum_{i=1}^{n}(d_i - \mu_d)^2 \cdot f_i$

$u_i = \frac{d_i}{h}$

$\cancel{\frac{1}{n}\sum_{i=1}^{n} u_i}$ $\mu_u = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{1}{n}\sum_{i=1}^{n}\frac{d_i}{h}$

$= \frac{1}{h}\left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)$

$= \frac{1}{h}\cdot \mu_d$

$\therefore \boxed{\mu_u = \frac{1}{h}\cdot \mu_d} \longrightarrow$ Also (proved)

$(u_i - \mu_u) = \frac{d_i}{h} - \frac{\mu_d}{h} = \frac{1}{h}(d_i - \mu_d)$

$\Rightarrow h(u_i - \mu_u) = (d_i - \mu_d)$

$\Rightarrow \sum_{i=1}^{n} h^2 (u_i - \mu_u)^2 = \sum_{i=1}^{n}(d_i - \mu_d)^2$

$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} h^2 (u_i - \mu_u)^2 = \frac{1}{n}\sum_{i=1}^{n}(d_i - \mu_d)^2$

$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} h^2 \cdot f_i (u_i - \mu_u)^2 = \frac{1}{n}\sum_{i=1}^{n} f_i (d_i - \mu_d)^2$ (proved).

$h^2\left[\frac{1}{n}\sum_{i=1}^{n} f_i (u_i - \mu_u)^2\right] = \frac{1}{n}\sum_{i=1}^{n} f_i (d_i - \mu_d)^2$

where, $\mu_u = \frac{1}{n}\sum_{i=1}^{n} u_i$

$= \frac{1}{h}\left[\mu_x - \frac{1}{\cancel{n}}\cdot \cancel{n}\, a\right]$

$\Rightarrow \boxed{\mu_u = \frac{1}{h}[\mu_x - a]}$

Variance for sinde/untrouped frequency data

$\therefore u_i - \mu_u = \frac{x_i - a}{h} - \frac{\mu_x - a}{h}$

$= \frac{1}{h}(x_i - \mu_x)$, for, $i = 1, 2 \ldots n$

$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} h(u_i - \mu_u)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^2$

Also,

$\frac{1}{n}\sum_{i=1}^{n} f_i (u_i - \mu_u)^2 = \frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \mu_x)^2$

$\therefore \frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \mu_x)^2 = \frac{1}{n}\sum_{i=1}^{n} f_i (d_i - \mu_d)^2 = \frac{1}{h}\sum f_i (u_i - \mu_u)^2 \times h^2$

(proved)

we have seen the inter relationship between μ_x & μ_d and μ_x & μ_u. The inter relationship between μ_d & μ_u and $\frac{1}{n}\sum_{i=1}^{n} f_i (d_i - \mu_d)^2 = \frac{1}{n}\sum_{i=1}^{n} h^2 f_i (u_i - \mu_u)^2$ have to seen and prove respectively.

$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - 2f_i\mu^2 + f_i\mu^2$ $\left[\because \mu = \frac{\sum x_i}{n}\right]$

$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - f_i\mu^2$

$\therefore Var(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^2$

$= \frac{1}{n}\sum_{i=1}^{n}\{(x_i - a) - (\mu_x - a)\}^2$

$= \frac{1}{n}\sum_{i=1}^{n}(d_i - \mu_d)$

In case of freq distribution

$Var(x)\,(\sigma^2) = \frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \mu_x)^2$

$= \frac{1}{n}\sum_{i=1}^{n} f_i \{(x_i - a) - (\mu_x - a)\}^2$

$= \frac{1}{n}\sum_{i=1}^{n} f_i (d_i - \mu_d)^2$

Prove that, $\frac{1}{n}\sum_{i=1}^{n} f_i (u_i - \mu_u)^2 h^2 = \frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \mu_x)^2$

$u_i = \frac{d_i}{h} = \frac{x_i - a}{h}$, $i = 1, 2 \ldots n$

$\mu_u = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - a}{h}$

$\Rightarrow \mu_u = \frac{1}{h}\left[\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} a\right]$

where, $\mu_u = \frac{1}{n}\sum_{i=1}^{n} u_i$

$= \frac{1}{h}\left[\mu_x - \frac{1}{n}\cdot n a\right]$

$\Rightarrow \boxed{\mu_u = \frac{1}{h}[\mu_x - a]}$

$\sigma^2 =$

$\sigma^2 = h^2$

Prove that $\frac{1}{n}\sum_{i=1}^{n} f_i (d_i - \mu_d)^2 = \frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \mu_x)^2$

1st way

$\mu_x = \frac{\sum x_i}{n}$

$\mu_d = \frac{\sum d_i}{n}$

$\mu_u = \frac{\sum u_i}{n}$

$d_i = x_i - a, \quad i = 1, 2, \dots, n$

$\mu_d = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)$

$\Rightarrow \mu_d = \frac{1}{n}\left\{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} a\right\}$

$\Rightarrow \mu_d = \frac{1}{n}\left\{\sum_{i=1}^{n} x_i - n.a\right\}$

$\Rightarrow \boxed{\mu_d = \mu_x - a}$ $\quad \left[\because \mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i\right]$

$\therefore d_i - \mu_d = (x_i - a) - (\mu_x - a)$
$= x_i - \mu_x$ for, $i = 1, 2, \dots, n$

$\therefore \sum_{i=1}^{n}(d_i - \mu_d)^2 = \sum_{i=1}^{n}(x_i - \mu_x)^2$

$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}(d_i - \mu_d)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^2$

Also,
$\frac{1}{n}\sum_{i=1}^{n} f_i (d_i - \mu_d)^2 = \frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \mu_x)^2$

2nd way — Also proved $\boxed{\mu_d = \mu_x - a}$

$d_i = x_i - a$

$\therefore \sum_{i=1}^{n} d_i = \sum_{i=1}^{n}(x_i - a) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} a = \sum_{i=1}^{n} x_i - na$

$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - a) = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{\cancel{n}}{\cancel{n}} a$

$\Rightarrow \boxed{\mu_d = \mu_x - a}$, where $\mu_x = \sum_{i=1}^{n} x_i$
$\mu_d = \frac{1}{n}\sum_{i=1}^{n} d_i$

10/2/23

Individual observation:

$$Var(\delta^2) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$

Simple / ungrouped freq distribution

$$\boxed{Var(\delta^2) = \frac{1}{n}\sum_{i=1}^{n} f_i \cdot (x_i-\mu)^2} \text{ — formula (a)}$$

Grouped freq distribution:

$x_i \to$ Class-marks of class-intervals
mid-points of the classes

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

① $d_i = x_i - a$ [a = arbitrary assumed mean] — ①

② $u_i = \frac{x_i - a}{h} = \frac{d_i}{h}$ [h = common division (class-size)] — ②

Similarly

$$\boxed{Var(\delta^2) = \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - \left(\frac{1}{n}\sum f_i x_i\right)^2} \text{ formula (b)}$$

using ① we will get, putting [$d_i + a = x_i$]

$$\boxed{Var(\delta^2) = \frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 - \left(\frac{1}{n}\sum f_i d_i\right)^2} \text{ — formula (d)}$$

$$\boxed{Var(\delta^2) = \frac{1}{n}\sum_{i=1}^{n} f_i (d_i - \mu)^2} \text{ — formula (c)}$$

using ② we will get

$$\boxed{Var(\delta^2) = h^2 \left[\frac{1}{n}\sum_{i=1}^{n} f_i (u_i - \mu)^2\right]} \text{ — formula (e)}$$

Putting $x_i - a = u_i \cdot h$
$\Rightarrow \boxed{x_i = u_i \cdot h + a}$

$$\boxed{Var(\delta^2) = h^2\left[\frac{1}{n}\sum_{i=1}^{n} f_i u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i u_i\right)^2\right]} \text{ — formula (f)}$$

Ex-①

x: 4.5 14.5 24.5 34.5 44.5 54.5 64.5

f: 1 5 12 22 17 9 4

(median has to be found out from ans) $a = 34.5$ — $h = 10$

x_i (ans)	f_i	$d_i = x_i - a$	$u_i = \frac{d_i}{h}$	$f_i d_i^2$	$f_i d_i$	$f_i u_i^2$
4.5	1	-30	-3	900	-30	9
14.5	5	-20	-2	2000	-100	20
24.5	12	-10	-1	1200	-120	12
34.5	22	0	0	0	0	0
44.5	17	10	1	1700	170	17
54.5	9	20	2	3600	180	36
64.5	4	30	3	3600	120	36
	$\sum f_i = 70$					

[a = 34.5, median lies in between observation]

$f_i u_i$
-3
-10
-12
0
17
18
12

$\left[\frac{\sum f_i x_i}{\sum f_i}, \frac{\sum f_i d_i}{\sum f_i}\right]$

formula,

i) $\frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i x_i\right)^2$

ii) $\frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i d_i\right)^2$

iii) $h^2\left[\frac{1}{n}\sum_{i=1}^{n} f_i u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i u_i\right)^2\right]$

are applicable for both ~~individual~~ or simple or ungrouped and grouped freq. distribution data]

$[n = \sum f_i]$

$h^2\left[\frac{1}{n}\sum_{i=1}^{n} f_i u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i u_i\right)^2\right]$

$= 10^2\left[\frac{1}{70} \times 130 - \left(\frac{1}{70} \times 22\right)^2\right]$

$= 10^2[18.57 - 0.098]$

$= 1757.8$

~~$= 10^2[18.57 - 9.87]$~~

~~$= 426.3$~~

$\frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i d_i\right)^2$

$= \frac{1}{70} \times 13{,}000 - \left(\frac{1}{70} \times 220\right)^2$

$= 185.71$ ~~987.7~~ 9.87

$= 175.8$

class interval	class mark (x_i)	f_i	$d_i = x_i - a$	$f_i d_i^2$	$f_i d_i$	$u_i = \frac{d_i}{h}$	$f_i u_i^2$	$f_i u_i$
3-6	4.5	5	-6	180	-30	-2	20	-10
6-9	7.5	7	-3	63	-21	-1	7	-7
9-12	10.5	4	0	0	0	0	0	0
12-15	13.5	3	3	27	9	1	3	3
15-18	16.5	8	6	288	48	2	32	16
		$\sum f_i = 27$		$\sum f_i d_i^2 = 588$	$\sum f_i d_i = 6$		$\sum f_i u_i^2 = 62$	$\sum f_i u_i = 2$

If class interval & f_i are given then we have to find out the rest of the thing

$a = 10.5$

$h = 3$

$\frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i d_i\right)^2$

$\frac{1}{27} \times 588 - \left(\frac{1}{27} \times 6\right)^2$

$\left[\frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 = \frac{1}{27} \times 588\right]$

$= 20.677 - 0.049$

$\left[\frac{1}{n}\sum_{i=1}^{n} f_i d_i = \frac{1}{27} \times 6\right]$

$= 20.6187$

$h^2\left[\frac{1}{n}\sum_{i=1}^{n} f_i u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i u_i\right)^2\right]$

$9\left[\frac{1}{27} \times 62 - \left(\frac{1}{27} \times 2\right)^2\right]$

$= 9\left[2.29 - (5.48 \times 10^{-3})^2\right]$

$= 20.617$

Q: Find the m.D or find the A.m or median??

mean Deviation

Here this mean is anyone of the central tendancy among the Arithmatic mean (A.m) or median.

It is denoted as $= \frac{1}{n}\sum_{i=1}^{n}|x_i - A|$ ↳ either arithmatic mean / median

$= \frac{1}{n}\sum_{i=1}^{n}|d_i|$

$[d_i = x_i - A]$

$$\boxed{\frac{1}{n}\sum_{i=1}^{n}|x_i - A| = \frac{1}{n}\sum_{i=1}^{n}|d_i|}$$

Ex-1

3, 7, 5, 8, 6, 11, 2.

Find out the mean deviation about the mean.

$\mu = A = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{42}{7} = 6$

$\therefore m.D = \frac{1}{7}\sum [|3-6| + |7-6| + |5-6| + |8-6| + |6-6| + |11-6| + |2-6|]$

$= \frac{1}{7}[3 + 1 + 1 + 2 + 0 + 5 + 4]$

$= \frac{16}{7} = 2.28$

Ex-2 median

Sort in ascending order

2, 3, 5, 6, 7, 8, 11 n = 7 (odd)

loc of the median $= \frac{(n+1)}{2} = \frac{8}{2} = 4$

median = 6

Hence mean & median both are = 6

$\therefore m.D = 2.28$

$[\frac{1}{7}|2-6| + \dots + |11-6|]$

If n = even

$\therefore$ mean of $x[\frac{n}{2}] + x[\frac{n+1}{2}]$ = median

Ex-3

3, 7, 5, 8, 6, 11, 2, 17

Find out the mean deviation about mean & median.

About mean,

$\mu = A = \frac{3+7+5+8+6+11+2+17}{8}$

$= 7.375$

$\therefore m.D = \frac{1}{8}\sum |3-7.4| + |7-7.4| + |5-7.4| + |8-7.4| + |6-7.4| + |11-7.4| + |2-7.4| + |17-7.4|$

$= \frac{27.8}{8} \approx 3.475$

About median,

2, 3, 5, 6, 7, 8, 11, 17

for ef

median $= \frac{8+1}{2} = \frac{9}{2} = 4.5$

$\therefore$ median $= \frac{x[4]+x[5]}{2}$

$= \frac{6+7}{2}$

$= \frac{13}{2}$

$= 6.5$

$\therefore M.D = \frac{1}{8}\sum |2-6.5| + |3-6.5| + |5-6.5| +$

$|6-6.5| + |7-6.5| + |8-6.5| +$

$|4-6.5| + |7-6.5|$

$= 3.375$

prove

(1) $\delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i(d_i-\mu_d)^2 = \frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i d_i\right)^2$

we know,

$\delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i(x_i-\mu_x)^2$ — (ii)

$d_i = x_i - a$

$\Rightarrow d_i + a = x_i$ — (1)

Putting (1) in (ii) we get,

$\delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i(d_i+a-\mu_x)^2$

$= \frac{1}{n}\sum_{i=1}^{n} f_i\left[d_i^2 + 2(a-\mu)\cdot d_i + (a-\mu)^2\right]$

$= \frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 + \frac{2(a-\mu)}{n}\sum_{i=1}^{n} f_i d_i + \frac{1}{n}\sum_{i=1}^{n} f_i (a-\mu)^2$

$= \frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 + \frac{2f_i(a-\mu)}{n}\sum_{i=1}^{n}(x_i-a) + \frac{1}{n}\sum_{i=1}^{n} f_i(a-\mu)^2$

$= \frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 + \frac{2f_i(a-\mu)}{n}$

$= \frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 = \frac{2(a-\mu)}{n}$

$= \frac{1}{n}\sum_{i=1}^{n} f_i(d_i + a - \mu_x + a)$

$\left[d_i = x_i - a \Rightarrow x_i - d_i = a \quad \mu_x = \frac{\sum x_i}{n}\right]$

$= \frac{1}{n}\sum_{i=1}^{n} f_i(d_i + a - \mu_d - a)^2$

$= \frac{1}{n}\sum_{i=1}^{n} f_i(d_i-\mu_d)^2$

$\left[\because \mu_d = \sum_{i=1}^{n}\frac{d_i}{n}\right.$ As, when we are using 'd' insted of 'x' then we have to calculate mean using d_i values, not 'x'$\left.\right]$

$= \frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 + \frac{1}{n}\sum f_i \mu_d^2 - \frac{1}{n}\sum_{i=1}^{n} f_i 2 d_i \mu_d$

$= \frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 + \frac{\mu_d^2}{n}\sum_{i=1}^{n} f_i - \frac{2\mu_d}{n}\sum_{i=1}^{n} f_i d_i$

$\left[= \frac{1}{n}\sum_{i=1}^{n}(x_i-a) = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} a = \mu_x - \frac{1}{n}\cdot na = \mu_x - a\right]$

$= \frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 + \frac{\mu_d^2}{n}\cdot n - \frac{2\mu_d}{n}\cdot \mu_d n$

$\left[\because \mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i\right]$

$= \frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 + \mu_d^2 - 2\mu_d^2$

