Probability ($0 \le P(A) \le 1$)

(1) Addition ~~Tho~~ Theorem

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

If they are mutually exclusive

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

(2) Conditional Probability

A given B $\Rightarrow$ $P(A/B) = \dfrac{P(A \cap B)}{P(B)}$

B given A $\Rightarrow$ $P(B/A) = \dfrac{P(A \cap B)}{P(A)}$

(3) Multiplication Theorem

$$P(A \cap B) = P(A/B) \times P(B)$$

OR

$$= P(B/A) \times P(A)$$

(4) Independent Events

If A & B are independent events

$$P(A \cap B) = P(A) \times P(B)$$

(5) Probability of not occuring an event

$$\boxed{P(\bar{A}) = P(A') = P(A^c) = 1 - P(A)}$$

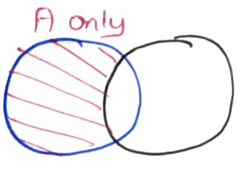
(6) De-Morgan's Laws

$$P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B})$$

$$P(\overline{A \cap B}) = P(\bar{A} \cup \bar{B})$$

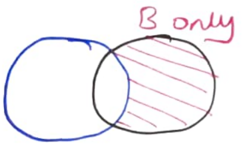
(7) (4.) P(A only)

$$\boxed{P(A \cap \bar{B}) = P(A) - P(A \cap B)}$$



(b) P(B only)

$$\boxed{P(\bar{A} \cap B) = P(B) - P(A \cap B)}$$

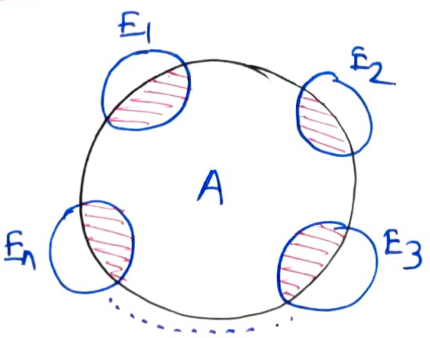


(c) P(A & B happening once)

$$\boxed{P(A \cup B) - P(A \cap B)}$$

(8) Baye's Theorem

$$\boxed{P(E_i/A) = \frac{P(A/E_i) \times P(E_i)}{\Sigma P(A/E_i) \times P(E_i)}}$$



* Discrete Random Variable

Denoted by $\Rightarrow P(x)$ (Probability Mass Function)

C.D.F $\Rightarrow F(X=x) = \sum_{x=-\infty}^{x} P(x)$; $\sum_{x=-\infty}^{\infty} P(x) = 1$

* Continuous Random Variable

Denoted by $\Rightarrow$ ~~f(x~~ $f(x)$; $\int_{-\infty}^{\infty} f(x) = 1$

CDF $\Rightarrow F(X=x) = \int_{-\infty}^{x} f(x) \cdot dx$

* Joint Probability

Discrete R.V

(i) Denoted by $P(x,y)$

(ii) Marginal probability

$$P(x) = \sum_{y=-\infty}^{\infty} P(x,y)$$

$$P(y) = \sum_{x=-\infty}^{\infty} P(x,y)$$

Continuous R.V

(i) Denoted by $f(x,y)$

(ii) Marginal Probability

$$f(x) = \int_{y=-\infty}^{\infty} f(x,y) \cdot dy$$

$$f(y) = \int_{x=-\infty}^{\infty} f(x,y) \cdot dx$$

NOTE: ~~It~~ Even if the question doesn't asks to find marginal probability find it for every such question

* Stochastic Independence $[P(A \cap B) = P(A) \times P(B)]$

$$P(x, y) = P(x) \cdot P(y)$$

$$F(x, y) = f(x) \cdot F(y)$$

* Conditional Probability $[P(A/B) = P(A \cap B)/P(B)]$

Discrete R.V

(i) $P(x/y) = \dfrac{P(x, y)}{P(y)}$

OR

$P(y/x) = \dfrac{P(x, y)}{P(x)}$

Continuous R.V

(ii) $f(x/y) = \dfrac{f(x, y)}{f(y)}$

OR

$f(y/x) = \dfrac{f(x, y)}{f(x)}$

* CDF To PDF Conversion in continuous RV

$$\begin{tikzcd} \text{PDF} \; [f(x)] \; [f(x,y)] \arrow[r, bend left, "\int \text{(Integration)}"] & \text{CDF} \; [F(x)] \; [F(x,y)] \arrow[l, bend left, "\delta \text{(Derivatives)}"] \end{tikzcd}$$

