f_i = class frequency

x_i = class mark

$\bar{x}$ = average ;

ie $\bar{x} = \dfrac{\Sigma(fx)}{\Sigma f}$

Example:

The Revenue generated by xxx-Inc was as shown.

Month	Jan	Feb	mar	April	May	Jun
Revenue (USD)	100-200	200-300	300-400	400-600	600-800	800-1200
units	4	6	2	3	8	2

Determine the mean absolute Deviation.

Recall:

$$M.D = \frac{\Sigma\left[f \cdot |x_i - \bar{x}|\right]}{\Sigma f}$$

Revenue USD	f	x_i	fx	$\bar{x}$	$(x_i - \bar{x})$	$\lvert x_i - \bar{x}\rvert$	$f \cdot \lvert x_i - \bar{x}\rvert$
100-200	4	$\frac{100+200}{2}$ 150	600	476	-326	326	1304

200-300	6	250	1500	476	-226	226	1356
300-400	2	350	700	476	-126	126	252
400-600	3	500	1500	476	24	24	72
600-800	8	700	5600	476	224	224	1792
800-1200	2	1000	2000	476	524	524	1048
	25		11900				5824

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{11900}{25} = 476$$

$$MAD = \frac{5824}{25} = 232.96$$

$$MAD = 232.96 \text{ USD}$$

class 6

Friday

July 19, 2024

Variance and standard deviation

✓ Variance refers to the mean of the square deviation btn the data set values and their average;

✓ Standard Deviation refers to the square root of Variance; it also gives the measure for the risks in a given variable [Investment]

Case I: Variance for ungrouped Data

✓ This is given by: σ^2

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

OR:

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

$\bar{x}^2$

μ^2

where:

x_i = data set values

$\bar{x}$ = mean of data set

N = size of the data set

Example:

The weights of 5 participants at a course were as noted:

Weight (Kg): 80, 65, 70, 75, 90

Determine the variance and standard deviation in the weight.

Recall:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N}$$

x_i Kg	$\bar{x}$ Kg	$(x_i - \bar{x})$ Kg	$(x_i - \bar{x})^2$ Kg^2
80	76	4	16
65	76	-11	121 $(-11)^2$
70	76	-6	36
75	76	-1	1
90	76	14	196
380		Total	370 ✓✓

$$\bar{x} = \frac{\sum x}{N} = 380/5$$

$$\bar{x} = 76$$

$Var(x) = 370/5 = 74$

$$\underline{\sigma^2 = 74 \, (Kg)^2}$$

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

$$S.D = \sqrt{74} = \pm 8.6 \text{ Kg}$$

Case II: Variance for Grouped Data

✓ Here we shall apply the formular:

$$Var(x), \sigma^2 = \frac{\sum_{i=1}^{n}\left[f_i \cdot (\overset{100}{x} - \bar{x})^2\right]}{\sum_{i=1}^{n} f_i \rightarrow N}$$

OR

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

where;

f_i = class frequency

x_i = mid mark (value)

Example:

The sales to 50 clients were as noted:

Sales (00) units

~~41~~	~~42~~	~~52~~	~~55~~	~~60~~
~~72~~	~~74~~	~~80~~	~~85~~	~~90~~
~~89~~	~~45~~	~~75~~	~~84~~	~~80~~*
~~84~~	~~89~~	~~70~~	~~72~~	~~90~~
~~48~~	~~69~~	~~66~~	~~74~~	~~100~~
~~110~~	~~90~~	~~42~~	~~81~~	~~72~~
~~52~~	~~82~~	~~77~~	~~44~~	~~93~~
~~69~~	~~86~~	~~74~~	~~52~~	~~70~~
~~42~~	~~58~~	~~66~~	~~78~~	~~93~~
~~80~~	~~70~~	~~82~~	~~105~~	~~109~~

starting with

$40 < x \leq 50$

$50 < x \leq 6$

(a) Using a uniform class width of 10 starting with $40 < x \leq 50$, construct a frequency distribution

(b) Determine the variance and standard deviation in the sales

≤ 10

(a) sales	Tally	frequency
$40 < x \leq 50$	~~||||~~ ||	7

$50 < x \leq 60$	卌 |	6
$60 < x \leq 70$	卌 ||	7
$70 < x \leq 80$	卌 卌 ||	12
$80 < x \leq 90$	卌 卌 |	11
$90 < x \leq 100$	||||	4
$100 < x \leq 110$	|||	3