$\left[\because \sum_{i=1}^{n} f_i = n \quad \mu_d = \frac{1}{n}\sum_{i=1}^{n} d_i\right.$ now, in case of freq d_i $\left.\mu_d = \frac{1}{n}\sum_{i=1}^{n} f_i d_i\right]$

$= \frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 - \mu_d^2$

$= \boxed{\frac{1}{n}\sum_{i=1}^{n} f_i d_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i d_i\right)^2}$

(Proved)

(2) Prove that, $\delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i x_i\right)^2$

$$\delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \mu)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 + \frac{1}{n}\sum_{i=1}^{n} f_i \mu^2 - \frac{1}{n}\sum_{i=1}^{n} f_i \cdot 2x_i \mu$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 + \frac{\mu^2}{n}\sum_{i=1}^{n} f_i - \frac{2\mu}{n}\sum_{i=1}^{n} f_i x_i$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 + \frac{\mu^2}{\cancel{n}} \cdot \cancel{n} - \frac{2\mu}{\cancel{n}} \cdot \mu \cdot \cancel{n} \quad \left[\because \mu_x = \frac{1}{n}\sum x_i,\ \text{now } \mu = \frac{1}{n}\sum f_i x_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - \mu^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - \left(\frac{1}{n}\sum f_i x_i\right)^2 \quad \left[\because \sum_{i=1}^{n} f_i = n\right]$$

(Proved)

(3) Prove that, $\delta^2 = \left[\frac{1}{n}\sum_{i=1}^{n} f_i u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i u_i\right)^2\right] h^2$

$$\delta^2 = h^2 \left[\frac{1}{n}\sum_{i=1}^{n} f_i (u_i - \mu)^2\right]$$

$$= h^2 \left[\frac{1}{n}\sum_{i=1}^{n} f_i (u_i^2 + \mu^2 - 2u_i \mu)\right]$$

$$= h^2 \left[\frac{1}{n}\sum_{i=1}^{n} f_i u_i^2 + \frac{1}{n}\sum_{i=1}^{n} f_i \mu^2 - \frac{1}{n}\sum_{i=1}^{n} f_i \cdot 2 u_i \mu\right]$$

$$= h^2 \left[\frac{1}{n}\sum_{i=1}^{n} f_i u_i^2 + \frac{\mu^2}{n}\sum_{i=1}^{n} f_i - \frac{2\mu}{n}\sum_{i=1}^{n} f_i u_i\right]$$

$$= h^2 \left[\frac{1}{n}\sum_{i=1}^{n} f_i u_i^2 + \frac{\mu^2}{n} \cdot n - \frac{2\mu}{n} \cdot \mu \cdot n\right]$$

$\left[\because \sum_{i=1}^{n} f_i = n\right]$ $\left[\because \mu = \frac{1}{n}\sum x_i,\ \text{now } \mu = \frac{1}{n}\sum f_i x_i\right]$

We know, $\delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \mu_x)^2$ — (ii)

and, $u_i = \frac{x_i - a}{h}$ — (i) $\Rightarrow u_i \cdot h + a = x_i$ $\left[\because u_i = \frac{d_i}{h} = \frac{x_i - a}{h}\right]$

Putting (i) in (ii) we get,

$$\delta^2 = \frac{1}{n}\sum_{i=1}^{n} f_i (u_i \cdot h + a - \mu_x)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} f_i (u_i \cdot h + \cancel{a} - \mu_u \cdot h - \cancel{a})^2 \quad \left[\because \mu_u = \frac{1}{h}[\mu_x - a]\right]$$

$$= \frac{1}{n} h^2 \sum_{i=1}^{n} f_i (u_i^2 + \mu_u^2 - 2u_i \mu_u)$$

$$= h^2 \left[\frac{1}{n}\sum_{i=1}^{n} f_i u_i^2 + \frac{1}{n}\sum_{i=1}^{n} f_i \mu_u^2 - \frac{2\mu_u}{n}\sum_{i=1}^{n} f_i u_i\right] \quad \left[\because \sum_{i=1}^{n} f_i = n \text{ in this case } \mu_u = \frac{1}{n}\sum_{i=1}^{n} f_i u_i\right]$$

$$= h^2 \left[\frac{1}{n}\sum_{i=1}^{n} f_i u_i^2 + \frac{1}{\cancel{n}} \cdot \cancel{n}\, \mu_u^2 - \frac{2\mu_u}{\cancel{n}} \cdot \mu_u \cdot \cancel{n}\right]$$

PTO (1)

17/2/23

Mean Deviation

Central tendency (Arithmetic mean, median)

$$\boxed{M \cdot D = \frac{1}{n}\sum_{i=1}^{n} |x_i - A|}$$

$\hookrightarrow$ mean, median $\quad d_i = x_i - A$

$$\boxed{M \cdot D = \frac{1}{n}\sum_{i=1}^{n} |d_i|}$$

Simple / ungrouped freq. Distribution data

$$\boxed{M \cdot D = \frac{1}{n}\sum_{i=1}^{n} f_i |x_i - A|}$$

Grouped / continuous / class - Interval Freq. Distribution Data

$$\boxed{M \cdot D = \frac{1}{n}\sum_{i=1}^{n} f_i |x_i - A|}$$

Here, $x_i \rightarrow$ Class-marks (mid points of the class interval)

$= \frac{1}{n} \cdot h^2 \cdot \sum_{i=1}^{n} f_i \cdot (u_i^2 + \mu_u^2 - 2u_i \cdot \mu_u)$

$= h^2 \left[\frac{1}{n} \sum_{i=1}^{n} f_i \cdot u_i^2 + \frac{1}{n} \sum_{i=1}^{n} f_i \mu_u^2 - \frac{2\mu_u}{n} \sum_{i=1}^{n} f_i \cdot u_i \right]$ $\left[\because \sum_{i=1}^{n} f_i = n \text{ \& in this case } \mu_u = \frac{1}{n} \sum_{i=1}^{n} f_i \cdot u_i \right]$

$= h^2 \left[\frac{1}{n} \sum_{i=1}^{n} f_i u_i^2 + \frac{1}{\cancel{n}} \cdot \cancel{n}\, \mu_u^2 - \frac{2\mu_u}{\cancel{n}} \cdot \mu_u \cdot \cancel{n} \right]$ $\boxed{\text{PTO}}$ 1

Mean Deviation

17/2/23

Cen

$\boxed{\text{PTO}}$

$= h^2 \left[\frac{1}{n} \sum_{i=1}^{n} f_i u_i^2 + \mu_u^2 - 2\mu_u^2 \right]$

[M. D

$= h^2 \left[\frac{1}{n} \sum_{i=1}^{n} f_i u_i^2 - \mu_u^2 \right]$

M. D

$= h^2 \left[\frac{1}{n} \sum_{i=1}^{n} f_i u_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} f_i u_i \right)^2 \right]$

(Proved)

simpl

$\left[M.D = \frac{1}{n} \sum_{i=1}^{n} f_i |x_i - A| \right]$

deviation about the mean

now $\sum_{i=1}^{n} f_i = n$?

let ~~Ob~~ servations = 1, 2, 3 ----- n

taken an ex,

1) observations = 3, 5, 10, 12, 14, 18,
(~~stu~~ marks obtained by stu.)

7, 19, 13, 1 total n = 10 observations

now, ~~C.I~~ ~~Freq~~ (the no. of student getting marks between that C.I)

0 − 5 ⟶ 2 (f_{0-5})

5 − 10 ⟶ 2 (f_{5-10})

10 − 15 ⟶ 4 (f_{10-15})

15 − 20 ⟶ 2 (f_{15-20}).

now, $\Sigma f_i = 10$ = no. of observations

elaborately $\sum_{i=1}^{n} f_i = \sum_{i=1}^{10} f_i = \boxed{f_{(0-5)} + f_{(5-10)} + f_{(10-15)} + f_{(15-20)}}$.

$f_1 + f_2 + f_3 \cdots + f_{10} = 2+2+4+2 = 10 = n$

(proved)

Ex ① Individual observation

(Sms)

3, 7, 5, 8, 6 $n = 5$

1. Sort the data in ascending order

Now

3, 5, 6, 7, 8

mean deviation about the median,

$M.D = \frac{1}{5}\Sigma|$

median = 6 = A

$M \cdot D = \frac{1}{5}\Sigma|3-6| + |5-6| + |6-6| + |7-6| + |8-6|$

$= \frac{1}{5}(3+1+0+1+2)$

$= \frac{1}{5} \times 7$

$= 1.4$

Sms ②

x_i	20	30	40	50
f_i	5	6	3	7

a) mean deviation about the mean,

mean $= \frac{1}{n}\Sigma f_i x_i$ where,

$n = \Sigma f_i$

$n = \Sigma f_i = 21$

mean $= \frac{1}{21}(5\times20 + 6\times30 + 3\times40 + 7\times50)$

$= 35.71$

$M.D = \frac{1}{21}[5\times|20-35.71| + 6\times|30-35.71| + 3|40-35.71| + 7\times|50-35.71|]$

$= \frac{1}{21}[78.55 + 34.26 + 12.87 + 100.03]$

$= 10.74$

b) mean deviation about the median

median of S.F.D

Ob (x_i)	Freq (f_i)	Cumulative freq (c.f_i)
20	5	5
30	6	11
40	3	14
50	7	21

$n = \Sigma f_i = 21$ (odd)

[cf is used to find out the median of grouped or ungrouped freq distribution data]

$\frac{n+1}{2} = \frac{21+1}{2} = 11$

median is located at the

$\frac{n}{2} = 10.5$

n = odd $\therefore x\left[\frac{n+1}{2}\right] = x[11]$

the c.f just above the 10.5 = 11

the corresponding obj = the value of the median.

median = 30

$$m.D = \frac{1}{21}\left[5\times|20-30| + 6\times|30-30| + 3\times|40-30| + 7\times|50-30|\right]$$

$$= \frac{1}{21}\left[50 + 30 + 140\right]$$

$$= 10.5 = 10.47$$

mean deviation for grouped freq. Distribution data

Ex - 1

Class-Interval (In m given)	Class marks (x_i)	Freq (In m given) (f_i)	cfi
10 - 20	15	3	3
20 - 30	25	7	10
30 - 40	35	5	15
40 - 50	45	6	21
	~~55~~		
	~~65~~	$\Sigma f_i = n = 21$	

Find out the mean deviation about a) mean b) median.

$$\text{mean} = \frac{1}{n}\Sigma f_i x_i$$

$$\text{mean} = \frac{1}{21}(3\times15 + 7\times25 + 5\times35 + 6\times45)$$

$$= 31.67$$

$$M.D = \frac{1}{21}(3\times|15-31.67| + 7\times|25-31.67| + 5\times|35-31.67| + 6\times|45-31.67|)$$

median $= \frac{n}{2} = 10.5$

the median =

$$= \frac{1}{21}(50.01 + 46.69 + 16.65 + 79.98)$$

$$= \cancel{6.349}\ 9.20 \simeq 9.67$$

$$\cancel{\simeq 6.35}$$

$$\text{median} = l + \frac{\left(\frac{n}{2} - cf\right)}{f}\times h$$

$$= 30 + \frac{(10.5 - 10)}{5}\times 10$$

$$= 31$$

$$M.D = \frac{1}{21}\left[3\times|15-31| + 7\times|25-31| + 5\times|35-31| + 6\times|45-31|\right]$$

$$= \frac{1}{21}(48 + 42 + 20 + 84)$$

$$= 9.23$$

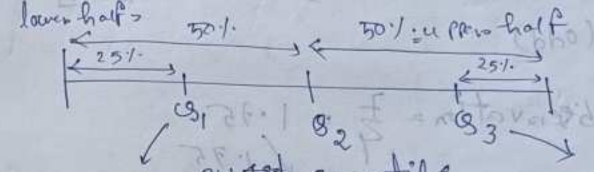
② Exceptional case

Class interval	Class-mark (x_i)	Freq (f_i)	cf_i
10-2			

No math

Quartile

If we divide our data set into 4 equal parts all the parts are known as quartile. The values which split that dataset into 4 equal parts are known as quartiles.



1st quartile / lower quartile

2nd quartile / middle quartile

upper quartile / 3rd quartile

Quartile : - 3 equal parts
Septiles : - 7 equal parts
Octiles : - 8 equal parts
Deciles : - 10 equal parts
Percentiles : - 100 equal parts

$$\boxed{IQR = Q_3 - Q_1}$$

$$\boxed{\text{Quartile Deviation (Q.D.)} = \frac{1}{2}(Q_3 - Q_1)}$$

Quantiles for individual observations:

1) Sort the data in ascending order.

2) $\boxed{\frac{iN+1}{4}}$ th observation value [for finding the quantiles]

$i = 1, 2, 3$ for Q_1, Q_2, Q_3 respectively

[$N = n$ same notation]

Ex-①

10, 15, 25, 31, 37, 45, 55.

$N = 7$ (odd)

$\frac{n}{4}$ th observation $= \frac{7}{4} = 1.75$

1.75 = 1 + 0.75

$= 10 + \frac{3}{4}(15 - 10)$ 1st observation $\frac{3}{4}$ of (2nd obs − 1st obs)

$= 10 + \frac{3}{4} \times 5$

$= 10 + \frac{15}{4}$

$Q_1 = 13.75$

$\frac{2N}{4}$ th observation $Q_2 = \frac{2N}{4}$

$Q_2 = \frac{N}{2} = \frac{7}{2} = 3.5$

$= 3^{rd} + 0.5(4^{th} - 3^{rd})$

$= 25 + 0.5(31 - 25)$

$= 28$

$Q_3 = \frac{3N}{4} = \frac{3 \times 7}{4} = \frac{21}{4} = 5.25$

$5^{th} + 0.25(6^{th} - 5^{th})$

$= 37 + 0.25(45 - 37)$

$= 39$

$IQR = 39 - 13.75$

$= 25.25$

$Q.D = \frac{1}{2}(Q_3 - Q_1)$

$= 12.625$ → quantile deviation

21/2/23 // Not correct

~~How to find out the median quantile for~~

1) Findout quantile for individual observation:

whenever,

$Q_2 \Rightarrow i = 2$

$x\left[\frac{in+1}{4}\right]$ n → odd

Arithmetic mean $\left(x\left[\frac{iN}{4}\right], x\left[\frac{iN}{4}+1\right]\right)$ → n = even

Q_1; $i = 1$	Pos. of quantile $= \frac{in+1}{4}$, $i = 1$
Q_2: Pos of quantile	$\frac{in+1}{4}$ → n → odd, $i = 2$; A.M. $\left(\frac{in}{4}, \frac{in}{4}+1\right)$ → even (nth obs)

$\boxed{Q_3: \quad i=3, \quad \frac{in+1}{4} = \text{pos of quantile}}$

[no need of (f_i)]
[Individual obs]

Ex-① 3rd

n → odd

1, 5, 2, 4, 11, 8, 10

Sort: 1, 2, 4, 5, 8, 10, 11

$Q_1, \; i=1, \; x\left[\frac{1\times 7+1}{4}\right]$ = value of quantile

$\Rightarrow Q_1 = x\left[\frac{7+1}{4}\right]$

$= x[2]$

$\boxed{Q_1 = 2}$

$Q_2: \; i=2, \; x\left[\frac{2\times 7+1}{4}\right] = Q_2$

$\Rightarrow Q_2 = x\left[\frac{15}{4}\right]$

$x[3.75]$

∴ value of 2nd quantile, $Q_2 = x[3] + 0.75(x[4] - x[3])$

$= 4 + 0.75(5-4)$

$\boxed{Q_2 = 4 + 0.75 = 4.75}$

$Q_3: \; i=3 \quad Q_3 = x\left[\frac{3\times 7+1}{4}\right]$

$= x[5.5]$

$\therefore Q_3 = x[5] + 0.50[x[6] - x[5]]$

$= 8 + 0.5(10 - 8)$

$\boxed{Q_3 = 9}$

$IQR = Q_3 - Q_1$

$= 9-2$

$= 7$

$Q.D = \frac{7}{2} = 3.5$

Ex-② 3rd

n → even

1, 5, 2, 4, 11, 8, 10, 12

Sort, 1, 2, 4, 5, 8, 10, 11, 12

$Q_1, \quad Q_1 = x\left[\frac{1\times 8+1}{4}\right]$

$= x[2.25]$

$\therefore Q_1 = 2 + 0.25(4-2) = 2 + 0.5 = 2.5$

$Q_2, \quad x\left[\frac{x[4] + x[5]}{2}\right]$

$= x\left[\frac{5+8}{2}\right]$

$= x[6.5]$

$Q_2 = 10 + 0.5(11-10)$

$= 10.5$

$Q_3, \quad Q_3 = x\left[\frac{24+1}{4}\right]$

$= x[6.25]$

$Q_3 = 10 + 0.25(11-10)$

$= 10.25$

$IQR = 10.25 - 2.5$

$= 7.75$

$Q.D = 3.875$

How to findout quantiles for freq distribution data

1st Quantiles for simple or ungrouped or discrete freq distribution data

[In this ungrouped freq distribution data we will always be provided with x_i (obs), f_i (freq)]

