Unit - 2

Regression - Regression analysis is a set of statistical processes for estemating the relationship b/w the variables.

It includes many techniques for modeling & analysing several variables, when the focus is on the relationship b/w a dependent variable & one or more independent variables (predictor).

Properties of Regression :-

1. Regression describes how an independent var. is associated with dependent var.
2. It helps to find the best fit line and estimate one var. based on another variable.
3. In a linear regression line the equation is given by-

$$y = b_0 + b_1 x$$

here, b_0 = constant
b_1 = regression co-efficient.

- The formula for regression co-efficient is given by-

$$\boxed{b_1 = \frac{\Sigma [(x_i - x)] [y_i - y]}{\Sigma (x_i - x)^2}}$$

where x_i & y_i are the observed dataset and x and y are the mean values of the dataset variables.

The regression cofficient of y on x is b_{yx}
The regression coefficient of x on y is b_{xy}

$$\boxed{b_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N \Sigma Y^2 - (\Sigma Y)^2}} \qquad \boxed{b_{yx} = \frac{N \Sigma XY - \Sigma X \Sigma Y}{N \Sigma X^2 - (\Sigma X)^2}}$$

* The point of intersection of two regression lines is mean value of given data.

4. If one of these regression coefficient is greater than one then the other will be less than one.

5. The airthmatic mean of both regression coefficients is greater than or equals to the coefficient of co-relation.

6. The geometric mean b/w two regression coefficient is equals to the co-relation coefficient.

7. If one coefficient is +ve then other is also +ve.

Que. Calculate Regression coefficient & obtain the line of regression for the following data.

x	y	x^2	y^2	xy
1	9	1	81	9
2	8	4	64	16
3	10	9	100	30
4	12	16	144	48
5	11	25	121	55
6	13	36	169	78
7	14	49	196	98
$\Sigma x = 28$	$\Sigma y = 77$	Σx^2 140	Σy^2 875	Σxy 334
$\bar{x} = 4$	$\bar{y} = 11$			

$N = 7$

$\therefore$ coeff. $b_1 = \dfrac{\Sigma[(x_i - \bar{x})(y_i - \bar{y})]}{\Sigma (x_i - \bar{x})^2}$

$y = b_0 + b_1 x$

X on Y

$$b_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma Y^2 - (\Sigma Y)^2}$$

$$b_{xy} = \frac{7\times 334 - 28\times 77}{7\times 875 - (77)^2}$$

$$= \frac{2338 - 2156}{6125 - 5929}$$

$$b_{xy} = \frac{182}{196} = 0.929$$

eq of x on y :-

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

$$(x-4) = 0.929\,(y-11)$$

$$x - 4 = 0.929y - 10.219$$

$$x = 0.929y - 6.219$$

Y on X

$$b_{yx} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma X^2 - (\Sigma X)^2}$$

$$b_{yx} = \frac{7\times 334 - 28\times 77}{7\times 140 - (28)^2}$$

$$= \frac{2338 - 2156}{930 - 784}$$

$$b_{yx} = \frac{182}{196} = 0.929$$

eq. of y on x :-

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

$$(y-11) = 0.929\,(x-4)$$

$$y - 11 = 0.929x - 3.716$$

$$y = 0.929x + 7.284$$

Corelation Co-efficient

$$\boxed{r^2 = b_{xy} \cdot b_{yx}}$$

also, $b_{xy} = r\dfrac{\sigma_x}{\sigma_y}$

$b_{yx} = r\dfrac{\sigma_y}{\sigma_x}$

Que. In a partially destroyed laboratory record of an analysis of corelation data of the following results are found

Variance of $x = 9$

Regression equations :-

$$8x - 10y + 66 = 0 \quad \text{— (1)}$$
$$40x - 18y = 214 \quad \text{— (2)}$$

Find the mean values of x & y. correlation b/w x and y and S.D of Y.