(b) Variance, σ^2

Recall:

$$\sigma^2 = \frac{\overset{'0000}{\sum(f \cdot x^2)}}{\sum f} - \left[\frac{\overset{'00}{\sum(f \cdot x)}}{\sum f}\right]^2$$

sales '00	f	x_i '00 units	fx '00 units	x^2 '0000 units^2	$f \cdot x^2$ '0000 units^2
40 – 50	7	45	315	2025	14175
50 – 60	6	55	330	3025	18150
60 – 70	7	65	455	4225	29575

70 – 80	12	75	900	5625	67500
80 – 90	11	85	935	7225	79475
90 – 100	4	95	380	9025	36100
100 – 110	3	105	315	11025	33075
Total	50		3630		278050

$$Var(x),\ \sigma^2 = \frac{278050}{50} - \left[\frac{3630}{50}\right]^2$$

$$= 5561 - 5270.76$$

$$\sigma^2 = 290.24 \times 10000$$

$$\sigma^2 = 2902400\ [\text{units}]^2$$

standard Deviation = $\sqrt{\text{Variance}}$

$$S.D = \sqrt{2902400} = 1703.64$$

$$\sigma = \pm 1703.64$$

Coefficient of variation: CV

✓ This refers to the ratio of the standard deviation to the mean expressed as a percentage.

i.e. $CV = \frac{\text{standard deviation}}{\text{Mean}} \times 100$ ✓✓

$CV = s/\bar{x} \times 100$; $\sigma/\mu \times 100$

✓ It is used to compare rates of changes in items of different units.

Height Age
m yrs
(0.23m) (0.23yrs)

Example:

The average weight in Co A is known to be 65 Kg with a S.D of 8 Kg ; whereas the average pay was known to be 820 USD with a S.D of 16 USD;
Compare the rates of change in weight & Pay.

① Weight (Kg)	② Pay (USD)
✓ $\bar{x}_1 = 65$	$\bar{x}_2 = 820$
✓ $s_1 = 8$	$s_2 = 16$

$CV = s/\bar{x} \times 100$

$CV_1 = 8/65 \times 100$

$CV_2 = \frac{16}{820} \times 100$

$CV_2 = \underline{1.95}$

$CV_1 = \underline{\underline{12.31\%}}$

Quartile Deviation:

✓ This is defined to be half the difference btn the upper Quartile & the lower Quartile.

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

where Q_3 = upper Quartile
Q_1 = lower Quartile

✓ Some times its referred to as the semi-inter Quartile range.

Class 7

Monday
July 22, 2024

Quartile Coefficient of Deviation / Dispersion

✓ This is defined to be the ratio of the difference of the upper Quartile & lower Quartile to their sum.