Here $\boxed{\frac{in}{4}\text{th}, \ i = 1, 2, 3 \ldots \quad Q_1, Q_2, Q_3 \text{ respectively}}$ —(I)

x_i	f_i	c.f
2	5	5
→ 4	7	12 ←
→ 5	6	18 ←
→ 7	5	23 ←
8	4	27
	$n = \sum f_i = 27$	

Q_1, $\frac{in}{4}$, $i = 1$

$\frac{1 \times 27}{4}$

$= 6.75$

c.f just above 6.75 = 12

Corresponding obs = $Q_1 = 4$

Q_2, $\frac{2 \times 27}{4}$

$= 13.5$

c.f just above 13.5 = 18

$\therefore Q_2 = 5$

Q_3, $\frac{3 \times 27}{4} = 20.25$

$Q_3 = 7$

$IQR = 7 - 4$

$= 3$

$Q.D = \frac{3}{2} = 1.5$

2nd Quantiles for grouped/Continuous freq distribution data

[In — — — x_i (obs), f_i (freq)]

Here, is added with (I)

$$\boxed{l + \left(\frac{\frac{in}{4} - c.f}{f}\right) \times h} \quad \text{—(II)}$$

$i = 1, 2, 3$ respectively for Q_1, Q_2, Q_3

Ex - ①

C.I	C.m (x_i)	f_i	cf
10-20	15	6	6
20-30	25	6	12
30-40	35	8	20
40-50	45	6	26
		$n = \sum f_i = 26$	

Q_1, $30 + \left(\frac{\cancel{1 \times 26} - 12}{8}\right) \times 10$

$= 47.5$

~~Q_2; $30 + \left(\frac{1 \times 26}{4} - 12\right)$~~

$i = 1$, $l + \left(\frac{i \times n}{4} - c.f\right) \times \frac{h}{f}$

$l + \left(\frac{n}{4} - cf\right) \times l$

$\frac{1 \times 26}{4} = 6.5$

just above 6.5 = 7

~~$Q_1 = 15$~~ median class = 10-20

Q_2, $\frac{2 \times 26}{4} = 13$

$i = 2$

$= 10 + \left(\frac{6.5 - 7}{7}\right) \times 10$

$Q_1 = 20 + \left(\frac{6.5 - 6}{5}\right) \times 10$

$= 21$

Q_2, $i = 2$, $\frac{2 \times 26}{4} = 13$

just above 13 = 20

median class = 30-40

$\therefore Q_2 = 30 + \left(\frac{13 - 12}{8}\right) \times 10$

$=$ ~~40.38.75~~ 31.25

Q_3, $i = 3$, $\frac{3 \times 26}{4} = 19.5$

just above 19.8 = 20

median class = 30-40 ⟶ fresh

P.T.O. →

$\therefore Q_3 = 30 + \left(\frac{19.5 - 12}{8}\right) \times 10$

$= 39.375$

$\therefore$ I.Q.R $= 39.375 - 21$

$= 18.375$

Q.D $= \frac{1}{2} \times 18.375$

$= 9.1875$

Ex-①

C.I.	C.m	f_i	c.f (cumulative freq)
10-20	15	6	6
20-30	25	6	12 ←
30-40	35	8	20 ← ←
40-50	45	6	26
		$n = \Sigma f_i = 26$	

Q_1, $i = 1$, $\frac{1 \times n}{4} = \frac{1 \times 26}{4} = 6.5$

c.f just above ~~value of~~ 6.5 = 12

median class = corresponding C.I. (class interval)

= 20-30

$\therefore l + \left(\frac{\frac{in}{4} - c.f}{f}\right) * h$ $h = 10$, c.f = the c.f of preceding class

$= 20 + \left(\frac{6.5 - 6}{6}\right) * 10$

$= 20.83$

$\therefore Q_1 = 20.83$

Q_2, $i = 2$, $\frac{2 \times 26}{4} = 13$

c.f just above 13 = 20

the corresponding C.I = 30-40

$\therefore l + \left(\frac{\frac{in}{4} - c.f}{f}\right) * h$ $i = 2$

$= 30 + \left(\frac{13 - 12}{8}\right) * 10$

$= 31.25$

$\therefore Q_2 = 31.25$

Q_3, $i = 3$, $\frac{3 \times 26}{4} = 19.5$

c.f just above 19.5 = 20

the corresponding C.I = 30-40

$\therefore l + \left(\frac{\frac{in}{4} - c.f}{f}\right) * h$ $i = 3$

$= 30 + \left(\frac{19.5 - 12}{8}\right) * 10$

$= 39.375$

$Q_3 = 39.375$

$\therefore IQR = Q_3 - Q_1 = 39.375 - 20.83 = 18.545$

$Q.D = \frac{1}{2} \times IQR = \frac{1}{2} \times 18.545 = 9.3$

24/2/23

Geometric mean (G.m)

For individual observation

set of observation - $x_1, x_2, \ldots, x_n$ of set of obs of

G.m = nth root of the product of the observations.

$$\boxed{G = (x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}}} \quad \text{---} (1)$$

$$\therefore \log G = \frac{1}{n} \log(x_1 \cdot x_2 \cdots x_n)$$
$$= \frac{1}{n}(\log x_1 + \log x_2 \cdots + \log x_n)$$

$$\boxed{\log G = \frac{1}{n}\sum_{i=1}^{n} \log x_i} \quad \text{---} (2)$$

$$\therefore \boxed{G = \text{Antilog}\left(\frac{1}{n}\sum_{i=1}^{n}\log x_i\right)} \quad \text{---} (3)$$

[Antilog > suppose $\log G = \frac{1}{n}\sum_{i=1}^{n}\log x_i = 2$

$G = \text{Antilog}(2)$

Antilog = 10 (as base=10) $= 10^2$]

If we have freq data (f_i) along with some obs(x_i)

Simple / ungrouped / discrete freq distribution data [$\sum f_i = N$]

$$\boxed{G = \left(x_1^{f_1} \cdot x_2^{f_2} \cdots x_n^{f_n}\right)^{\frac{1}{N}}} \quad \text{---} (1)$$

$$\log G = \frac{1}{N}\log\left(x_1^{f_1} \cdot x_2^{f_2} \cdots x_n^{f_n}\right)$$
$$= \frac{1}{N}\left\{\log(x_1^{f_1}) + \log(x_2^{f_2}) + \cdots + \log(x_n^{f_n})\right\}$$
$$= \frac{1}{N}\left[f_1 \log x_1 + f_2 \log f_2 + \cdots + f_n \log x_n\right]$$

$$\boxed{\log G = \frac{1}{N}\sum_{i=1}^{N} f_i \log x_i} \quad \text{---} (2)$$

$$\boxed{G = \text{Antilog}\left(\frac{1}{N}\sum_{i=1}^{N} f_i \log x_i\right)} \quad \text{---} (3)$$

Grouped / continuous freq distribution data

$$\boxed{G = \left(x_1^{f_1} \cdot x_2^{f_2} \cdots x_n^{f_n}\right)^{\frac{1}{N}}} \quad \text{---} (1)$$

$$\boxed{\log G = \frac{1}{N}\sum_{i=1}^{N} f_i \cdot \log x_i} \quad \text{---} (2)$$

$$\boxed{G = \text{Antilog}\left(\frac{1}{N}\sum_{i=1}^{N} f_i \log x_i\right)} \quad \text{---} (3)$$

[Here, everything will be same but x_i will indicate the class-marks (mid points)]

~~Uses~~ **uses of G.m** - In economical discussion it is useful ~~more~~ and efficient more than A.m.

Ex-(1) suppose we have to find out the General index and group of indeces in case of economics (i.e. G.m)

should be vertical

Group	A	B	C	D	E	F	Total
Group Index (x)	118	120	97	107	112	93	
weight (f)	4	1	2	6	5	2	$\sum f_i = N = 20$

[Have to find out general and avg index of all groups]

$$\boxed{\log G = \frac{1}{N}\sum_{i=1}^{N} f_i \cdot \log x_i} \quad \text{— (2)}$$

[but x_i will indicate the class-marks] (mid points)

$$\boxed{G = \text{Antilog}\left(\frac{1}{N}\sum_{i=1}^{N} f_i \log x_i\right)} \quad \text{— (3)}$$

In calculator $\frac{1}{N}\sum_{i=1}^{n} f_i \cdot \log x_i$ → ~~not~~ giving correct ans → so can do this for shortcut

have to find out general and avg index of cou[…] grou[…]

					20
				2	
			5		
		6			
		2			

$$G = \left(x_1^{f_1} \cdot x_2^{f_2} \cdots\right)$$

$$\log G = \frac{1}{N}\sum_{i=1}^{N} f_i \cdot \log x_i \quad \text{— (2)}$$

$$\text{Antilog}\left(\frac{1}{N}\sum_{i=1}^{N} f_i \cdot \log x_i\right) \quad \text{— (3)}$$

[Here, everything will be same but x_i will indicate the class-marks (mid points)]

Variable / Group	Group Index (x_i)	weight (f_i)	$\log(x_i)$	$f_i \cdot \log x_i$
A	118	4	2.072	8.288
B	120	1	2.079	2.079
C	97	2	1.986	3.972
D	107	6	2.029	12.174
E	111	5	2.045	10.225
F	93	2	1.968	3.936
Total		$N = \Sigma f_i = 20$		$\Sigma f_i \cdot \log x_i = 40.68$

the table should be like this

$\log x_n)$

(2)

(3)

along with so

distribution data $[\Sigma f_i =$

$+ \cdots + \log(x_n^{f_n})]$

Group Index (x)	118	120					$\Sigma f_i = N = 20$
weight		1	2	6	5	2	

have to find out general and avg ind of gr

1.73
In calculator
= 40.68

$$\boxed{\begin{aligned} N &= \Sigma f_i = 20 \\ \log G &= \frac{1}{N}\sum_{i=1}^{N} f_i \cdot \log x_i \end{aligned}} \rightarrow \text{2 main formula}$$

$\log x_i$	2.0720	2.079	1.986	2.029	2.045	1.968	
$f_i \cdot \log x_i$	8.288	2.079	3.972	12.174	10.225	3.936	40.68

$\log G = \frac{1}{N}\sum_{i=1}^{N} f_i \log x_i$

$= \frac{1}{20} \times 40.68$

$= 2.0337$

$G = \text{Antilog}(2.0337)$

$= 108.068 \simeq 108.145$

Ex-

(2) The Geometric mean of 10 observations on a certain variable was calculated to be 16.2. It was later discovered that one of the observation was wrongly recorded as 12.9 when infact it was 21.9. Apply the appropriate correction & calculate the correct geometric mean.

Calculated G.M = 16.2

the true Obs = 21.9

the wrong Obs = 12.9

$G = (x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}}$ [nth observation]

$\boxed{x_1 \cdot x_2 \cdots x_n = G^n}$ (4) use this particular formula for this

$G = 16.2$ [G = calculated mean]

$\Rightarrow G^{10} = (16.2)^{10}$ [G_1 = correct mean]

$G_1^{10} = \frac{(16.2)^{10}}{12.9} \times 21.9$

$\Rightarrow 10 \log G_1 = 10 \log 16.2 + \log(21.9) - \log(12.9)$

$= 12.09 + 1.34 - 1.11$

$= 12.32$

$\Rightarrow G_1 = 10^{\frac{12.32}{10}} = 10^{1.232}$

$= 17.06$

Corrected G.M = 17.06

(or) $G_1^{10} = \frac{(16.2)^{10}}{12.9} \times 21.9$

$G_1^{10} = 2.11 \times 10^{12}$

$\Rightarrow 10 \times \log G_1 = \log(2.11 \times 10^{12})$

$\Rightarrow G_1 = 17.08 \simeq 17.06$

[As we have to find out the correct mean only not the correct mean of both Obs]

$\Rightarrow \log \log G_1 = \log \log 16.2 + \log(1.1.9) - \log($

Logic

Calculated G.M = 16.2,

wrong obs = 12.9

true/actual obs = 21.9

$\Rightarrow$ G = calculated G.M = 16.2

Say G' = corrected G.M

We know, $\boxed{G^n = x_1 \cdot x_2 \ldots x_n}$

Here given ~~(G)~~ 10, we get the wrong

Correct G.M at 10th obs, $\therefore (G)^{10} = (1.62)^{10}$

~~we get~~ For obs 12.9 we get $(1.62)^{10}$

" " 1 " $(1.62)^{10}$ / 12.9

" " 21.9 .. $\frac{(1.62)^{10}}{12.9} \times 21.9 = G_1^{10}$

(or) $G_1^{10} =$

[As we have

28/2/23

Geometric mean

Ex-③ The weighted G.M of 4 no. 8, 25, 17, 30 is 15.3, if the weights of the 1st 3 no. are 5, 3, 4 respectively. Find the weight of the 4th no.

weight = freq

x_i	f_i	$\log x_i$	$f_i \cdot \log x_i$
8	5	0.90	4.51
25	3	1.39	4.19
17	4	1.23	4.92
30	w	1.47	$1.47w$
	$N = \sum f_i = 12 + w$		$\sum f_i \log x_i = 13.62 + 1.47w$

$\Rightarrow \frac{13.62 + 1.47w}{12+w} = \log(15.3)$

given, $\therefore G = 1.53$, $\log G = \log(15.3)$

$\Rightarrow 13.62 + 1.47w = 1.18(12+w)$

$\Rightarrow 1.47w - 1.18w = 0.54$

$\Rightarrow 0.29w = 0.54$

$\Rightarrow w = 1.862 \simeq 2$

(the weight will be in integer)

$\Rightarrow 1.18470 - 1.477w$

$= 13.1308 -$

Ex-④ The production of a particular item increased from 99 million to 176 million from 2015 to 2022. Assuming that the production increased at a constant annual rate, find the avg annual rate of increase.

P (product value)   i (rate of increment)   n (no of years)

Here, $\boxed{A = P(1+i)^n}$ [P = initial product value, A = final product value]

$P = 99 \times 10^6$, $A = 176 \times 10^6$, $n = 2015 - 2022 = 7$ yrs

$\Rightarrow \left(\frac{176 \times 10^6}{99 \times 10^6}\right)^{\frac{1}{7}} - 1 = i$

interest, simple interest, compound interest, rate of interest = this formula

$\Rightarrow i = 0.08566$

(or) (process) taking log of both sides,

$\Rightarrow \log 176 = \log 99 + 7\log(1+i)$

$\Rightarrow 0.2498 = 7\log(1+i)$

$\Rightarrow i = 0.08566$

$= 8.56\%$

$\approx 8.6\%$ Ans

main point increase

[Here the rate of interest is same, but the rate of production of a particular product isn't constant, it's increasing, so we have to apply G.M not A.M or H.M] ~~[Am just reverse]~~ (kinda same with A.m but little diff. from A.m)

Ex-5

The sum of money was invested for 4 years at 2% per annum and then accumulated sum was invested for 3 years at 3% per annum. What is the avg compound interest rate?