Sol given $\sigma_x^2 = 9 \quad \therefore \sigma_x = 3$

eq
$$8x - 10y + 66 = 0 \quad \text{— (1)}$$
$$40x - 18y = 214 \quad \text{— (2)}$$

$$10y = 8x + 66 \quad \text{using eq (1) } b_{yx}$$
$$y = \frac{8}{10}x + \frac{66}{10}$$

$$y = 0.8x + 6.6$$

also $y = b_{yx} x + b_0$

$$b_{yx} = 0.8$$

Similarly
$$40x = 18y + 214 \quad \text{bxy using eq (2)}$$
$$x = \frac{18}{40}y + \frac{214}{40}$$

$$x = 0.45y + \frac{214}{40}$$
$$\therefore b_{xy} = 0.45$$

$\because \quad r^2 = b_{xy} \cdot b_{yx}$

$r^2 = 0.45 \times 0.8$

$r^2 = 0.36$

$r = \pm\sqrt{0.36}$

$\boxed{r = \pm 0.6}$

* if we take bxy from -① then we should take byx from eq-② not from same equation else your corelation will be '1' that is wrong.

also, $b_{xy} = r \dfrac{\sigma_x}{\sigma_y}$

$\dfrac{18}{40} = \dfrac{3}{5} \times \dfrac{3}{\sigma_y}$

$\sigma_y = \dfrac{8 \times 40}{5 \times 18}$

$\sigma_y = 4$

Multiplying eq-① by then subtracting ① from -②

$40x - 18y = 214$

$40x - 50y = -330$

$- \quad + \quad +$

$32y = 544$

$y = 17$

$\therefore \quad x = 0.45 \times 17 + \dfrac{214}{40} = 13$

thus $\bar{X} = 13$ & $\bar{Y} = 17$ = Ans now bxy using eq - ①

$$8x = 10y - 66$$

$$x = \frac{10}{8}y - \frac{66}{8}$$

$$x = \frac{5}{4}y - \frac{33}{4}$$

$$\therefore b_{xy} = \frac{5}{4}$$

now byx using eq - ②

$$18y = 40x - 214$$

$$y = \frac{40}{18}x - \frac{214}{18}$$

$$y = \frac{20}{9}x - \frac{107}{9}$$

$$\therefore b_{yx} = \frac{20}{9}$$

Now $r^2 = b_{xy} \cdot b_{yx}$

$$r^2 = \frac{5}{4} \times \frac{20}{9}$$

$$r^2 = \frac{25}{9}$$

$$r = \pm\sqrt{\frac{25}{9}}$$

$$\boxed{r = \pm 5/3}$$

V.I

Q for. The following table shows the sales & advertisement expenditure of a firm

	Sales (x)	Advertisement (y)
Mean	40	6
SD	10	1.5

Coefficient of co-relation $r = 0.9$. Now estimate the likely sales for a proposed advertisement expenditure of ₹10cr.

Soln Sales — x $\bar{x} = 40$ $\sigma_x = 10$

Advertisement - y $\bar{y} = 6$ $\sigma_y = 1.5$

Regression eq of x on y

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

here, $b_{xy} = r\dfrac{\sigma_x}{\sigma_y}$

$$\therefore \quad (x-\bar{x}) = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

$$(x-40) = 0.9 \times \frac{10}{1.5}(y-6)$$

$$x-40 = 6(y-6)$$

$$x-40 = 6y-36$$

$$\Rightarrow x = 6y+4$$

Now, for 10cr. on advert. $x = 6(10)+4 = 64$ crore.

Que For 5 pairs of observation in the following results are obtained $\Sigma x = 15$, $\Sigma y = 25$, $\Sigma x^2 = 55$, $\Sigma y^2 = 135$, $\Sigma xy = 83$. Find the equation of lines of regression & Estimate the value of x on the first line when $y = 12$ & value of y on the 2^{nd} line if $x = 8$.