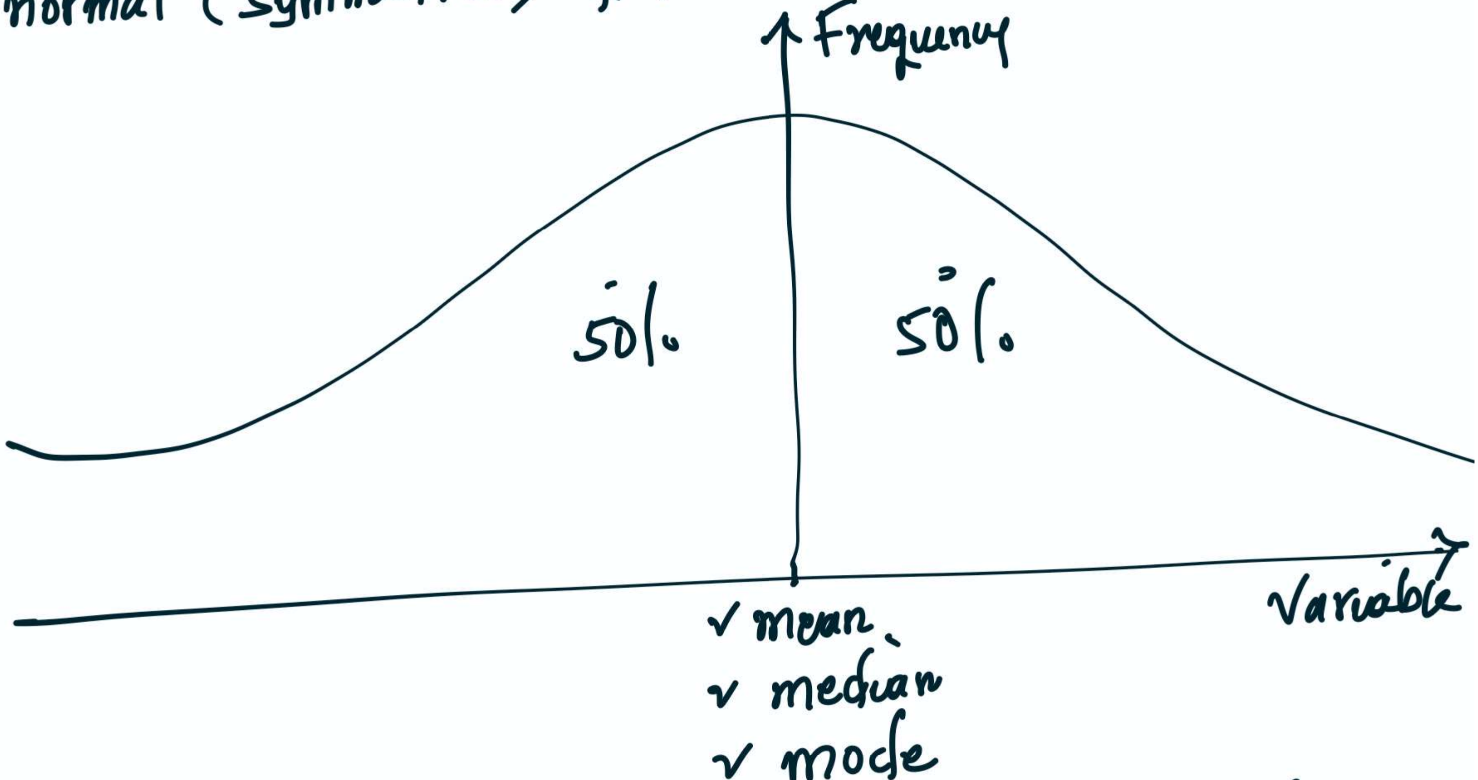
✓ It is given by:

$$QCD = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Q_3 = upper Quartile
Q_1 = Lower Quartile

Skewness:

✓ When a data set is evenly distributed about a given measure of location, then we have a normal (symmetric) distribution.