Here,

4 yrs → 2%
3 yrs → 3%

A → Final value
P → initial value

$$\boxed{A = P(1+i_1)^{n_1}(1+i_2)^{n_2}} \quad \text{(i)}$$

where,
$n_1 = 4$
$i_1 = \frac{2}{100}$
$n_2 = 3$
$i_2 = \frac{3}{100}$

$n = n_1 + n_2$ yrs

Now,

$A = P(i+1)^7$ (ii)

Equating (i) & (ii)

$\Rightarrow P(i+1)^7 = P(1+\frac{2}{100})^4 \cdot (1+\frac{3}{100})^3$

$\Rightarrow P(i+1)^7 = (1+0.02)^4(1+0.03)^3$

$\Rightarrow i = 0.02424$ or 2.42% Ans

[Here the rate of interest is given, but the product of the particular amount isn't constant, so we have to apply this formula]

$\boxed{\text{avg rate of } i, \text{ c.i. etc} = \text{this formula}}$

3/3/23

Harmonic mean (H.m)

It is reciprocal of the A.m. of the reciprocal of the observations.

Suppose, $x_1, x_2, \ldots x_n$

$$\boxed{H \cdot m = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}}$$

If x_i & f_i are there, where, $N = \sum_{i=1}^{n} f_i$

x_i / f_i

$$H \cdot m = \frac{1}{\frac{1}{N}\sum_{i=1}^{n}\frac{f_i}{x_i}}$$

$$\boxed{H \cdot m = \frac{N}{\sum_{i=1}^{n}\frac{f_i}{x_i}}}$$

← for simple / ungrouped / discrete freq distribution

This H.m is used (it's uses are very rare) for fractions such as rates or multiples (when the observations are expressed of rates, speeds, prices etc).

For grouped / continuous freq distribution

$$\boxed{H \cdot m = \frac{N}{\sum_{i=1}^{n}\frac{f_i}{x_i}}}$$

where, x_i = class marks (mid point of class intervals)

It is reci...
of the operation observations

suppose, $x_1, x_2, \ldots x_n$

$$\boxed{H.m = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}}$$

If x_i & f_i are there, where, $N = \sum_{i=1}^{n} f_i$

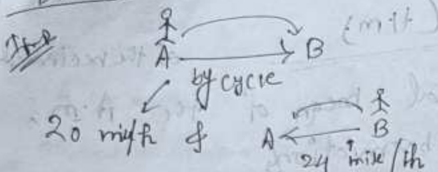
$x_i \mid f_i$

$$H.m = \frac{1}{\frac{1}{N}\sum^{n} \frac{f_i}{\ldots}}$$

Harmonic mean = The H.m. is defined as "A value that is the reciprocal of the mean of the reciprocals of a set of numbers or observations.

Ex-1

mile/h =



Find out the avg speed of the person.

[Here the distance is same so we can't use A.m. so we have to use H.m to find out the ans of the qns]

Avg mean = H.m (20, 24)

$$\therefore H.m = \frac{1}{\frac{1}{2}\left(\frac{1}{20}+\frac{1}{24}\right)}$$

$$= \frac{2 \times 20 \times 24}{20+24}$$

$= 21.818$ m/h — Ans

Ex-2

If the interest paid on each of the 3 different sums of money yeilding 5%, 6% & 8% simple interest per anum (year) respectively is the same, then what is the avg interest on the total sum invested.

ammount x_1 x_2 x_3

5% 6% 8%

i - interest i i

Same interest

[If x_1, x_2 + x_3 and the percentage of interest are diff, then we have to use H.m].

avg interest rate = H.m (5, 6, 8)

$$\therefore H.m = \frac{3}{\left(\frac{1}{5}+\frac{1}{6}+\frac{1}{8}\right)}$$

$= 6.101\%$ — Ans

Ex-3 Qns

A man travels 12 miles (miles = distance) at 4 mi/h (speed) and again 10 miles at 5 mi/h. So what is the avg mean.

[mixture of speed = H.m]

x_i	f_i	f_i/x_i
4	12	
5	10	2
Total (distance covered)	$n = \sum f_i = 22$	

H.m (avg speed) $= \frac{22}{3+2} = \frac{22}{5} = 4.4$ mi/h — Ans

[Avg speed = $\frac{\text{total distance}}{\text{total time}}$

Here, $\frac{f_i}{x_i} = \frac{mi}{mi/h} = h$]

Ex-4 Imp

A man travel 12 hrs at 4 m/h (time) and again johns 5 m/h. what is the avg speed.

[A.m will be applied here, not H.m

mixed hrs = A.m

x	f	fx
4	12	48
5	10	50
	$\Sigma f = N = 22$	$\Sigma fx = 98$

Here Avg speed = Total distance / Total time

but here time is already given, no need to $\frac{\sum f_i}{\sum \frac{f_i}{x_i}}$

$= \frac{\sum f_i x_i}{\sum f_i}$

$\frac{\sum f_i x_i}{N} = \frac{48+50}{22} = \cancel{50.27}$ 4.45

H.m > G.m > A.m

The ans of H.m = A.m = H.m

Ex-5 Imp

A person purchased 300 rupees worth of orange from 5 markets at rupees 5, Rs 6, Rs 8, Rs 10, Rs 12 per orange respectively. what is the avg price of an orange?

H.m (5, 6, 8, 10, 12)

no. of oranges = not given

The price of oranges of each market is fixed

H.m

You can't apply A.m

300/- is fixed main thing fixed = H.m

$\Rightarrow$ Hm = $\frac{5}{\left(\frac{1}{5}+\frac{1}{6}+\frac{1}{8}+\frac{1}{10}+\frac{1}{12}\right)}$

(avg price per orange) = 7.407 /-

Ex-6 Imp

What would be the avg, if we had purchased 50 oranges from each market. (similar to prev ques)

given = no of oranges, price of an of each market is same A.m

total no = no fixed, main thing isn't fixed, but no of orange fixed = A.m

A.m (5, 6, 8, 10, 12)

avg prize per orange (A.m) = $\frac{5+6+8+10+12}{5}$

= 8.2 /-

for multiple groups

main things, products or any thing are not same, but of things varying = A.m, just opposite = H.m

main thing isn't fixed, but secondary thing same fixed = Apply A.m (kind of h.m but it is slightly different from h.m)

Imp Mean of composite Groups

Group 1 $\rightarrow n_1 \rightarrow$ no of observation, $x_1 \rightarrow$ mean

Group 2 $\rightarrow n_2 \rightarrow$ no of obs, $x_2 \rightarrow$ mean

What will be the mean of this 2 group as whole -

mean of means

$$\boxed{\bar{x} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}}$$

(mean of the composite groups, if we have 2 groups)

1) Arithmethic mean = Rest of cases

2) G.m = Primarily →

Cases
Ratios
Rate of change
Percentage change
Population size ...etc

3) H.m = Primarily →

Cases
Speed
Cost ... etc

or $\boxed{N = n_1 + n_2}$ [no. of groups. no. of observation eg = 2]

$$\boxed{N\bar{x} = n_1 x_1 + n_2 x_2}$$

[N = no. of groups / observations] [eg. 2 obs — n_1, n_2]

averages from each

2, 3 — (no. of groups) [N = total obs]

$$\boxed{\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_k x_k}{n_1 + n_2 + \cdots + n_k}}$$

$$\boxed{= \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i}}$$ ← for multiple groups.

m.A > m.G > m.H

m.A = m.G = m.H

Ex-① ~~Imp~~

There are 2 branches of an establishment employing 100 & 80 person respectively. If the A.m of the monthly salaries paid by the 2 branches are RS-35,000/- & RS-30,000/- respectively. Find the A.m of the salaries of the employees of the establishment as a whole?

$n_1 = 100$, $n_2 = 80$

$x_1 = 35,000$ $x_2 = 30,000$

$$\bar{x} = \frac{100 \times 35,000 + 80 \times 30,000}{180}$$

RS = 32778/-

② ~~Imp~~ The avg weekly salary of a dept in a farm was rs-5200 & that of another dept is rs-4200. The mean salary of all the employees was rs. 5000. Find the ratio & percentage of employees of the departments.

$x_1 = 5200/-$ $x_2 = 4200/-$ $\bar{x} = 5000/-$

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$$

$$\Rightarrow 5000 = \frac{n_1 \times 5200 + n_2 \times 4200}{n_1 + n_2}$$

$$\Rightarrow 5000 n_1 - 5200 n_1 = 4200 n_2 - 5000 n_2$$

$$\Rightarrow -200 n_1 = -800 n_2$$

$$\Rightarrow \frac{n_1}{n_2} = \frac{8}{2} = \frac{4}{1}$$

∴ the percentage of employees of 1st dept $\frac{4}{5} \times 100 = 80\%$.

2nd

∴ $\frac{1}{5} \times 100 = 20\%$.

Say there are k no. of groups in this case

k ← 1, 2, ... k

m_1, m_2 ... m_k

n_1, n_2 ... n_k

(3) The mean age of a group of 100 children was 9.35 yrs. The mean age of 25 of them was 8.75 years. and that of another 65 was 10.51 yrs, what was the mean of the remaining them?

$\Rightarrow 9.35 = \frac{25 \times 8.75 + 65 \times 10.51 + 10 \times x_3}{\cancel{25+65}\ 100}$

$\Rightarrow 935 = \cancel{90+10}\ 901.9 + 10 x_3$

$\Rightarrow x_3 = 3.31$ yrs

17/3/23

Geometric mean ~~for~~ composite group

Group 1	Group 2	Composite G.m
G.m $\rightarrow G_1$	G_2	$G = (G_1^{n_1} \cdot G_2^{n_2})^{\frac{1}{N}}$
No. of obs $\rightarrow n_1$	n_2	$N = n_1 + n_2$

$$\boxed{= \sqrt[N]{G_1^{n_1} \cdot G_2^{n_2}}}$$

Say there are K no. of groups, in this case –

K $\rightarrow$ 1, 2, — — , K

G_1, G_2 — — , G_K

n_1, n_2, — — n_K

$$\boxed{\text{Composite G.m} = G = \sqrt[N]{G_1^{n_1} \cdot G_2^{n_2} \cdot G_3^{n_3} \cdots G_K^{n_K}}}$$

[where $N = n_1 + n_2 + \cdots + n_K$]

$\Rightarrow \log G = \frac{1}{N}(n_1 \log G_1 + n_2 \log G_2 + \cdots + n_K \log G_K)$

$\Rightarrow \boxed{\log G = \frac{1}{N} \sum_{i=1}^{K} n_i \log G_i}$ [where $N = \sum_{i=1}^{K} n_i$]

EX-① (Composite standard deviation from included)

3 groups of obs contain 8, 7 & 5 obs. Their G.m are 8.52, 10.12, 7.75 respectively. Find the G.m of the 20th obs in the single group formed by pooling the 3 groups.

$n_1 = 8$, $n_2 = 7$, $n_3 = 5$

$G_1 = 8.52$, $G_2 = 10.12$, $G_3 = 7.75$

$\log G = \frac{1}{20}(8 \cdot \log 8.52 + 7 \cdot \log 10.12 + 5 \cdot \log 7.75)$

$\Rightarrow \log G = 0.9463$

$\Rightarrow G = 8.837$

(2) n no. of obs, have to find out A.m, G.m, H.m, then, [relation between Am, Gm, Hm]

i) $\boxed{A.m \geq G.m \geq H.m}$ (for any given set of observation)

$\rightarrow$ exactly equal when all observations are exactly the same.

Find the G.M of the 20 obs in the single group formed by pooling the 3 groups.

$n_1 = 8,\ n_2 = 7,\ n_3 = 5$

$G_1 = 8.52,\ G_2 = 10.12,\ G_3 = 7.75$

$$\log G = \frac{1}{20}\left(8 \cdot \log 8.52 + 7 \cdot \log 10.12 + 5 \cdot \log 7.75\right)$$

$\Rightarrow \log G = 0.9463$

$\Rightarrow G = 8.8371$

(Q)

or

$$G = \left(8.52^{8} * 10.12^{7} * 7.75^{5}\right)^{\frac{1}{20}}$$

$= 8.837$

iii) If we have only 2 obs, in that case -

$$\boxed{\frac{A.m}{G.m} = \frac{G.m}{H.m}} \rightarrow \text{only for 2 observation}$$

$$\Rightarrow G.m^2 = A.m \times H.m$$

iii) Standerd deviation for Composite groups

Group 1 → n_1 (obs) → $\bar{x}_1$ (mean) → δ_1 (standerd deviation)

Group 2 → n_2 → $\bar{x}_2$ → δ_2

Composite Group → $N = n_1 + n_2$ → $\bar{x}$ → δ

$$N\bar{x} = n_1\bar{x}_1 + n_2\bar{x}_2$$

$$\boxed{\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}}$$

$N = n_1 + n_2$

$$\boxed{N\delta^2 = (n_1\delta_1^2 + n_2\delta_2^2) + (n_1 d_1^2 + n_2 d_2^2)}$$

where $d_1 = \bar{x}_1 - \bar{x}$

$d_2 = \bar{x}_2 - \bar{x}$

$N = n_1 + n_2$

In case of k no. of groups, (generalized form)

$$N\delta^2 = \sum_{i=1}^{k} n_i\delta_i^2 + \sum_{i=1}^{k} n_i d_i^2$$

where, $N = \sum_{i=1}^{k} n_i$

$d_i = \bar{x}_i - \bar{x}$

EX-2 2nd

① For a group of 50 boys the mean score & standerd deviation of scores on a test is 59.5 & 8.38. For a group of 40 girls the same results are 54.0 & 8.23. Find the mean & standard deviation of the compiled group of 90 children.

Boys

$\bar{x}_1 = 59.5$

$n_1 = 50$

$\delta_1 = 8.38$

Girls

$\bar{x}_2 = 54.0$

$n_2 = 40$

$\delta_2 = 8.23$

$$N = 90, \quad \bar{x} = \frac{59.5 \times 50 + 54.0 \times 40}{90} = 57.056$$

$d_1 = 59.5 - 57.056 = 2.444$, $d_2 = 54.0 - 57.056 = -3.056$

$$\delta^2 = \frac{[50 \times (8.38)^2 + 40 \times (8.23)^2] + [50 \times (2.444)^2 + 40 \times (-3.056)^2]}{90}$$

$= 76.586$

$\Rightarrow \delta = 8.751$

$\therefore \delta = 8.751, \ \bar{x} = 57.056$

② Prob

the mean & variance of a group of 100 obs. are 6.5 & 3.0 respectively. 55 of this obs. have mean is 6.6 & standard deviation is 1.5 (δ_1). Find the mean & standard deviation of the remain 45 obs (n_2).

$\bar{x} = 6.5$
$\delta^2 = 3.0$

$n_1 = 55$ $\quad n_2 = 45$
$\bar{x}_1 = 6.6$ $\quad \bar{x}_2 = ?$
$\delta_1 = 1.5$ $\quad \delta_2 = ?$

$$\bar{x}_2 = 6.37, \qquad \delta_2 = 1.97$$

$$6.5 = \frac{55 \times 6.6 + 45\bar{x}_2}{100}$$

$\Rightarrow 363 + 45\bar{x}_2 = 650$
$\Rightarrow \bar{x}_2 = 6.37$

$d_2 = 6.37 - 6.5 = -0.13$
$d_1 = 6.6 - 6.5 = 0.1$

$$\Rightarrow 100 \times 3 = [55 \times (1.5)^2 + 45\delta_2^2] + [55 \times (0.1)^2 + 45 \times (-0.13)^2]$$

$\Rightarrow 45\delta_2^2 = 174.0395$

$\Rightarrow \delta_2 = 1.97$

②

21/3/23

Chw Prob

Absolute measures of dispersion - Range, Quantile deviation, mean deviaⁿ, Standard devⁿ, variance.

Relative measures of Dispersion

1) Coefficient of variance (C.V)

$$\boxed{= \frac{\delta}{\bar{x}} \times 100}$$ [δ = Standard devⁿ]

2) Coefficient of quantile deviation

$$= \boxed{\frac{Q.D}{\text{median}} \times 100}$$

3) Coefficient of mean deviation =

$$\boxed{\frac{M.D.}{\text{mean or median}} \times 100}$$

C.V. is widely used to compare the given data set.