Sol given $\Sigma x = 15$ $\Sigma y = 25$ $n = 5$ $\Sigma XY = 83$

$\Sigma \bar{x} = 3$ $\Sigma \bar{y} = 5$

$\Sigma x^2 = 55$ $\Sigma y^2 = 135$

Reg. eq of x on y. Reg. eq. of y on x

$$(x - \bar{x}) = bxy\,(y - \bar{y}) \qquad (y - \bar{y}) = byx\,(x - \bar{x})$$

$$bxy = \frac{N\Sigma XY - \Sigma X.\Sigma Y}{N\Sigma y^2 - (\Sigma y)^2} \qquad byx = \frac{N\Sigma XY - \Sigma X.\Sigma Y}{N\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{5\times83 - 15\times25}{5\times135 - (25)^2} \qquad = \frac{5\times83 - 15\times25}{5\times55 - (15)^2}$$

$$= \frac{415 - 375}{675 - 625} \qquad = \frac{415 - 375}{275 - 225}$$

$$bxy = \frac{40}{50} = 0.8 \qquad byx = \frac{40}{50} = 0.8$$

Lines $(x-3) = 0.8(y-5)$ Line $(y-5) = 0.8(x-3)$

$x - 3 = 0.8y - 4$ $y - 5 = 0.8x - 2.4$

$x = 0.8y - 1$ $y = 0.8x + 2.6$

for $y = 12$ $x = 0.8\times12 - 1$ for $x = 8$ $y = 0.8\times8 + 2.6$

$x = 8.6$ $y = 9.$

Camlin Page

Date / /

Que.1 The two regression lines are $3x+2y=26$ & $6x+3y=31$ find the co-relation coefficient

Que.2 For the given lines of regression $3x-2y=5$ & $x-4y=7$ find the a) Regression coefficients
b) Coefficient of co-relation

Sol (1) given

$$3x+2y=26 \quad \text{—} (1)$$
$$6x+3y=31 \quad \text{—} (2)$$

finding bxy using eq - (1)

$$3x+2y=26$$
$$3x=-2y+26$$
$$x=\frac{-2}{3}y+\frac{26}{3}$$

$$\therefore \; bxy=-\frac{2}{3}$$

finding byx using eq - (2)

$$6x+3y=31$$
$$3y=-6x+31$$
$$y=\frac{-6}{3}x+\frac{31}{3}$$

$$\therefore \; byx=-2$$

thus $r^2=bxy.byx=-\frac{2}{3}\times -2=\frac{4}{3}$

$$r^2=4/3 \quad \therefore \quad r=\pm\frac{2}{\sqrt{3}}$$

finding bxy using eq - ②

$$6x = -3y + 31$$

$$x = -\frac{3}{6}y + \frac{31}{6}$$

$$\therefore b_{xy} = -\frac{1}{2}$$

finding byx using eq - ①

$$2y = -3x + 26$$

$$y = -\frac{3}{2}x + \frac{26}{2}$$

$$\therefore b_{yx} = -\frac{3}{2}$$

$$\therefore \quad r^2 = b_{xy} \cdot b_{yx}$$

$$r^2 = \frac{-1}{2} \times \frac{-3}{2}$$

$$r^2 = \frac{3}{4}$$

$$r = \pm \frac{\sqrt{3}}{2}$$

(2) given $3x - 2y = 5 \quad —(1)$

$x - 4y = 7 \quad —(2)$

Find a) Regression coefficients

b) Coefficient of correlation

finding bxy using eq-(1)

$3x - 2y = 5$

$3x = 2y + 5$

$x = \frac{2}{3}y + \frac{5}{3}$

$\therefore b_{xy} = \frac{2}{3}$

finding byx using eq-(2)

$x - 4y = 7$

$\therefore 4y = x - 7$

$y = \frac{1}{4}x - \frac{7}{4}$

$\therefore b_{yx} = \frac{1}{4}$

Since,

$r^2 = b_{xy} . b_{yx}$

$r^2 = \frac{\cancel{2}}{3} \times \frac{1}{\cancel{4}_2}$

$r^2 = 1/6$

$r = \pm \frac{1}{\sqrt{6}}$

finding bxy using eq-(2)

$x - 4y = 7$

$x = 4y + 7$

$\therefore b_{xy} = 4$

finding byx using eq-(1)

$3x - 2y = 5$

$2y = 3x - 5$

$y = \frac{3}{2}x - \frac{5}{2}$

$\therefore b_{yx} = \frac{3}{2}$

Since,

$r^2 = b_{xy} . b_{yx}$

$r^2 = \cancel{4}^{2} \times \frac{3}{\cancel{2}}$

$r^2 = 6$

$r = \pm\sqrt{6}$

Probability Distribution

Discrete

It has PMF
(Probability Mass Function)