(i) $f(x) = \dfrac{d}{dx} F(x)$ & $F(x) = \displaystyle\int_{x=0}^{x} f(x) \cdot dx$

(ii) $f(x, y) = \dfrac{d}{dx} \cdot \dfrac{d}{dy} F(x, y)$ & $F(x, y) = \displaystyle\int_{x=0}^{x} \int_{y=0}^{y} f(x, y) \cdot dx \cdot dy$

i.e $= \displaystyle\int_{x=0}^{x} f(x) \cdot dx \cdot \int_{y=0}^{y} f(y) \cdot dy$

* Expectation

~~Ra~~ Discrete RV = $\sum x \cdot p(x)$

Continuous RV = $\int_{-\infty}^{\infty} x \cdot f(x) \cdot dx$

* Variance

$$V(x) = E(x^2) - [E(x)]^2$$

where $E(x^2) = \sum x^2 \cdot P(x)$

* Properties of E(X) & V(X)

(i) $E(x \pm y) = E(x) \pm E(y)$

(ii) $E(ax \pm by) = aE(x) \pm bE(y)$

(iii) $E(ax \pm b) = aE(x) \pm b$

(i) $V(x \pm y) = V(x) \pm V(y)$

(ii) $V(ax \pm by) = a^2V(x) \pm b^2V(y)$

(iii) $V(ax \pm b) = a^2V(x)$

* Probability Distribution

(i) Bernaulli's Distribution

(a) Mean of of Bernaulli's ~~Dist~~ Distribution = p

(b) Variance of Bernaulli's Distribution = pq.

(ii) Binomial Distribution

$$\boxed{\begin{aligned} P(X=x) &= {}^nC_x \cdot p^x \cdot q^{n-x} \quad , x = 0,1,2,\ldots,n \\ &= 0 \quad , \text{otherwise} \end{aligned}}$$

Here, p = success probability
$q = 1-p$ (failure probability)

a.) Mean of Binomial Distribution = np

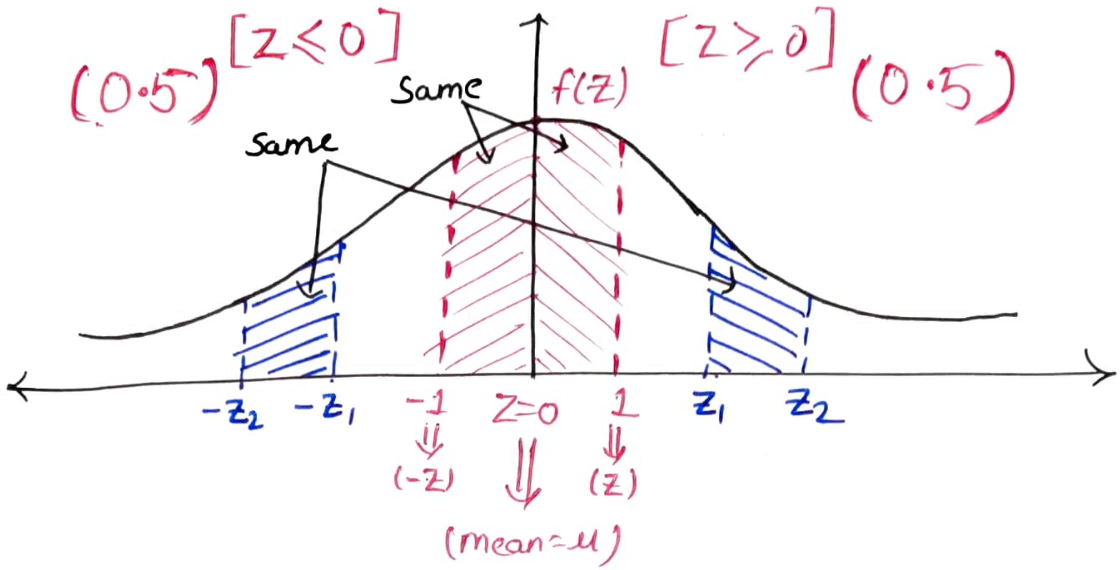
b.) Variance of Binomial Distribution = npq