$A \longrightarrow \bar{x} \quad \delta_1$ — $\bar{x}$ is same for both cases.
$B \longrightarrow \bar{x} \quad \delta_2$

$$\frac{C.V(A)}{C.V(B)} = \frac{\frac{\delta_1}{\bar{x}} \times 100}{\frac{\delta_2}{\bar{x}} \times 100} = \frac{\delta_1}{\delta_2}$$

$\boxed{\delta \text{ - sigma}}$

less C.V $\rightarrow$ Data set highly homogenous i.e. the diff. between obs of the data set is less (less variable)

large C.V $\longrightarrow$ inc. the diff of the obs of the data set is large (highly variable)

Also higly scatter around the mean

Say, C.V. (A) = 5, C.V(B) = 8, then this

C.V(A) Dataset is heighly homogeneous compared to other Data set & C.V(B) Data set is higly $\boxed{\text{variable}}$

Ex-①

2 farms A & B belonging to the same industry. pays ~~avgr~~ avg monthly salaries as follows -

	Farm A	Farm B
No. of workers	1000	1200
Avg monthly wages (salary)	RS 9800	RS 9800
variance of distribution of wages	100	121

Which particular farm higly homogenous salary to the workers? - Ans - Farm A. as it has lesser variance, so lesser standard deviat, so lesser C.V. (also the avg monthly salaries are exactly the same) we don't have to show calculation but if we want, we can)

② Heights & weights of students.

	Height	weight
mean =	162.6 cm	52.36 kg
variance	127.69 cm²	23.1361 kg²

~~C.V. (high~~ standard devi^n, $\delta_1 = \sqrt{127.69}$
$= 11.3$

Standard $\delta_2 = \sqrt{23.1361}$
$= 4.81$

$C.V(height) = \frac{11.3 \times 100}{162.6} = 6.95$

$C.V.(weight) = \frac{4.81 \times 100}{52.36} = 9.18$

$C.V.(weight) > C.V(height)$

∴ the weight is heigly variable.

∴ the height is highly homogenous.

∴ weight shows high variability than heights.

③ The following is the record of goals scored by team A in a foot ball season.

A

No. of goals scored (x_i)	0	1	2	3	4
No. of matches (f_i)	1	9	7	5	3

[As it is constant, it is not]

for team B, mean no. of goals scored per matchs was 2 with a standerd deviation of 1.25 goals. Find which team may be considered more consistent in scoring goals?

$mean(A) = \frac{\sum f_i x_i}{\sum f_i}$

$= \frac{50}{25}$

$\bar{x}_A = 2$

$S_A^2 = \frac{1}{N}\sum f_i x_i^2 - \left(\frac{1}{N}\sum f_i x_i\right)^2$

$= \frac{1}{25} \times 130 - \left(\frac{1}{25} \cdot 50\right)^2$

$= 1.2$

$S_A = 1.095$

$\bar{x}_B = 2, \; S_B = 1.25$

As the mean is exactly the same the team B is highly variably than A.

$\left[\begin{array}{l} C.V(A) = \frac{1.095 \times 100}{2} = 54.77 \\ C.V.(B) = \frac{1.25 \times 100}{2} = 62.5 \end{array} \right]$

∴ Team A is more consistent in scoring goals. [As A is highly homogenious]

Relation between mean, median & mode

1) $\boxed{mean - mode = 3(mean - median)}$ for only Data set is unimodal

This is for when your data set is very less skewed

2) If your data set is normally distributed then, $\boxed{mean = median = mode}$

Ex-① mean = 26.8, median = 27.9, estimate mode.

⇒ $26.8 - mode = 3(26.8 - 27.9)$

⇒ $mode = 30.1$

Relation between Standard devⁿ, quartile devⁿ & mean devⁿ, Range

1) If the dataset is normally distributed, then (4) $\boxed{S.D \geq M.D}$

Normal Distribution

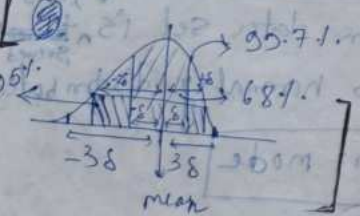
99.7% data are $\pm 3\delta$ away from the mean of the data set. (mean $\pm 3\delta$)

(1) $\boxed{Range = 6\delta}$

& $\frac{2}{3}$ or 68% of the data contains within 2 quartiles of your data

(2) $\boxed{Q.D = \frac{2}{3}\delta}$

(3) $\boxed{M.D = \frac{4}{5}\delta}$ = 95% of data



CW 24/3/23

Skewness, moments, kurtosis

Central tendency (mean)

measures of dispersion (var / s.d.)

A	B
↓	↓
$\bar{x}$	$\bar{x}$
δ_1	δ_2

$\bar{x}$ is same for both A & B

if, $\delta_1 > \delta_2$ then A is highly scattered

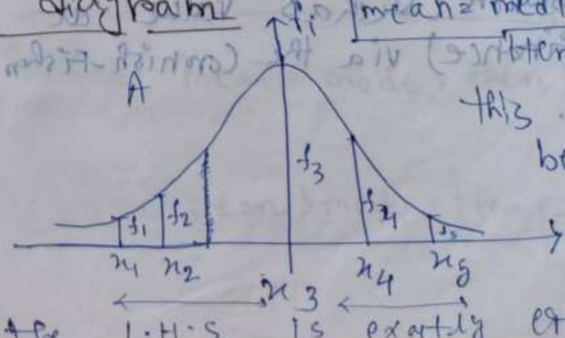
if,

A	B
$\bar{x}$	$\bar{x}$
δ	δ

Sometimes $\bar{x}$ & δ both are same for this case skewness, moments, kurtosis are arised.

Skewness :- Basically measures how the freq polygon / distribution is symmetric around the central point of a freq distribution of a particular variable.

Freq distribution curve / polygon / diagram



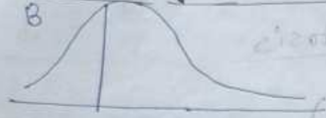
mean = median = mode = $\bar{x}$ for this case

there, skewness = 0

this particular will be known as symmetrical distribution.

the L.H.S is exactly equal to R.H.S

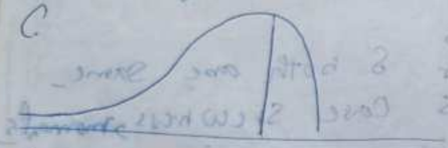
Positively skewed distribution



RHS has more observation.
In this case,

$\boxed{\text{mean} > \text{median} > \text{mode}}$

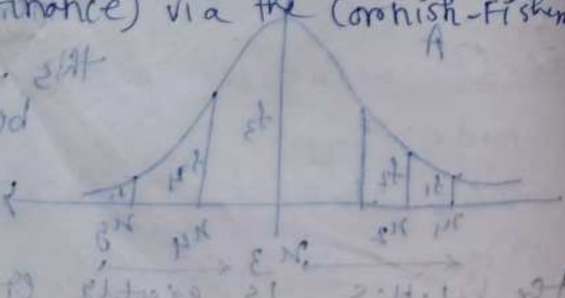
Negatively skewed distribution



LHS has more obs
In this case,

$\boxed{\text{mode} > \text{median} > \text{mean} \text{ or } \text{mean} < \text{median} < \text{mode}}$

Use of skewness- It is used to obtain approximate probabilities and quantiles of distribution (such as value at risk in finance) via the Cornish-Fisher expansion.



Absolute measures of skewness

when,
i) mean = mode = median $\longleftrightarrow$ symmetrical data set
ii) mean > mode $\longrightarrow$ Positively skewed
iii) mean < mode $\longrightarrow$ Negatively skewed

When,
i) $Q_3 - Q_2 = Q_2 - Q_1$ or $Q_2 - Q_1 = Q_3 - Q_2$ $\longrightarrow$ symmetrical data set
median, 1st quartile, 3rd quartile
ii) $Q_3 - Q_2 > Q_2 - Q_1 \longrightarrow$ positively skewed
iii) $Q_3 - Q_2 < Q_2 - Q_1 \longrightarrow$ negatively skewed

Relative measures of skewness

① Karl Pearson's coefficient of skewness.

$$\boxed{S_{kp} = \frac{\text{mean} - \text{mode}}{\text{standard deviation}}}$$

generally skewness varies from $[-3, 3]$ (theoritically)

i) when, mean = mode, then $\boxed{S_{kp} = 0}$ → symmetrical & not skewed dataset

ii) when mean > mode then $\boxed{S_{kp} > 0}$ → Positively skewed data

iii) when, mean < mode, $\boxed{S_{kp} < 0}$ → negatively skewed

S_{kp} measures the quantity of skewness.
S_{kp} measures of skewness.
~~EX-IP~~
Some times mode can be ill-defined (not defined)

then, $\boxed{\text{mean} - \text{mode} = 3(\text{mean} - \text{median})}$

~~IMP~~
~~Ex-~~
① ~~meat~~ In a data set mean is observed to be 150, mode is 56 & S.D of a data set is 15. Findout the pearson C. of skewness.

$S_{kp} = \frac{150 - 56}{45}$

$= -0.4$ slightly

∴ Data set is negatively skewed.
(negatively skewed can be 2 = slightly
ii) more negatively skewed compared to others)

IMP
② In a particular data set P.C.S = 0.4, S.D of the data set = 8, & mean = 30. Estimate mode of the data set.

⇒ $0.4 = \frac{30 - \text{mode}}{8}$

⇒ mode = 26.8

Data set is positively skewed. (as $S_{kp} = 0.4$)

IMP
③ In a data set mean = 45, median = 48. moderately skewed (ill-defined mode) $S_{kp} = -0.4$, then findout the S.D.

~~⇒ $0.4 = \frac{45 - 48}{S.D}$~~
~~⇒ S.D = 7.5~~
, this formula is only applicable for moderately skewed data set (when mode is ill-defined)

⇒ ~~45~~ $S_{kp} = \boxed{\frac{3(\text{mean} - \text{median})}{S.D.}}$

⇒ $-0.4 = \frac{3(45 - 48)}{S.D.}$

⇒ SD = 22.5

(~~positively~~ ⇒ negatively skewed)

$$\boxed{S_{kp} = \frac{\text{mean} - \text{mode}}{S.D} = \frac{3(\text{mean} - \text{median})}{S.D}}$$

(9) In a m. s. data set mean is = 20, median = 18.5 & coeff of varriance = 30%. Find out SKP.

$$C.V = \frac{s}{\bar{x}} \times 100 \quad \left[CV = \frac{s}{\bar{x}} \times 100\% \right] \left[CV_s \text{ unit } \% \right]$$

$$\Rightarrow 30 = \frac{s}{20} \times 100$$

$$\Rightarrow s = 6$$

$$\therefore S_{KP} = \frac{3(20 - 18.5)}{6} = 0.75$$

Slightly positively skewed.

(2) Bowley's skewness coefficient

It is based on the quartiles.

$$\boxed{S_{KB} = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}}$$

when, $Q_3 - Q_2 = Q_2 - Q_1$ → no skewness

then $\boxed{S_{KB} = 0}$ → symmetrical distribution

when, $Q_3 - Q_2 > Q_2 - Q_1$

then, $\boxed{S_{KB} > 0}$ → +vely skewed

when, $Q_3 - Q_2 < Q_2 - Q_1$

then, $\boxed{S_{KB} < 0}$ → -vely skewed

while

$$S_{KB} = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} > 0$$

$$\Rightarrow (Q_3 - Q_2) - (Q_2 - Q_1) > 0$$

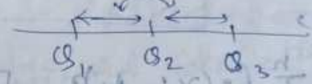
$$\Rightarrow Q_3 - Q_2 > Q_2 - Q_1 \longrightarrow \text{for the}$$

Similarly,

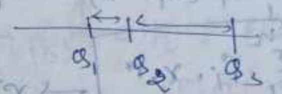
S_{KB}, $Q_3 - Q_2 < Q_2 - Q_1$ → for -ve

equal = equidistance

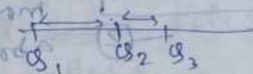
for, $S_{KB} = 0$,



for, $S_{KB} > 0$,



for, $S_{KB} < 0$,



Ex-

(1) In a dataset the diff among the quartiles = 15, & their sum is = 35, the median of the data set = 20. Find B.S.C.

$Q_2 = 20$. $Q_3 + Q_1 = 35$, $Q_3 - Q_1 = 15$

$$= \frac{35 - 2 \times 20}{15}$$

$$= -0.333$$

Imp

moments -

Suppose,

$X \quad x_i / f_i$

Imp

$$\boxed{\mu_p' = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - A)^p} \quad (1)$$

→ pth moment about the point $x = A$

where, $p = 0, 1, 2, 3$

when, $\boxed{A = 0}$ → moment about the origin

In this case,

$$\mu_p' = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - A)^p \quad \left[N = \sum_{i=1}^{n} f_i\right]$$

$$\boxed{\mu_p' = \frac{1}{N}\sum_{i=1}^{n} f_i \cdot x_i^p} \quad (2)$$

→ pth moment about the origin

Imp

$$\boxed{\mu_p' = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^p} \quad (3)$$

→ pth moment about the mean or Central moment

when, $p = 0$, A = present

then, $\mu_0' = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - A)^0$

$= \frac{1}{N}\sum_{i=1}^{n} f_i$

$= \frac{1}{N} \cdot N$

$\boxed{\mu_0' = 1}$ → 0th moment about $x = A$

when, $p = 1$, A

$$\boxed{\mu_1' = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - A)^1}$$

$\cancel{\frac{1}{N}\sum f_i}(x_i - A)$

$\mu_1' = x_i - A$

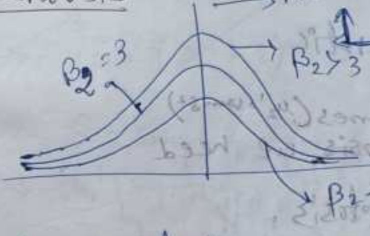
Similarly,

$$\boxed{\mu_n' = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - A)^n}$$

Imp

Kurtosis -

→ Normal distribution / normal curve



Here mean, S.D & median is exactly the same

It Kurtosis measures the peakedness of the data.

$$\boxed{\beta_2 = \frac{\mu_4'}{\mu_2'^2}} \quad (1)$$

$[\mu_4'$ = 4th central moment, μ_2' = 2nd central moment$]$

if, $\boxed{\beta_2 = 3}$, then our dataset is normal,

i.e, not flat, not peak.

if, $\boxed{\beta_2 < 3}$, then our d.s is flat.

if, $\boxed{\beta_2 > 3}$, then our d.s is peak.

if $\beta_2 = 3$, then data set or polygon is called
mesokurtic

$\beta_2 < 3$, ---- platykurtic

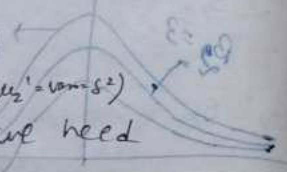
$\beta_2 > 3$, ---- leptokurtic

$$\boxed{\gamma_2 \text{ (gamma)} = \beta_2 - 3} \quad \text{— (2)}$$

then, $\boxed{\gamma_2 = 0} \rightarrow$ mesokurtic

$\boxed{\gamma_2 > 0} \rightarrow$ ~~lepto~~ leptokurtic

$\boxed{\gamma_2 < 0} \rightarrow$ platykurtic



To compute the measures ($\mu_2' = var = s^2$) ~~kurtosis~~ we need both moment & kurtosis.