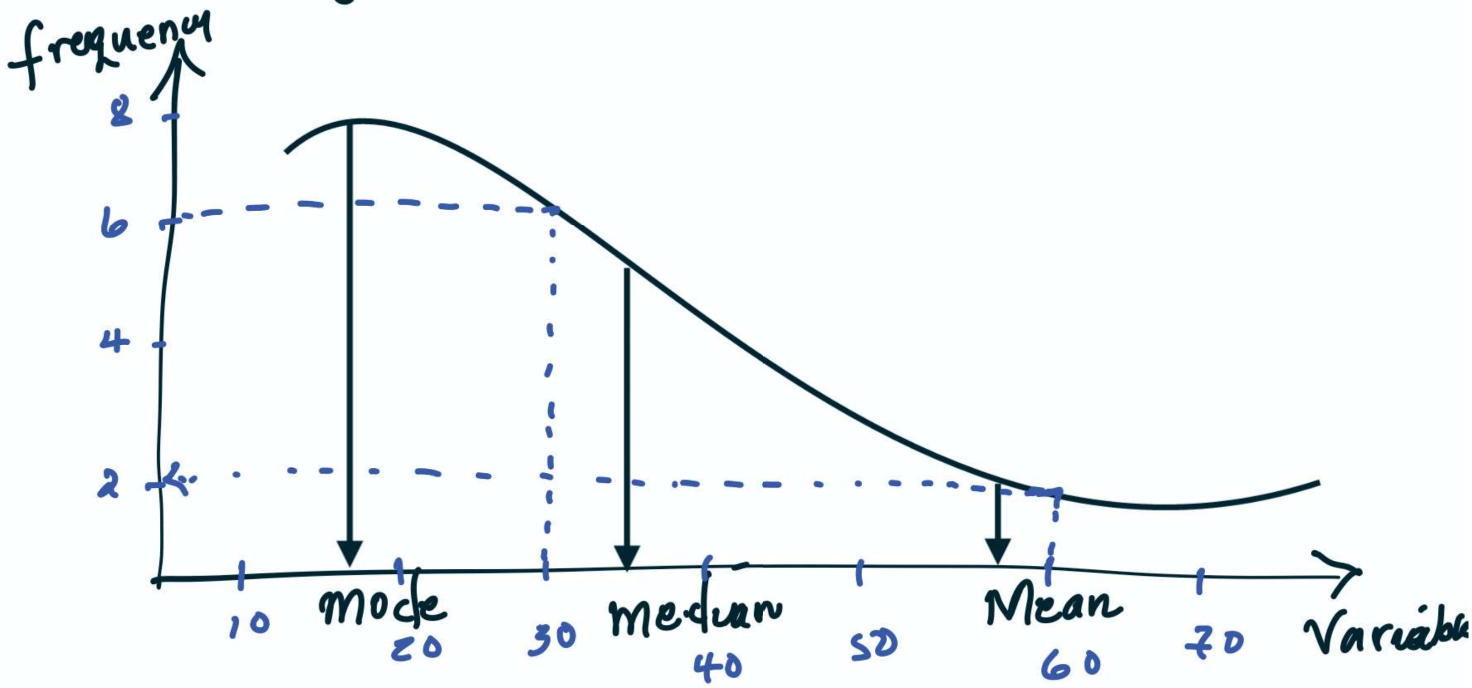


✓ Where the data set is unevenly distributed, then we have a skewed distribution; hence skewness refers to the uneven distribution of the data set about a given measure of location.

✓ There are two types of skewness
- Positive skewness
- Negative skewness

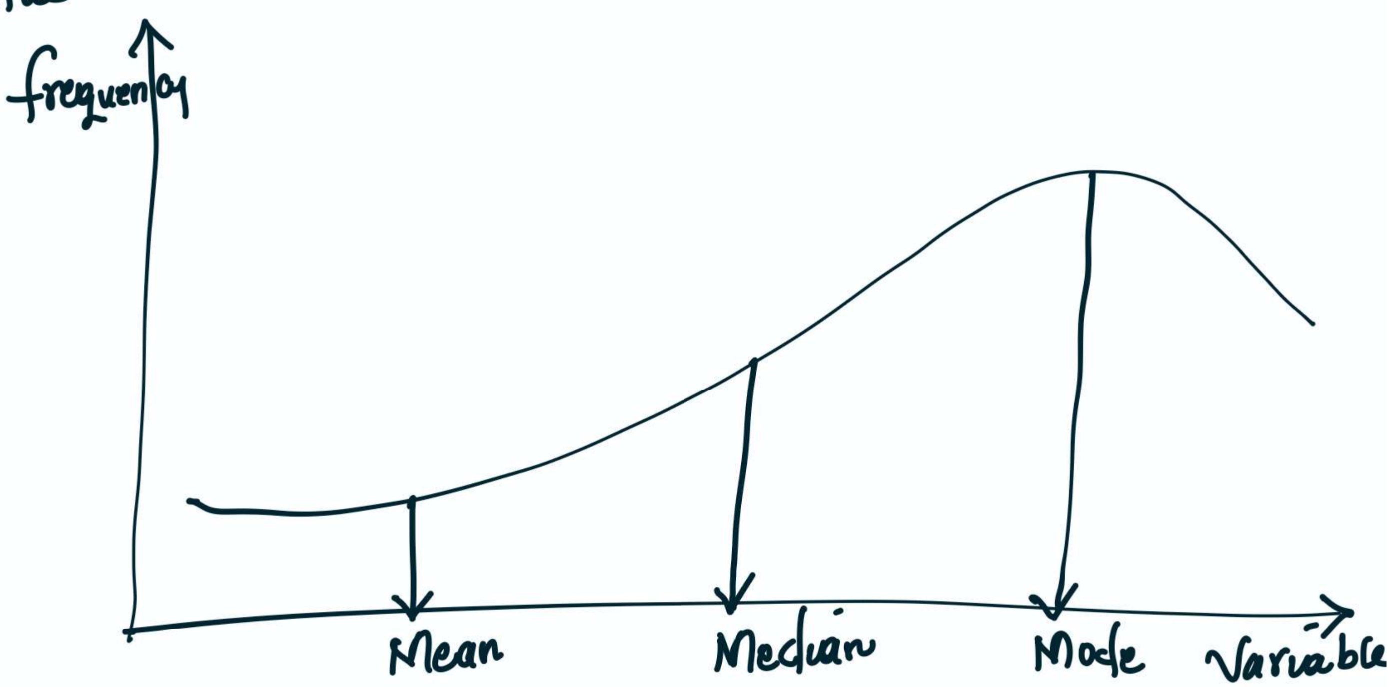
(a) Positive skewness

✓ This occurs when there are few items of a high value; here the distribution has a longer tail to the right.