(iii) Poisson's Distribution

$\boxed{\text{Mean} = \lambda = np}$ where n = no of observations
p = ~~success~~ Given probability

$$\therefore \boxed{\begin{aligned} P(X=x) &= \frac{e^{-\lambda} \cdot \lambda^x}{x!} \quad , \text{where } x = 0,1,2,\ldots\ldots \\ &= 0 \quad , \text{otherwise} \end{aligned}}$$

* Standard Normal Distribution



Properties

(i) Mean of Standard Normal value $= 0$

(ii) Standard Deviation of SNV $= 1$

(iii) The graph of $f(Z)$ is bell shaped having peak at $Z = 0$.

(iv) The graph of $f(Z)$ is symmetric about the vertical axis (i.e $Z = 0$)

(v) $P(Z \leq 0) = P(Z \geq 0) = 0.5$

(vi) $P(-z \leq Z \leq 0) = P(0 \leq Z \leq z)$

(vii) $P(-z_2 \leq Z \leq -z_1) = P(z_1 \leq Z \leq z_2)$

* To Solve the Problems in SNV

We cannot solve the problems on Normal Variable as it is

Therefore, we transform given normal variable to Standard Normal Value

~~tie~~ i.e $Z = \frac{X-\mu}{\sigma}$

~~Eg~~ Eg:- $P(X>65) = P\left(\frac{X-\mu}{\sigma} > \frac{65-\mu}{\sigma}\right)$

$= P\left(Z > \frac{65-60}{5}\right)$ .... [If $\mu=60$ & $\sigma=5$]

$= P(Z>1)$

conversion from X (Normal variable) to Z (Standard Normal variable)

Measures / Data type	Mean	Median	Mode
Simple Raw Data	$\boxed{\frac{\Sigma x_i}{n}}$ (No Arrangement)	(Arrangement Required: Sorted Data) (i) If n is odd $\left(\frac{n+1}{2}\right)^{th}$ value (ii) If n is even $\frac{\left(\frac{n}{2}\right)^{th}_{value} + \left(\frac{n}{2}+1\right)^{th}_{value}}{2}$	Value having most no of repetitions
Discrete Data	$\boxed{\frac{\Sigma f_i \cdot x_i}{\Sigma f_i}}$ (No Arrangement)	CUMMULATIVE FREQUENCY (CF) Total CF = N (i) If N = ODD $\left(\frac{N+1}{2}\right)^{th}$ value (ii) If N = EVEN (Same as above) CHECK FOR CF having Highest Ceiling value	Same as Above
Grouped Frequency Data	(i) $\frac{\Sigma f_i x_i}{\Sigma f_i}$ where x_i is classmark $\left[x_i = \frac{l_i + u_i}{2}\right]$ (ii) Assumed Mean Method Assumed Mean = a Class width = c Define $u_i \Rightarrow u_i = \frac{x_i - a}{c}$ $\boxed{\text{Mean} = a + c \times \frac{\Sigma f_i u_i}{\Sigma f_i}}$	Step 1:- Total CF ⇒ N Step 2:- Calculate N/2 value Step 3:- Check the N/2 value in the C.F column and take the closest ceiling value of it. The interval having N/2 value will be the median class. Step 4:- We get, Median class $= (l_1 - l_2)$ Median Freq = f & Pre-Median C.F = F $\boxed{\text{MEDIAN} = l_1 + \left(\frac{l_2 - l_1}{f}\right)\left(\frac{N}{2} - F\right)}$	Step 1:- Taking Highest frequency value into consideration from the table Step 2:- Interval having highest frequency will be modal class • f = Freq of modal class • f_1 = Freq of pre-modal class • f_2 = Freq of post-modal class • $d_1 = f - f_1$ & $d_2 = f - f_2$ $\boxed{\text{MODE} = l_1 + \left(\frac{d_1}{d_1 + d_2}\right) \cdot (l_2 - l_1)}$