28/3/23

Prob

Ex - moment

(1) The 1st 3 moments of a distribution about the value or point ~~to~~ 2 of the variable are 1, 16, -40 respectively, show that mean = 3, variance = 15, $\mu_3 = -86$.

$$\mu_n' = \frac{1}{n}\sum_{i=1}^{n} f_i \cdot (x_i - A)^n$$

moment about the ~~mean~~ Point A

$$\mu_n' = \frac{1}{N}\sum^{N} (x_i - A)^n \quad \text{— (1)}, \ A = 2$$

1st 3 moments, μ_1', μ_2', μ_3' (not 0th moment)

$\mu_1' = 1$, $\mu_2' = 16$, $\mu_3' = -40$

from (1), we get

$\mu_1' = (x_i - 2)^1$

$\Rightarrow 1 = x_i - 2$

$\Rightarrow x_i = 3$

$\cancel{\mu_3' = (x_i - 2)^3}$

$\Rightarrow \cancel{-40 = (x_i - 2)^3}$

$\cancel{\mu_2' = (x_i - 2)^2}$

$\cancel{16 = x_i - 2}$

$\cancel{\Rightarrow x_i - 2 = \pm 4}$

$\cancel{\Rightarrow x_i = 6, -2}$

$$\mu_1' = \frac{1}{N}\sum_{i=1}^{N} f_i (x_i - A)^1$$

$$= \frac{1}{n}\sum_{i=1}^{N} f_i x_i - \frac{1}{n}\sum_{i=1}^{N} f_i \cdot A$$

$$= \bar{x} - \frac{N}{N} A$$

$$= \bar{x} - A$$

$\Rightarrow 1 = \bar{x} - 2$

$\Rightarrow \bar{x} = 3$

mean $(\bar{x}) = 3$

$$\mu_2' = \frac{1}{N}\sum_{i=1}^{N} f_i (x_i - A)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} f_i x_i^2 - \frac{2}{N}\sum_{i=1}^{N} f_i \cdot A + \frac{A^2}{N}\sum_{i=1}^{N} f_i$$

$=$

general formula,

$$\boxed{\mu_r = \mu'_r - {}^rC_1\, \mu'_{r-1}\mu'_1 + {}^rC_2\, \mu'_{r-2}\mu_1'^2 - {}^rC_3\, \mu'_{r-3}\mu_1'^3 + \cdots (-1)^r \mu_1'^r}$$

r-th moment about the arbitary point

1st moment about the arbitary point

from that,

$$\Rightarrow \mu_2 = \mu'_2 - {}^2C_1 \cdot \mu'_1 \cdot \mu'_1 + {}^2C_2 \cdot \mu'_0 \mu_1'^2$$

$$= \mu'_2 - 2\mu_1'^2 + \mu_1'^2$$

$$= \mu'_2 - \mu_1'^2$$

$$= 16 - 1$$

$$= 15$$

$$\mu_2 = \frac{1}{N}\sum_{i=1}^{N} f_i(x_i - \bar{x})^2 \rightarrow \text{(central moment)}$$

$$= S^2 \qquad \left[\because S^2 = \frac{1}{N}\sum_{i=1}^{N} f_i(x_i - \bar{x})^2\right]$$

$$\Rightarrow S^2 = 15$$

$\therefore$ variance = 15

from that,

$r = 3$

$$\Rightarrow \mu_3 = \mu'_3 - {}^3C_1 \cdot \mu'_2 \cdot \mu'_1 + {}^3C_2 \cdot \mu'_1 \cdot \mu_1'^2 - {}^3C_3 \cdot \mu'_0 \mu_1'^3$$

$$= \mu'_3 - 3\cdot\mu'_2\cdot\mu'_1 + 3\cdot\mu_1'^3 - 1\cdot 1\cdot \mu_1'^3$$

$$= \mu'_3 - 3\mu'_2\mu'_1 + 2\mu_1'^3$$

$$= -40 - 3\cdot 16\cdot 1 + 2\cdot 1$$

$$= -86 \text{ (Proved)}$$

Ex- Kurtosis

Imp

① The 2nd, 3rd & 4th central moments are 2, 0.6 & 18.25. Find out the kurtosis.

$$\boxed{\mu'_2 = 2,\ \mu'_3 = 0.6,\ \mu'_4 = 18.25}$$

$$\beta_2 = \frac{18.25}{2^2}$$

$$= 4.5625 > 3$$

$\therefore$ this curve is leptokurtic (peaked than the normal curve).

Imp

② The 1st 4 central moments $0,\ 2.5,\ 0.7,\ 18.25$. Describe the nature of the curve.

$$\mu_1 = 0,\ \mu_2 = 2.5,\ \mu_3 = 0.7,\ \mu_4 = 18.25$$

$$\beta_2 = \frac{18.25}{(2.5)^2}$$

$$= 2.92 < 3$$

$\therefore$ this curve is mesokurtic (almost normal curve)

Imp Imp Def Covariance, correlation (from DS, DM, DL)

It is used to describe the relation among a pair of datas/variables/quantities.

If Covariance +ve, then if the val. of one variable is has increased then another

will
one also increase. If it is -ve
the reverse will happen. If the
covariance is ~~equal~~ zero, then the variables
are independent.

[Normal formula]

$$\boxed{cov(x, y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}$$

for sample data

When $\boxed{cov(x, y) = 0}$ → x & y are independent
∴ one variable doesn't affect
another variable.

$\boxed{cov(x, y) > 0}$ → x & y both increases
and decreases together

$\boxed{cov(x, y) < 0}$ → x ~~..~~ increases &
y decreases & vice
-versa

this formula can also be expressed as

$$\boxed{cov(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}$$

Proof it ↓ ↓ for population data

$$\boxed{cov(x, y) = \frac{1}{n} \sum x_i y_i - \left(\frac{1}{n} \sum x_i\right)\left(\frac{1}{n} \sum y_i\right)}$$

for faster computation of cov. then
It is the best formula.

[Here, n, N are different
$\bar{X} = \bar{x}$
$\bar{Y} = \bar{y}$]

Ex - Imp

① Calculate the cov. of x, y, if summation of $x_i = 50$, summation of $y_i = -30$, $\sum x_i y_i = -115$, $n = 10$.

$$= \frac{1}{10}(-115) - \left(\frac{1}{10} \times 50\right)\left(-\frac{1}{10} \times 30\right)$$

$= +~~26.5~~\ 3.5$, we have positive ~~correlation~~ covariance.

Prove that

$$cov(x, y) = \frac{1}{n} \sum x_i y_i - \left(\frac{1}{n} \sum x_i\right)\left(\frac{1}{n} \sum y_i\right)$$

We know that

$$cov(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{n} \sum_{i=1}^{n} (x_i y_i - x_i \bar{y}_i - \bar{x}_i y_i + \bar{x}\bar{y})$$

$$= \frac{1}{n} \sum_{i=1}^{n} (x_i y_i) - \frac{1}{n} \sum_{i=1}^{n} x_i \bar{y}_i - \frac{1}{n} \sum \bar{x}_i y_i + \frac{1}{n} \sum_{i=1}^{N} \bar{x}\bar{y}$$

$$= \frac{1}{n} \sum_{i=1}^{n} (x_i y_i) - \cancel{n}\ \bar{x}_i \bar{y}_i - \bar{x}_i \bar{y}_i + \bar{x}\bar{y}$$

$[\because \frac{1}{n} \sum_{i=1}^{n} \bar{x}\bar{y} = \bar{x}\bar{y}]$

$$= \frac{1}{n} \sum_{i=1}^{n} (x_i y_i) - \bar{x}\bar{y} - \cancel{\bar{x}\bar{y}} + \cancel{\bar{x}\bar{y}}$$

$[\because \frac{1}{n} \sum x_i \bar{y}_i = \bar{x}\bar{y}]$

$$= \frac{1}{n}(x_i y_i) - \left(\frac{1}{n} \sum x_i\right)\left(\frac{1}{n} \sum y_i\right)$$

$[\because \frac{1}{n} \sum_{i=1}^{n} (x_i y_i) = \frac{1}{n}(x_i y_i)]$

$$\text{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_iy_i - x_i\bar{y} - y_i\bar{x} + \bar{x}\bar{y})$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n}x_iy_i - \bar{y}\sum_{i=1}^{n}x_i - \bar{x}\sum_{i=1}^{n}y_i + \sum_{i=1}^{n}\bar{x}\bar{y}\right\}$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n}x_iy_i - \bar{y}(n\bar{x}) - \bar{x}(n\bar{y}) + n\bar{x}\bar{y}\right\} \quad \left[\because \sum_{i=1}^{n}x_i = n\bar{x},\ \sum_{i=1}^{n}y_i = n\bar{y}\right]$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n}x_iy_i - n\bar{x}\bar{y}\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_iy_i - \bar{x}\bar{y}$$

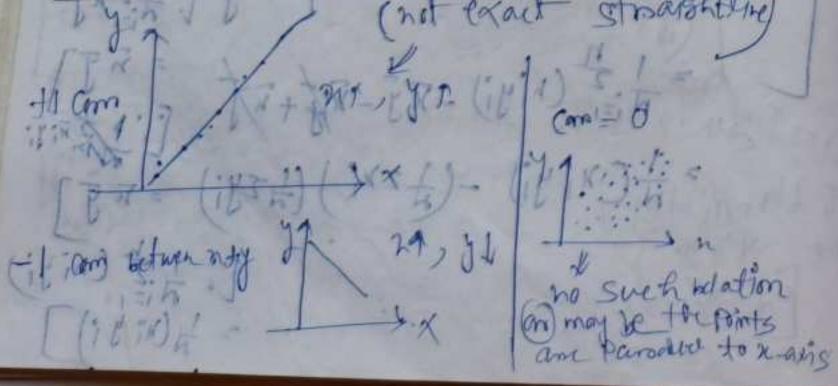
$$= \frac{1}{n}\sum_{i=1}^{n}x_iy_i - \left\{\frac{1}{n}\sum_{i=1}^{n}x_i\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}y_i\right\}$$

$$\therefore \text{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}x_iy_i - \left\{\frac{1}{n}\sum_{i=1}^{n}x_i\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}y_i\right\}$$

(Proved)

31/3/23

Scatter plot



31/3/23

Covariance

$$\boxed{\text{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}$$

$$\boxed{\text{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}x_iy_i - \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right)}$$

Correlation - Standardise relation to find out the relationship between variables. It is a statistical measure that expresses the extent to which two variables are linearly related (they can change together at a constant rate).

Pearson's correlation

$$\text{Pearson's Correlation} = \frac{\text{Cov}(x,y)}{\sqrt{\text{Var}(x)}\sqrt{\text{Var}(y)}} = \frac{\text{Cov}(x,y)}{S_x \cdot S_y}$$

$$= \frac{\frac{1}{n}\sum x_iy_i - \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (2)$$

$$P_{corr}(x,y) = \frac{\frac{1}{n}\sum_{i=1}^{n}x_iy_i - \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}y_i\right)^2}} \quad (3)$$

[Corr = Correlation, P = Pearson's]

$$P_{corr}(x,y) = \frac{n\sum_{i=1}^{n}x_iy_i - \left(\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i\right)}{\sqrt{n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2}\cdot\sqrt{n\sum_{i=1}^{n}y_i^2 - \left(\sum_{i=1}^{n}y_i\right)^2}} \quad (4)$$

[multiplying by n^2 both Nr & Dr]

For faster correlation we will use this formula –

$$\boxed{\rho_{corr}(x,y) = \frac{\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}}} \quad (1)$$

Advantage of using correlation.

The correlation value always lies between [-1, +1], where -1 indicates that we have a perfectly negative correlation between the variable x & y, and +1 indicates we have a perfectly positive correlation. (meaning of perfectly (+1) & (-1) – previous).

when $Corr(x,y) = 0 \rightarrow x, y$ are completely independent.

(P.C.C) – Pearson Correlation Coefficient

Ex – Prob

(1) Find the coeff. of correlation between x & y when $cov(x,y) = -16.5$ & $var(x) = 2.89$ & $var(y) = 100$

$\Rightarrow \rho_{corr}(x,y) = \frac{-16.5}{\sqrt{2.89}\cdot\sqrt{100}} = -0.97$

nearly perfectly negative correlation. [-1]

Prob

(2) no. of obs (n) = 25, $\sum x = 125$, $\sum y = 100$, $\sum x^2 = 650$, $\sum y^2 = 460$, $\sum xy = 508$. find out the correlation between x & y.

$$\rho_{corr}(x,y) = \frac{25\times 508 - 125\times 100}{\sqrt{25\times 650 - (125)^2}\cdot\sqrt{25\cdot 460 - (100)^2}}$$

$$= \frac{200}{25\times\sqrt{1500}}$$

Same data $= 0.206$ Slightly positively corr.

Prob (3) It was later discovered that at the time of recording the data set instead of the recording

x	y
6	14
8	6

pairs, the corrected value of x-y pairs should have been

x	y
8	12
6	8

Now compute the corrected correlation between x & y.

Correct,

x	y
8	12
6	8

Incorrect,

x	y
6	14
8	6

corrected $\Sigma x = 125 - (6+8) + (8+6)$

$= 125$

$\Sigma y = 100 - (14+6) + (12+8)$

$= 100$

$\cancel{\Sigma x^2 = (125)^2 = 15625}$

$\cancel{\Sigma y^2 = (100)^2 = 10000}$

$\Sigma xy = 508 - (8\times12 + 6\times8)$

$+ (6\times14 + 8\times6)$

$= \cancel{496}\ 520$

$\Sigma x^2 = 650 - (6^2+8^2) + (8^2+6^2)$

$= 650$

$\Sigma y^2 = 460 - (14^2+6^2) + (12^2+8^2)$

$= \cancel{436}\ \cancel{520}\ 436$

$\cancel{496}\ 520$

$\cancel{520}$

$$P_{cor(x,y)} = \frac{25 \times \cancel{496} - 125\times100}{\sqrt{25\times650-(125)^2}\cdot\sqrt{25\times436 - (100)^2}}$$

$= \cancel{\frac{100}{25\times 54.77}}$

$= 0.66$

Positive cor.

(3)

Probability = It is the measure of the likelihood of an event to happen.

It basically measures how likely a particular event is going to happen.

Probability ~~for~~ (Event) = P(Event) → in general

P(Event) = $\frac{\text{Outcomes that meet our criteria}}{\text{all possible outcomes}}$
P(Event) = $\frac{\text{Outcomes favourable to our criteria}}{\text{all possible outcomes}}$

$\boxed{P = \frac{m}{n}}$

$P(H) = \frac{1}{2} = \frac{1}{2} = 50\% = 0.5$

Head $\left(\frac{1}{2}, \frac{1}{2}\right)$

$= \frac{|\{H\}|}{|\{H,T\}|}$

Probability lies - [0, 1]

$P(E_1) = 0 \longrightarrow$ Impossible event

$P(E_2) = 1 \longrightarrow$ Sure / Certain event

$P(H \text{ or } T) = 1 = \frac{|\{H,T\}|}{|\{H,T\}|}$

Tools of Probability

Random Experiment whenever you perform a particular action repeatedly & the ~~ot~~ outcome (of it) isn't certain but it is anyone of all possible out comes, then this type of experiment is called Random Experiment.

Ex — $P(H \text{ or } T) = 1$

Usually, we may get a different number of outcomes from an experiment. However, when an experiment satisfies the following 2 conditions, it is called a random experiment. (i) It has more than 1 possible outcome, (ii) It isn't possible to predict the outcome in advance.

Event - It is any subset of all possible outcomes of a random exp.

Ex—① {HH, HT, TH, TT} = Sample space = all possible outcomes

Any subset, $P(HH) = \frac{1}{4} = 0.25$

The no. of favourable outcomes to the total no. of outcomes is defined as the probability of occurrence of any event.

Ex - ② 6-sided fair

$P(2 \text{ or } 4 \text{ or } 6) = \frac{3}{6} = 0.5$

$= \frac{|\{2,4,6\}|}{|\{1,2,3,4,5,6\}|}$

{2,1}, {1,2}, {2,2}, ~~{2,3}, {3,2}, {2,4}, {2,5}~~, {2,6}, {1

negation of event or complementary Event—

If, the event we are discussing about can't happen, then this type of event is called negation of event.

Ex—① ~~sum of~~ throw dice 2 times & getting the sum ≥ 3 → the ~~Samp~~ possible outcomes are very large. So this event can't happen, but it will take much time, so we will use complement or negation

{1,1} {1,2} ... {6,6}

{1,1}

$P(\bar{E}) = \frac{1}{36}$

$P(E) = 1 - P(\bar{E}) = 1 - \frac{1}{36} = \frac{35}{36}$

$$\boxed{P(E) + P(\bar{E}) = 1}$$

Playing cards

1. Total 52 cards except joker
2. 26 red & 26 Black
3. 4 suits each with 13 cards
4. a) spade ♠ – black
   b) Club ♣ – black
   c) heart ♥ – red
   d) diamond ♦ – red
5. Each suit contains 13 cards
   A (1, 14), 2, 3, 4, 5, 6, 7, 8, 9, 10, jack (knave on right), king, queen
6. 9 number cards in each suit
7. 3 face cards in each suit: jack, king, queen.
8. 12 face cards

Ex-① $P(\text{King}) = \frac{4}{52} = \frac{1}{13}$

Ex-② $P(\text{red queen}) = \frac{2}{52} = \frac{1}{26}$

4/4/23

Experiment – It is some occurrence that produce some outcomes.