(b) Negative skewness

✓ This occurs when there few item of very low value



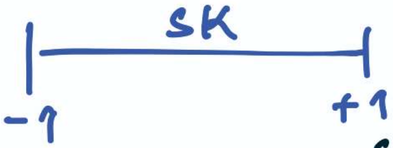
The dsn has a longer tail to the left.

Skewness can be categorised into 3 coefficients

① Karl Pearson's coefficient:

✓ This is given by:

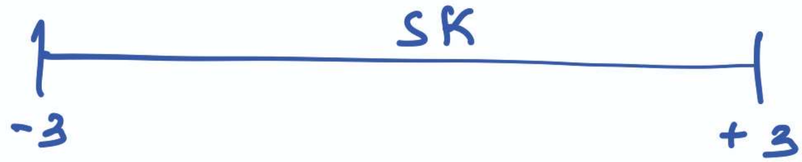
$$SK = \frac{\text{mean} - \text{mode}}{\text{standard Deviation}}$$



$$SK = \frac{\bar{x} - \text{mode}}{\sigma} \;; \quad -1 \leq SK \leq +1$$

OR

$$SK = 3\left[\frac{\text{mean} - \text{median}}{\text{standard Deviation}}\right]$$



$$-3 \leq SK \leq +3$$

② Bowley's coefficient of skewness

✓ This uses Quartiles for skewness; it is given by:

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

③ Kendall's coefficient of skewness

✓ This uses Percentiles for skewness; it is given by:

$$K_{SK} = \frac{P_{90} + P_{10} - 2P_{50}}{P_{90} - P_{10}}$$

Relationship btn the mean, median & Mode

✓ From Pearson's coefficient of skewness, we have:

$$SK = \frac{\text{mean} - \text{mode}}{S.D}$$ ✓✓

$$SK = 3\left[\frac{\text{mean} - \text{median}}{S.D}\right]$$ ✓✓

$$\Rightarrow \frac{\text{mean} - \text{mode}}{\cancel{S.D}} = \frac{3(\text{mean} - \text{median})}{\cancel{S.D}}$$

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

$$\text{mean} - \text{mode} = 3\text{mean} - 3\text{median}$$

$$\text{mean} - 3\text{mean} = -3\text{median} + \text{mode}$$

$$\frac{-\cancel{2}\text{mean}}{-\cancel{2}} = \frac{-3\text{median} + \text{mode}}{-2}$$

$$\text{Mean} = \frac{-3\text{median}}{-2} + \frac{\text{mode}}{-2}$$

$$= \frac{3\text{median}}{2} - \frac{\text{mode}}{2}$$

$$\text{Mean} = \frac{3\text{median} - \text{mode}}{2}$$

Variance & standard Deviation of a sample

Case I: For ungrouped data

✓ The variance for a random sample of variable x shall be given by:

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

OR:

$$Var(x),\ s^2 = \frac{\sum_{i=1}^{n} x^2}{n-1} - \frac{\left(\sum_{i=1}^{n} x\right)^2}{n(n-1)}$$

where:

x_i = sample values

$\bar{x}$ = sample mean

n = sample size

Case II: For Grouped Data.

✓ Here the class mark is required & then we apply the formular:

$$Var(x),\ s^2 = \frac{\sum_{i=1}^{n}\left[f_i \cdot (x_i - \bar{x})^2\right]}{\sum f - 1}$$ *

OR

$$\boxed{Var(x),\ s^2 = \frac{\sum_{i=1}^{n}(f_i \cdot x_i^2)}{\sum_{i=1}^{n} f_i - 1} - \frac{\left[\sum_{i=1}^{n}(f_i \cdot x_i)\right]^2}{\sum f_i (\sum f_i - 1)}}$$

where;

x_i = class mark

f_i = class frequency

$\bar{x}$ = mean; $= \dfrac{\sum(f \cdot x)}{\sum f}$.

Variance of x with Probabilities

✓ Here we apply the formular:

$$Var(x),\ \sigma^2 = \sum_{i=1}^{n}\left[x_i^2 \cdot P_i(x)\right] - \left[\sum_{i=1}^{n}\left[x_i \cdot P_i(x)\right]\right]^2$$

where;

x_i = data set values

$P_i(x)$ = probability of the random variable x

Example:

The weights of participants were as noted together with their likelihood.