Quartiles

$$\boxed{Q_1 = l_1 + \left(\frac{l_2 - l_1}{f}\right) \cdot \left(\frac{N}{4} - F\right)} \Rightarrow \text{Quartile 1}$$

$$\boxed{Q_2 = l_1 + \left(\frac{l_2 - l_1}{f}\right) \cdot \left(\frac{N}{2} - F\right)} \Rightarrow \text{Quartile 2}$$

$$\boxed{Q_3 = l_1 + \left(\frac{l_2 - l_1}{f}\right) \cdot \left(\frac{3N}{4} - F\right)} \Rightarrow \text{Quartile 3}$$

Relation between Mean, Median & Mode

$$\boxed{\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median})}$$

If class is not continuous ?

For eg:-

Classes	1-10	11-20	21-30	31-40	41-50
Freq	...	.....	...	.....	...

$$\boxed{\text{Class Difference} = \frac{\text{Lower limit of next class} - \text{upper Limit of previous class}}{2}}$$

So in every class interval,
(i) ~~We~~ Add the class difference value to the upper limit
(ii) Subtract the class difference value from the lower limit

∴

Classes	0·5 – 10·5	10·5 – 20·5	20·5 – 30·5	30·5 – 40·5	40·5 – 50·5
Freq.	....	....	....	....	...

Combined Mean $= \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$ i.e $\dfrac{\sum n_i x_i}{\sum n_i}$

* Measures of Dispersion

I) Absolute Measures

(1) Range : High − Low = Max value − Min value

(2) Mean Deviation from M (M = Mean, Median & Mode)

AND

Standard Deviation (σ)

Measures / Data types	Mean Deviation	Standard Deviation ($\bar{x}$ = mean)
(i) Raw Data	$\frac{\Sigma(\lvert x_i - M \rvert)}{n}$ Where, M = Mean, median, mode n = Total observations	$\sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$ OR $\sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$
(ii) Discrete Frequency Data	$\frac{\Sigma f_i(\lvert x_i - M \rvert)}{\Sigma f_i}$ Where M = mean, median, mode Σf_i = Total observations	$\sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}}$ OR $\sqrt{\frac{\Sigma f_i \cdot x_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i x_i}{\Sigma f_i}\right)^2}$
(iii) Grouped Frequency Data	$\frac{\Sigma f_i(\lvert x_i - M \rvert)}{\Sigma f_i}$ Where M = mean, median, mode x_i = class-mark	$\sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}}$ OR $\sqrt{\frac{\Sigma f_i x_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i x_i}{\Sigma f_i}\right)^2}$ where x_i = classmark

(3) Quartile Range

$Q_3 - Q_1 \Rightarrow$ Quartile Range

Semi-Inter Quartile Range $\Rightarrow \dfrac{Q_3 - Q_1}{2}$
(Also known as Quartile Deviation)

(II) Relative Measures

(1) Coefficient of Range $= \dfrac{H-L}{H+L}$

(2) Coefficient of M.D $= \dfrac{MD}{Mean}$

(3) Coefficient of QD $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

(4) Coefficient of Variance $= \dfrac{S.D}{Mean}$
(Standard Deviation)

(III) Change of scale method in S.D.

$\sigma_x = c \cdot \sigma_u$ where c = class width

$$= c \cdot \sqrt{\frac{\Sigma f_i u_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right)^2}$$

(IV) Standard Deviation for combined groups

$$\sigma = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$ where $d_1 = \bar{x} - \bar{x}_1$

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

SKEWNESS

(1) Karl Pearson's coefficient of skewness

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

(2) Bowley's Coefficient of skewness

$$\frac{Q_3 + Q_1 - 2M}{S.D}$$

Correlation

(1) Karl Pearson's coefficient of Correlation (r)

$$r = \frac{\frac{\Sigma x_i y_i}{n} - \left(\frac{\Sigma x_i}{n} \times \frac{\Sigma y_i}{n}\right)}{\sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2} \cdot \sqrt{\frac{\Sigma y_i^2}{n} - \left(\frac{\Sigma y_i}{n}\right)^2}}$$