Random experiment – Done

Deterministic Experiment – The exps. which have only one possible result or outcome i.e whose result is certain or unique.

Sample space, Event – Done

↓ Subset of a sample space

Elementary Events – Each of the individual outcomes of a sample. (An event having only one outcome or any single outcome of a trial (exp) ~~experiment~~ event)

mutually Exclusive Event – two or more events are performed they are said M.E. if the occurrence of one event prevents the occurrence of other event, then it is called.

Ex– $E_1 = \{H\}$ → It prevents E_2

$E_2 = \{T\}$ → It prevents E_1

Exhaustive Events.

let, S = Sample space, $A_1, A_2, \ldots A_n$ be some events which are subset of S $[A_i \subset S]$, in these case we will say that $A_1, A_2 — A_n$ is exhaustive events if, $\cancel{\sum_{i=1}^{n} P(A_i) = 1}$ $\boxed{A_1 \cup A_2 \cup \ldots A_n = S}$

$$\boxed{P(\bar{A}) = 1 - \frac{m}{n} = \frac{n-m}{n}}$$

Ex

① In a symaltanious toss of 3 coins. Find the probability of getting i) 2 H

ii) Exactly one H

iii) Exactly 2 T

iv) Exactly 1 T

v) no tails

$S = \{(H,H),(H,T),(T,H),(T,T)\}$

i) $P(2H) = \frac{2}{8} = \frac{1}{4}$

ii) $P(\text{Exactly } 1H) = \frac{4}{8} = \frac{1}{2}$

iv) $P(1T) = \frac{4}{8} = \frac{1}{2}$

v) $P(\text{no tails}) = \frac{2}{8} = \frac{1}{4}$

(iii) $P(\text{Exactly } 2T) = \frac{2}{8} = \frac{1}{4}$

② 3 coins are tossed once. find the probability of getting -

i) all heads iv) no H vii) A head on first coin.

ii) Atleast 2 H v) Exactly one tail

iii) Atmost 2H vi) Exactly 2 T

$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

i) $P(A) = \frac{1}{8}$

ii) $P(B) = \frac{4}{8} = \frac{1}{2}$

iii) $P(C) = \frac{7}{8}$

iv) $P(D) = \frac{1}{8}$

v) $P(E) = \frac{3}{8}$

vi) $P(F) = \frac{3}{8}$

vii) $P(G) = \frac{4}{8} = \frac{1}{2}$

③ Two dice are thrown symultaniously. Find the probability of getting -

i) An even no. as the sum.

ii) The sum is a prime no.

iii) a total of atleast 10

iv) a doublet of even no.

v) a multiple of 2 on dice one and a multiple of 3 on the dice two.

vi) Same no. on both dice

vii) a multiple of 3 as the sum

i) S = {1,1; 1,3; 1,5; 2,2; 2,4; 2,6; 3,1; 3,3; 3,5; 4,2; 4,4; 4,6; 5,1; 5,3; 5,5; 6,2; 6,4; 6,6}

$P(A) = \frac{18}{36} = \frac{1}{2}$

ii) S = {(1,2), (1,4), (1,6), (2,1), (2,3), (2,5), (3,2), (3,4), ~~(3,6)~~ (4,1), (4,3), (5,2), (5,6), (6,1), (6,5), (1,1)}

$P(B) = \frac{15}{36}$

iii) S = {~~(1,1), (1,2), (1,3), (1,4), (1,5)~~, (3,6) (4,4) (5,5) (5,6), (6,5), (6,6)}

$P(C) = \frac{6}{36}$

iv) S = {(2,2), (4,4), (6,6)}

$P(D) = \frac{3}{36} = \frac{1}{12}$

v) S = {(2,3), (4,6), (4,3), (2,6), (6,3), (6,6)}

$P(E) = \frac{6}{36} = \frac{1}{6}$

vi) $P(F) = \frac{6}{36} = \frac{1}{6}$

vii) S = {(1,2), (1,5), (2,4), (3,3), (3,6), (6,4), (2,1), (5,1), (4,2), (4,5), (6,3), (6,6)}

$P(G) = \frac{12}{36} = \frac{1}{3}$

(9) 4 cards are drawn at random ~~of back~~ of 52 playing cards. Pack

Find the p of getting -

i) All the 4 cards of the same suits.

ii) All the 4 cards of the same no.

iii) One card from each suit.

iv) two red cards and two black cards.

v) all cards of the same colour.

vi) all face cards.

i) $P(A) = \frac{{}^{13}C_4 + {}^{13}C_4 + {}^{13}C_4 + {}^{13}C_4}{{}^{52}C_4} = \frac{2860}{270725} = \frac{44}{4165}$

(ii) $P(B) = \frac{9}{{}^{52}C_4} = \frac{9}{270725}$

$= {}^4C_4 + {}^4C_4 + {}^4C_4 + \ldots + {}^4C_4$ (9 times) [∵ 9 no. and 4 same suits no. of]

(iii) $P(C) = \frac{{}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1}{{}^{52}C_4} = \frac{2197}{20825}$

(iv) $P(D) = \frac{{}^{26}C_2 \times {}^{26}C_2}{{}^{52}C_4} = \frac{325}{833}$

(vi) $P(F) = \frac{{}^{12}C_4}{{}^{52}C_4} = \frac{495}{270725} = \frac{99}{54145}$ [3 face cards from each 4 suits]

(v) $P(E) = \frac{{}^{26}C_4 + {}^{26}C_4}{{}^{52}C_4} = \frac{92}{833}$

$\boxed{P(\text{All face \& Ace cards}) = \frac{13}{52}}$

Addition Rule on Probability 11/4/23

Imp

"at least" one among those (all) events occurs or not.

Suppose a coin is tossed twice

$S = \{(H,H), (H,T), (T,H), (T,T)\}$ Find P of getting "Atleast" one head when a coin is tossed twice.

$H_1 = H$ on 1st trial, $H_2 = H$ on 2nd trial

$P(H_1) = \frac{2}{4} = \frac{1}{2}$ $P(H_1) + P(H_2) = \frac{1}{2} + \frac{1}{2}$ (wrong) as it denotes there is 100% chance of getting H, which isn't correct.

$P(H_2) = \frac{2}{4} = \frac{1}{2}$

Addition rule of Prob. says that,

$$\boxed{P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)}$$

$\quad\ \ \cup \qquad\qquad\qquad\qquad\qquad\quad \cap$

So,

$P(H_1 \text{ or } \cup H_2) = P(H_1) + P(H_2) - P(H_1 \cap H_2)$

$= \frac{1}{2} + \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{2}$

$= 1 - \frac{1}{4}$

$= \frac{3}{4}$

(1) $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

(2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

→ A & B are mutually exclusive events, $P(A \cap B) = 0$, $P(A \cup B) = P(A) + P(B)$

This is just for 2 events

∴ Similarly,

A, B, C

3) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$

→ (Prove)

$P(A \cup B \cup C)$, let $B \cup C = X$

$P(A \cup X) = P(A) + P(X) - P(A \cap X)$

$= P(A) + P(B \cup C) - P(A \cap (B \cup C))$

$= P(A) + P(B) + P(C) - P(B \cap C) -$

[$P(A \cap (B \cup C))$

$= P((A \cap B) \cup (A \cap C))$

$= P(A \cap B) + P(A \cap C)$ [∵ $A \cap A = A$] (Prove)

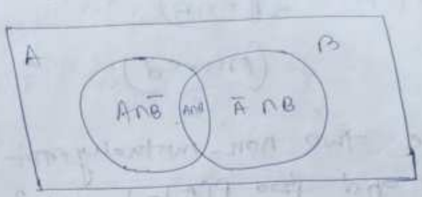
$- P(A \cap B \cap A \cap C)$

$= P(A \cap B) + P(C \cap A) - P(A \cap B \cap C)$]

$P(A \cap B) - P(C \cap A) + P(A \cap B \cap C)$

[∵ $P(A \cap B \cap A \cap C) = P(A \cap B \cap C)$]

~~Probability~~



$A \cap \bar{B}$ and $A \cap B$ are mutually exclusive

$(A \cap \bar{B}) \cup (A \cap B) = A$

$P(A \cap \bar{B}) \cup P(A \cap B) = P(A)$

4) $P(A \cap \bar{B}) = P(A) - P(A \cap B)$

$\bar{A} \cap B$ and $A \cap B$ are mutually exclusive and also,

$(\bar{A} \cap B) \cup (A \cap B) = B$

$\Rightarrow P(\bar{A} \cap B) \cup P(A \cap B) = P(B)$

5) $P(\bar{A} \cap B) = P(B) - P(A \cap B)$

6) $P((A \cap \bar{B}) \cup P(\bar{A} \cap B)) = P(A) + P(B) - 2P(A \cap B)$

Prove, $P((A \cap \bar{B}) \cup P(\bar{A} \cap B)) = P(A) - P(A \cap B) \cup P(B) - P(A \cap B)$

$(A \cap \bar{B})$ & $(\bar{A} \cap B)$ are mutually exclusive

$P(A\cap\bar{B}) \cup P(A\cap B)$

$\boxed{P(A\cap\bar{B}) = P(A) - P(A\cap B)}$

mutually exclusive

$$P\big(P(A\cap\bar{B}) \cup P(\bar{A}\cap B)\big) = P(A\cap\bar{B}) + P(\bar{A}\cap B) - 0$$

[$\because$ $\bar{A}\cap B$ & $A\cap\bar{B}$ are mutually exclusive]

$$= P(A) - P(A\cap B) + P(B) - P(A\cap B)$$

$$= P(A) + P(B) - 2P(A\cap B)$$

So,

$= P(A) + P(B) - 2P(A \cap B)$

(proved)

Ex-①

① A & B are two non-mutually exclusive events and $P(A) = \frac{1}{4}$, $P(B) = \frac{2}{5}$ & $P(A \cup B) = \frac{1}{2}$. Find $P(A \cap B)$ & $P(A \cap \bar{B})$.

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\Rightarrow \frac{1}{2} = \frac{1}{4} + \frac{2}{5} - P(A \cap B)$

$\Rightarrow P(A \cap B) = \frac{1}{4} + \frac{2}{5} - \frac{1}{2}$

$= \frac{3}{20}$

$P(A \cap \bar{B}) = P(A) - P(A \cap B)$

$= \frac{1}{4} - \frac{3}{20}$

$= \frac{5-3}{20}$

$= \frac{2}{20}$

$= \frac{1}{10}$

② If E & F are two events such that $P(E) = \frac{1}{4}$, $P(F) = \frac{1}{2}$, P(E and F) $= P(E \cap F) = \frac{1}{8}$. Find $P(E \text{ or } F) = P(E \cup F)$ & $P(\bar{E} \cap \bar{F})$.

~~P(A∪B)~~

$P(E \cup F) = \frac{1}{4} + \frac{1}{2} - \frac{1}{8}$

$= \frac{2+4-1}{8}$

$= \frac{5}{8}$

$P(\bar{E} \cap \bar{F}) = 1 - P(E \cup F)$

$= 1 - \frac{5}{8}$

$= \frac{3}{8}$

$P(\overline{E \cup F}) = 1 - P(E \cup F)$ $\boxed{P(\bar{A}) = 1 - P(A)}$

③

Sport	Male	Female	Total
Football	22	16	38
Basketball	13	8	21
Other	25	16	41
Total	60	40	100

Find out what is the Prob. of a person is chosen i.e. P(male or Football) = ?

is a male or prefer football

Let.

A = Person is a male

B = Prefer football

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A) = \frac{60}{100}$

$P(B) = \frac{38}{100}$

$P(A \cap B) = \frac{22}{100}$

$\therefore P(A \cup B) = \frac{60}{100} + \frac{38}{100} - \frac{22}{100}$

$= \frac{98 - 22}{100}$

$= \frac{76}{100} = \frac{19}{25}$

$= 0.76$

CW — 2/4/23

(4) Find the probability of getting an even no. on the 1st die or total of 8 in a single through of 2 dies.

S/-

A = the probability of getting an even no. in the 1st die

B = total of 8 in a single through of 2 dies.

$S_1 = \{(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)$, $(4,1)$, $(6,1)\}$

$P(A) = \frac{18}{36} = \frac{1}{2}$

$S_2 = \{\cancel{(1,3)}\ \cancel{(4,4)}, (2,6), (6,2), (3,5), (5,3), (4,4)\}$

$P(B) = \frac{5}{36}$

$P(A \cup B) = ?$

$P(A \cap B)$

$S = \{(2,6), (6,2), (4,4)\}$

$P(A \cap B) = \frac{3}{36} = \frac{1}{12}$

$\therefore P(A \cup B) = \frac{1}{2} + \frac{5}{36} - \frac{1}{12}$

$= \frac{18 + 5 - 3}{36}$

$= \frac{20}{36} = \frac{5}{9}$

(5) A die is thrown twice. What is the probability that atleast one of the 2 throws comes us with a no. 4.

$S = \{(1,4), (4,1), (2,4), (4,2), (3,4), (4,3), (4,4), (5,4), (4,5), (6,4), (4,6)\}$

A = getting 4 on the 1st trial

B = 4 on the 2nd trial

$A \cap B$.

$P(A) = \frac{6}{36}$

$P(B) = \frac{6}{36}$

$P(A \cap B) = \frac{1}{36}$

$P(A \cup B) = \frac{6}{36} + \frac{6}{36} - \frac{1}{36}$

$= \frac{12-1}{36}$

$= \frac{11}{36}$

(6) One no. is chosen from no.s 1 to 200. Find the probability that it is divisable by 4 or 6.

A = Divisiable by 4 $= \frac{200}{4} = 50$

B = " 6 $= \frac{198}{6} = 33$

$P(A \cup B) = ?$ $A \cap B = \frac{192}{12} = 16$

~~S = {4, 8, 12, 16, 20, 24, 28, 36, 40, 44, 48,~~

$P(A \cup B) = \frac{50 + 33 - 16}{200}$

$= \frac{67}{200}$

(7) 4 cards are drawn at a time, from a pack of 52 playing cards. Find the probabilities of getting - all the 4 cards of the same suit.

$P(A) = \frac{{}^{13}C_4 + {}^{13}C_4 + {}^{13}C_4 + {}^{13}C_4}{{}^{52}C_4}$ [∵ as all are independent]

$= \frac{2860}{270725}$

$= \frac{44}{4165}$

(8) 2 cards are drawn from a pack of 52 cards. What is the probability that either both are red or both are kings.

$P(A \cup B) = \frac{{}^{26}C_2 + {}^{4}C_2 - {}^{2}C_2}{{}^{52}C_2}$

$= \frac{331}{1326}$ $\frac{330}{1326} = \frac{55}{221}$

Conditional Probability

Imp

Independent Events, Dependent Events

A B

when the ~~out come~~ occurrence of A doesn't affect the outcome of B & vice versa.

When the outcome of the one (event) affects the outcome of another event.

Conditional Probability

whenever one exp is already done & depending upon it you have to get the probability of another exp then this is called conditional probability.