Participant	A	B	C	D	E
Weight (kg)	80	85	60	75	95
Expectation, P(x)	0.25	0.15	0.30	0.20	0.1

Determine the variance and standard deviation in the weight.

Recall:

$$Var(x), \sigma^2 = \sum[x^2 \cdot P(x)] - [\sum[x \cdot P(x)]]^2$$

x_i	$P(x)$	x_i^2	$x \cdot P(x)$	$x^2 \cdot P(x)$
80	0.25	6400	20	1600
85	0.15	7225	12.75	1083.75
60	0.30	3600	18	1080
75	0.20	5625	15	1125
95	0.10	9025	9.5	902.5
	Total		75.25	5791.25

$$Var(x), \sigma^2 = 5791.25 - (75.25)^2$$

$$\sigma^2 = 128.6875$$

$$\sigma^2 = 128.69\,(Kg)^2$$

Recall:

$$S.D, \sigma = \sqrt{\text{Variance}}$$

$$\sigma = \sqrt{128.6875} \rightarrow \sqrt{128.69}$$

$$\sigma = \pm 11.34$$

$$\sigma = \pm 11.34\,Kg.$$

Probability:

- "may occur"
- chance
- luck
- Possibility
- Likelihood
- unsure
- uncertain
- 50/50

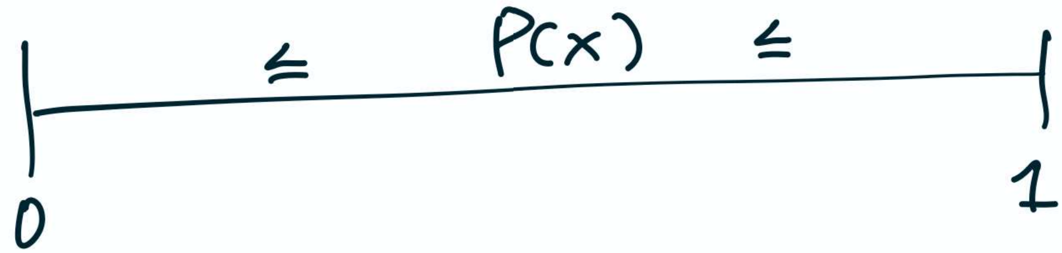
- This is concept that is concerned with the likelihood that a given event will occur.
- The probability of a random variable, x shall be given by:

$$P(x) = \frac{\text{Number of Possible outcomes of } x}{\text{Total sample space}}$$

$$\text{ie } P(x) = \frac{n(x)}{n(\mathcal{E})}$$

Note that the P(x) ranges from 0 to 1

ie $0 \leq P(x) \leq 1$



Example:

A bag contains 4 oranges, 6 mangoes & 5 apples; Determine the probability of picking an apple.

$$P(apple) = \frac{N^{\underline{o}} \text{ of apples}}{\text{Total } n^{\underline{o}} \text{ fruits}}$$

$$n(S) = 4 + 6 + 5$$
$$= 15$$

$$n(A) = 5$$

$$P(A) = \frac{n(A)}{n(S)} = 5/15 = 0.33$$

$$P(A) = \underline{\underline{0.33}}$$

Rules of Probability:

(a) The addition law of mutually exclusive events

✓ This states that the probability of one or other mutually exclusive events, shall be the sum of their individual probabilities

ie $P(A \text{ or } B) = P(A \cup B)$

$$= P(A) + P(B) - P(A \cap B)$$

where:

A, B = mutually exclusive events.

for 3 events:

$$\boxed{\begin{aligned} P(A \text{ or } B \text{ or } C) &= P(A \cup B \cup C) = \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ &\quad - P(B \cap C) + P(A \cap B \cap C) \end{aligned}}$$

Two or more events are said to be mutually exclusive if the occurrence of one affects the occurrence of the other. ie these are events that can't happen at the same time.

Example:

When a dice is thrown, determine the probability of having a 3 or a 4

$$P(3 \text{ or } 4) = P(3 \cup 4)$$

$$= P(3) + P(4) - P(3 \cap 4)$$

$$= 1/6 + 1/6 - 0$$

$$P(3 \text{ or } 4) = 2/6 = 1/3$$

$$P(3 \text{ or } 4) = \underline{\underline{0.333}}$$