(2) Spearman's Rank Correlation (R)

$$R = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$$

where $d_i = R_1 - R_2$ OR $R_2 - R_1$

n = n.o of observations

Types of Questions

1.) Ranks are given

2.) Ranks are not given

3.) Ranks are repetitive IMP

~~4) Ranks are more than 2~~

4.) More than 2 observations are given IMP

Ranks are repetitive

Step 1:- If the ranks are repeated then find the average rank of the n.o of occurence for that dataset

$$\text{Avg rank} = \frac{\text{Sum of all the ranks of repeated data}}{\text{N.o of occurences of the data}}$$

Step 2:- If there is repetition then find the correction factor for each repetition of x & y dataset

If rank is repeated "m" times then correction factor is given by

$$\text{Correction Factor} = \frac{m(m^2-1)}{12}$$

Step 3:- Find the corrected Σd_i^2

$$\therefore \Sigma d_i^2{}_{(\text{corrected})} = \Sigma d_i^2 + \Sigma \text{ All the corrected factors}$$

Step 4:- Input all the values in the formula

$$R = 1 - \frac{6\Sigma d_i^2{}_{(\text{corrected})}}{n(n^2-1)}$$

Regression

(i) Regression of y on x

$$\boxed{y - \bar{y} = b_{yx}(x - \bar{x})}$$

where $\bar{x} = \frac{\sum x_i}{n}$; $\bar{y} = \frac{\sum y_i}{n}$

$$b_{yx} = \frac{\frac{\sum x_i y_i}{n} - \left(\frac{\sum x_i}{n} \times \frac{\sum y_i}{n}\right)}{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$$

(ii) Regression of x on y.

$$\boxed{x - \bar{x} = b_{xy}(y - \bar{y})}$$

where,

$$b_{xy} = \frac{\frac{\sum x_i y_i}{n} - \left(\frac{\sum x_i}{n} \times \frac{\sum y_i}{n}\right)}{\frac{\sum y_i^2}{n} - \left(\frac{\sum y_i}{n}\right)^2}$$

NOTE :- $\boxed{b_{xy} \times b_{yx} = r^2}$

where r is K.P coefficient of correlation

Types of Question

1) Given Regression Lines (Equations)

To find

(i) Mean values of $\bar{x}$ & $\bar{y}$

[Solve the two linear equations & Find values of x & y. These are the values for $\bar{x}$ & $\bar{y}$]

2) Relation between Regression coefficient, Correlation coefficient & standard deviation.

$$\boxed{b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} \quad \& \quad b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}}$$

3) Find the Regression Equations

Using: (1) Regression of y on x &

(2) Regression of x on y

NOTE :- While calculating coefficient of correlation on x & y,

If b_{yx} & b_{xy} are negative then, r will also be negative

If b_{yx} & b_{xy} are positive then, r will also be positive

*Testing of Hypothesis

(1) Small sample Test / Student Test / T-test ($n < 30$)

Null Hypothesis (H_0) : Population mean = μ

Alternate Hypothesis (H_1) : Population mean $\neq \mu$

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}}$$

where μ = Assumed mean,

S = S.D

$\sqrt{n-1}$ = Degree of freedom

If: $|t| \leq t_\alpha \Rightarrow$ Null Hypothesis Accepted

Else: $\Rightarrow$ Null Hypothesis Rejected

(2) Large Sample Test / Z-test ($n \geq 30$)

Null & Alternate Hypothesis same as above

$$z = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

If $|z| \leq z_\alpha$ then null hypothesis is accepted

Else: Null hypothesis is rejected

(3) Chi-Square Test

~~Step 1~~ :- (χ_0^2)

Null Hypothesis : Given Data is uniformly distributed

Alternate Hypothesis : (χ_a^2) Given Data is not uniformly distributed

Step 1 :- Sample size $\Rightarrow N$ (Total observations)

N.o of class intervals $\Rightarrow n$

$$\therefore \text{Expected Frequency } (E_i) = \frac{N}{n}$$

Step 2 :-

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Step 3 :- If $\chi_0^2 \leq \chi^2_{\alpha, n-1}$ then,:

Null hypothesis is accepted

Else:

Null hypothesis is rejected