$P(A|B)$ = what is the $P(A)$ when B is already occured

Conditional probability works in the

$\boxed{\text{multiplication rule.}}$

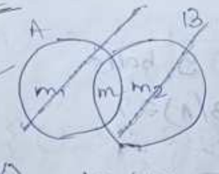
If A & B are independent, then the occurance of A doesn't affect the ~~out~~ outcome of B and vice versa.

$$\boxed{P(A|B) = P(A)\cdot P(B) = P(B/A)}$$ Independent events

dependent,

$$\boxed{\begin{aligned} P(A\cap B) &= P(A)\cdot P(B/A) \\ P(A\cap B) &= P(B)\cdot P(A/B) \end{aligned}}$$

Prove



total n elementary events

$A \to m_1$
$B \to m_2$ $A\cap B \to m$

~~$P(A) = \frac{m_1+m}{n}$~~ $P(A) = \frac{m_1}{n}$

~~$P(B) = \frac{m_2+m}{n}$~~ $P(B) = \frac{m_2}{n}$ $P(A\cap B) = \frac{m}{n}$

$$P(A\cap B) = \frac{m}{n} = \frac{m}{m_1}\cdot\frac{m_1}{n} = P(B/A)\cdot P(A)$$

~~$P(A/B) =$~~

$$P(A\cap B) = \frac{m}{n} = \frac{m}{m_2}\cdot\frac{m_2}{n} = P(A/B)\cdot P(B)$$

$$\boxed{\begin{aligned} P(A\cap B) &= P(B/A)\cdot P(A) \\ P(A\cap B) &= P(A/B)\cdot P(B) \end{aligned}}$$

← rule for conditional probability whenever the events are dependent

Ex

(1) Let there be a bag containing 5 white & 4 red balls, 2 balls are drawn from the bag one after another without replacement. Find p of drawing a red ball in the 2nd drawn given that a white ball has been already drawn in the 1st draw.

A = Drawing a white ball in the 1st drawn

B = ——— 2nd drawn

5 white balls } total 9 balls
4 red balls } $P(B/A) = ?$

$P(A) = \frac{4}{8}$, $P(B) = \frac{1}{8}$

$P(A \cap B) =$

$P(B/A) = \frac{P(A \cap B)}{P(A)}$ $\left[\frac{P(A \cap B)}{P(A)} = \frac{\frac{4}{8}}{\frac{5}{8}} = \frac{1}{2} \right]$

$= \frac{4}{8}$

$= \frac{1}{2}$

(2) A Pair of dies is thrown if the 2 nos appearing on them are different. Find the probability – i) the sum of the nos is 6.

ii) Sum of the nos is 4 or less

iii) Sum of the nos is 4.

Let us consider that

A = no. appearing on the dies are different

B = sum of the no. is 6

C = Sum of the no. is 4 or less

D = Sum of the no. is 4

$S_A = \{(1,2), (2,1), \ldots, (2,5), \ldots\}$

$P(A) = \frac{30}{36}$

$B = \{(1,5), (5,1), (2,4), (4,2), \cancel{(3,3)}\}$

$P(B) = \frac{4}{36}$

$S_C = \{(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)\}$

$P(C) = \frac{4}{36}$

$S_D = \{(1,3), (3,1)\}$

$P(D) = \frac{2}{36}$

i) $P(B/A) = \frac{4}{30}$

ii) $P(C/A) = \frac{4}{30}$

iii) $P(D/A) = \frac{2}{30}$

(3) A bag contains 10 white & 15 black balls. 2 balls are drawn in succession without replacement. What is the P that 1st is ~~each~~ white and the 2nd is black.

A = the 1st ball is white

B = 2nd ball is black

$P(A) = \frac{^{10}C_1}{^{25}C_1}$ $P(B) = \frac{^{15}C_1}{^{25}C_1}$ $P(A \cap B) = \frac{^{15}C_1}{^{24}C_1}$

$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{15}{24}$

$\Rightarrow \frac{15}{24}$

$\frac{10}{25}$

$= \frac{5}{24} \times \frac{25}{10}$

$= \frac{75}{48}$ $\quad \frac{25}{16}$

$= \frac{1 \times 25}{4 \times 102}$

$= \frac{1}{8}$

2/5/23

Ch Conditional Probability

Ex-

① Find the probability of drawing a diamond card in each of the two consecutive draws, from a well shuffled pack of cards. If the card drawn is not replaced after the 1st draw.

the required probability –

$P(A) \frac{^{13}C_1}{^{52}C_1} \times \frac{^{12}C_1}{^{51}C_1} \leftarrow P(B/A)$

$P(A \cap B) = \frac{1}{17}$

⑤ A bag contains 5 white, 7 red, & 8 black balls. If 4 balls are drawn 1 by 1, without replacement. Find the P of getting all white balls.

5 → white
7 → red
8 → black

A, B, C, D indicates that getting white ball 1st trial, ~~red 2nd~~, white 2nd, white 3rd trial, white 4th trial.

$P(A \cap B \cap C \cap D) = \frac{^5C_1}{^{20}C_1} \times \frac{^4C_1}{^{19}C_1} \times \frac{^3C_1}{^{18}C_1} \times \frac{^2C_1}{^{17}C_1}$

$P(A)$ $\quad P(B/A)$ $\quad P(C/A \cap B)$ $\quad P(D/A \cap B \cap C)$

$= \frac{1}{969}$

$= P(A) \cdot P(B|A) \cdot P(C|A \cap B) \cdot P(D|A \cap B \cap C)$

⑥ An urn contains 5 white, 8 black balls. 2 successive drawings of the 3 balls at a time are made such that the balls aren't replaced before the 2nd draw. Find the P that the 1st draw gives 3 white balls & the 2nd draw gives 3 black balls.

A → 1st drawn 3 white balls
B → 2nd draw 3 black

$P(A \cap B)$

$P(A) \cdot P(B/A)$

$= \frac{5_{C_3}}{13_{C_3}} \cdot \frac{8_{C_3}}{10_{C_3}}$

$= \frac{7}{429}$

(7) 2 balls are drawn from an urn containing 2 W. 3 R. & 4 B balls one by one without replacement. What is the p that atleast one ball is red.

A → 1 Red ball, B → 2 Red balls

$P(A) = \frac{3_{C_1}}{9_{C}} \cdot \frac{6_{C_1}}{8_{C_1}}$ $P(B) = \frac{3_{C_1}}{9_{C_1}} \cdot \frac{2_{C_1}}{8_{C_1}}$

$= \frac{1}{4}$ $= \frac{1}{12}$

$P(A \cup B) = \frac{1}{4} + \frac{1}{12}$

$= \frac{3+1}{12} = \frac{4}{12} = \frac{1}{3}$ not correct

~~2)~~ $P(R_1 \cup R_2) = ?$

$P(R_1) = \frac{3_{C_1}}{9_{C_1}} = \frac{3}{9}$ $P(R_2) = \frac{2_{C_1}}{9_{C_1}}$

$P(R_1 \cap R_2) = \frac{3_{C_1}}{9_{C_1}} \cdot \frac{2_{C_1}}{8_{C_1}}$ $\therefore P(R_1 \cup R_2) = \frac{3}{9} + \frac{3}{9} ...$

$= \frac{42}{72}$

R_1 → getting a red ball (in 1st drawn)

R_2 → getting a red ball (in 2nd drawn)

$P(R_1 \cup R_2) = 1 - \overline{P(R_1 \cup R_2)}$ [1 → total Probability]

$= 1 - P(\overline{R_1} \cap \overline{R_2})$

$= 1 - P(\overline{R_1}) \cdot P(\overline{R_2}/\overline{R_1})$

$P(\overline{R_1}) = \frac{6_{C_1}}{9_{C_1}} = \frac{2}{3}$

$P(\overline{R_2}/\overline{R_1}) = \frac{5_{C_1}}{8_{C_1}} = \frac{5}{8}$

$= 1 - \frac{2}{3} \cdot \frac{5}{8}$

$1 - \frac{5}{12}$

$= \frac{7}{12}$

Required Prob. $P(R_1 \cup R_2) = \frac{7}{12}$

(8) A black & a red die are rolled.

i) Find the C.P. of obtaining a sum > 9 given that Black die resultant in 5

ii) Find the C.P. of obtaining a sum = 8 given that red die resultant in a no. < 4.

i) A = sum > 9 ; B = Black die resultant in 5

C = Sum = 8 , D = Red die resultant in < 4

$A = \{(4,6), (6,4), (5,5), (5,6), (6,5), (6,6)\}$

$B = \{ (5,5), (5,6) \}, (5,1), (5,2) \ldots$

$C = \{ (2,6), (6,2), (3,5), (5,3), (4,4) \}$

$D = \{ \cancel{(4,4)}, (6,2), (6,3), (6,1), (5,1), (5,2), (5,3) \} \to 14$

(i) $P(A \cap B) = P(B) \cdot P(A/B)$

$\Rightarrow \frac{2}{36} = \frac{2}{36} \times \frac{3}{36} P(A/B)$

$\Rightarrow P(A/B)$

(ii) $P(C \cap D) = P(C) \cdot P(C/D)$

$\Rightarrow \frac{2}{36} = \frac{5}{36} \times P(C/D)$

$\Rightarrow P(C/D) = \frac{2}{5}$

i) $P(A \cap B) = P(B) \cdot P(B/A)$

$\Rightarrow \frac{2}{36} = \frac{6}{36} \times P(B/A)$

$\Rightarrow P(B/A) = \frac{1}{3}$

ii) $P(C \cap D) = P(D) \cdot P(C/D)$

$\Rightarrow \frac{2}{36} = \frac{18}{36} \times P(C/D)$

$\Rightarrow P(C/D) = \frac{2}{18} = \frac{1}{9}$

i) $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{2}{6} = \frac{1}{3}$

ii) $P(D/C) = \frac{P(C \cap D)}{P(C)} = \frac{2}{2}$

$P(C/D) = \frac{P(C \cap D)}{P(D)} = \frac{2}{18} = \frac{1}{9}$

Baye's Theorem

An event A corresponds to a no. of exhaustive events $B_1, B_2, \ldots, B_n$. If $P(B_i)$ & $P(A/B_i)$ are given, then

$$\boxed{P(B_i/A) = \frac{P(B_i) \cdot P(A/B_i)}{\Sigma P(B_i) \cdot P(A/B_i)}}$$

Proof - By the multiplication law of Probability,

$P(AB_i) = P(A) \cdot P(B_i/A) = P(B_i) \cdot P(A/B_i)$ — (1)

$P(B_i/A) = \frac{P(B_i)\, P(A/B_i)}{P(A)}$ — (2)

since the event A corresponds to $B_1, B_2, \ldots, B_n$ we have by the addition law of Probability,

$P(A) = P(AB_1) + P(AB_2) + \ldots + P(AB_n)$

$= \Sigma P(AB_i) = \Sigma P(B_i) \cdot P(A/B_i)$ [By (1)]

Hence from (2) we have,

$$P(B_i/A) = \frac{P(B_i)\, P(A/B_i)}{\Sigma P(B_i)\, P(A/B_i)}$$

(Proved)

That is known as the theorem of inverse probability.

The prob $P(B_i)$, $i=1,2 \ldots n$ are called a priori probabilities as these exist before we get any information from the experiment.

The prob. $P(A/B_i)$, $i=1,2 \ldots n$ are called posteriori probabilities as these are found after the experiment results are known.

Ex-

① Three machines m_1, m_2 & m_3 produce identical items. Of their respective output 5%, 4% & 3% of items are faulty. On a certain day, m_1 has produced 25% of the total o/p, m_2 has produced 30% & m_3 the remainder. An item selected at random is found to be faulty. What is the chances that it was produced by the machine with the highest output?

$A \rightarrow$ A faulty item from any of machine

$B_i \rightarrow$ item drawn at random produced by m_i

$B_1 \rightarrow$ item I

$B_2 \rightarrow$ item II

$B_3 \rightarrow$ item III

$P(B_1) = 25\% = 0.25$

$P(B_2) = 30\% = 0.30$

$P(B_3) = 45\% = 0.45$ $[0.25+0.30+0.45=1]$

We have to get, $P(B_3/A) = ?$

Now, we know, $P(A/B_1) = 0.05$ ($\because 5\%$)

$P(A/B_2) = 0.04$ ($\because 4\%$)

$P(A/B_3) = 0.03$ ($\because 3\%$)

By bayes's theorem,

$$P(B_3/A) = \frac{P(A/B_3) \cdot P(B_3)}{P(A/B_1) \cdot P(B_1) + P(A/B_2) \cdot P(B_2) + P(A/B_3) \cdot P(B_3)}$$

$$= \frac{0.03 \times 0.45}{0.25 \times 0.05 + 0.30 \times 0.04 + 0.45 \times 0.03}$$

$= 0.355$

(2) There are 3 bags, 1st containing 1W, 2R, 3G balls, 2nd 2W, 3R & 1G balls & 3rd 3W, 1R & 2G balls. 2 balls are drawn from a bag chosen at random. These are found to be one W & one R. Find the prob. that the balls so drawn came from the second bag.

~~1st bag total =~~

1st bag (bag $-b_1$)	2nd bag (bag $-b_2$)	3rd bag (bag $-b_3$)
1W, 2R, 3G	2W, 3R, 1G	3W, 1R, 2G
total = 6 balls	total = 6 balls	total = 6 balls

~~A → 2 balls are 1W & 1R~~

B_i → ~~2 Balls~~ Bag is chosen ~~are came from~~ ~~bag~~ b_i

B_1 → 2 Balls chosen came from bag 1 (b_1)

B_2 → ... 2 (b_2)

B_3 → ... 3 (b_3)

we have to get $P(B_2/A) = ?$, Now, $P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$

Now, $P(A/B_1) = \frac{^2C_1 \times {}^1C_1}{^6C_2} = \frac{2}{15}$

$P(A/B_2) = \frac{^2C_1 \times {}^3C_1}{^6C_2} = \frac{6}{15} = \frac{2}{5}$

$P(A/B_3) = \frac{^3C_1 \times {}^1C_1}{^6C_2} = \frac{3}{15} = \frac{1}{5}$

Now,

$P(A/B_1) =$

$P(A/B_2) =$

$P(A/B_3) =$

A → 2 balls are 1W & 1R

B_i → b_i bag is chosen $[i = 1, 2, 3]$

B_1 → 1st bag is chosen

B_2 → 2nd

B_3 → 3rd

We have to get $P(B_2/A) = ?$

Now, $P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$

Now, $P(A/B_1) = \frac{^2C_1 \times {}^1C_1}{^6C_2} = \frac{2}{15}$

$P(A/B_2) = \frac{^2C_1 \times {}^3C_1}{^6C_2} = \frac{6}{15} = \frac{2}{5}$

$P(A/B_3) = \frac{^3C_1 \times {}^1C_1}{^6C_2} = \frac{3}{15} = \frac{1}{5}$

∴ By Bayes theorem

$$P(B_2/A) = \frac{P(A/B_2)\cdot P(B_2)}{P(A/B_1)\cdot P(B_1) + P(A/B_2)\cdot P(B_2) + P(A/B_3)\cdot P(B_3)}$$

$$= \frac{\frac{6}{15} + \frac{1}{3}}{\frac{1}{3}\left[\frac{2}{15} + \frac{6}{15} + \frac{3}{15}\right]}$$

$$= \frac{\frac{6}{15}}{\frac{11}{15}}$$

$$= \frac{6}{11}$$

<u>multiplication rule (theorem)</u>

According to the multiplication rule of Prob., the Prob. of occurrence of both the events A & B is equal to the product of the probability of B occurring & the conditional Probability that event A occurring given that event B occurring.

$\Rightarrow P(A \cap B) = P(B) \cdot P(A/B)$ → If A & B are dependent

$\Rightarrow P(A \cap B) = P(A) \cdot P(B)$ → If A & B are independent

Proof- we know that the conditional prob. of event A given that B has occured is denoted by $P(A/B)$

Now $P(A/B) = \frac{P(A \cap B)}{P(B)}$, where $P(B) \neq 0$

$\Rightarrow P(A \cap B) = P(B) \cdot P(A/B)$ — ①

Similarly, $P(B/A) = \frac{P(B \cap A)}{P(A)}$, where $P(A) \neq 0$

$\therefore \Rightarrow P(A \cap B) = P(A) \cdot P(B/A)$ — ②

$[\because P(A \cap B) = P(B \cap A)]$

$\therefore$ from ① & ② we get,

$P(A \cap B) = P(B) \cdot P(A/B) = P(A) \cdot P(B/A)$

where $P(A) \neq 0$, $P(B) \neq 0$ → this is known as multiplication rules of prob.

$\therefore P(A \cap B) = P(A) \cdot P(B)$

If A & B are independent then it will be,

As, $P(B/A) = P(B)$.

$\therefore P(A \cap B) = P(A) \cdot P(B/A) = P(A) \cdot P(B)$ (proved)

④