When data is discrete we need to check for using PMF :

i) $P_i \geq 0$

ii) $\sum_{i=1}^{n} P_i = 1$

Here we have :-

1) Bernoullis Distribution
2) Binomial Distribution
3) Poisson Distribution

Continous

It has PDF
(Probablity Density Function)

when data is continous we need to check for using PDF

i) $P_i \geq 0$

ii) $\int_a^b f(x)\,dx = 1$

Here we have :-

1) Uniform Distribution
2) Exponential "
3) Normal Distribution

* Area under the curve is $\int f(x)\,dx = 1$.

Camlin	Page
Date	/ /

Binomial Distribution -

$$\boxed{P(X=x) = {}^{n}C_{x}\, p^{x} q^{n-x}}$$

x = for event we need to calculate probability
n: Total no. of events
p: probability of success
q: probability of failure

& $\boxed{p+q=1}$

→ Binomial distribution describes the no. of successes in a fine number of independent & identical bernoulli trials.

Bernoulli Distribution -

→ It deals with single binary experiment with two possible outcomes (success or failure) & assigns probablities to each outcome.

Bernoulli trials :-
These trials in probability where only two possible outcomes are obtained whether its success or failure, true or false, 0 or 1.

Ques. If a coin is tossed 5 times find the probabilities
1) exactly two heads
2) atleast four heads.
3) Atmost two heads

Solu given $n = 5$ event: tossing coin

$p = \frac{1}{2}$ (occurring heads)

$q = \frac{1}{2}$

1) exactly two heads $\therefore (x=2)$.

$$P(X=x) = {}^{n}C_{x}\, p^{x} q^{n-x}$$

$x = 2$

$$P(X=2) = {}^{5}C_{2}\, p^{2} q^{5-2}$$

$$P(X=2) = \frac{5!}{2!\,3!} \left(\tfrac{1}{2}\right)^{2} \left(\tfrac{1}{2}\right)^{3}$$

$$P(X=2) = 10 \times \frac{1}{32} = \frac{5}{16}$$

2) Atleast four heads $\therefore x \geq 4$

$$P(X \geq 4) = P(X=4) + P(X=5)$$

$$= {}^{5}C_{4}\, p^{4} q^{1} + {}^{5}C_{5}\, p^{5} q^{0}$$

$$= \frac{5!}{4!\,1!} \left(\tfrac{1}{2}\right)^{4} \tfrac{1}{2} + \frac{5!}{5!\,0!} \left(\tfrac{1}{2}\right)^{5} \left(\tfrac{1}{2}\right)^{0}$$

$$= \frac{5}{32} + \frac{1}{32} = \frac{6}{32}$$

$$P(X \geq 4) = \frac{3}{16}$$

3) Atmost two heads $\therefore x \leq 2$

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$

$$= {}^{5}C_{0}\, p^{0} q^{5} + {}^{5}C_{1}\, p^{1} q^{4} + {}^{5}C_{2}\, p^{2} q^{3}$$

$$= \frac{1}{32} + \frac{5}{32} + \frac{10}{32} = \frac{16}{32} = \frac{1}{2}$$

A coin is tossed 4 times what is probability of getting atleast two heads

$n = 4 \quad p = \frac{1}{2} \quad q = \frac{1}{2}$

$$P(\text{atleast two heads}) = P(X \geq 2)$$
$$= P(X=2) + P(X=3) + P(X=4) + P(X=5)$$
$$= 1 - [P(X=0) + P(X=1)]$$
$$= 1 - \left[{}^4C_0 p^0 q^4 + {}^4C_1 p^1 q^3\right]$$
$$= 1 - \left[\frac{4!}{0!\,4!}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 + \frac{4!}{1!\,3!}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3\right]$$
$$= 1 - \left(\frac{1}{16} + \frac{4}{16}\right)$$
$$P(X \geq 2) = 1 - \frac{5}{16} = \frac{11}{16}$$

→ Properties of Binomial Distribution

1) There are only two possible outcome, True or False, success or failure, 0 or 1.
2) The probability of success & failure remains same for each trial.
3) Only the number of success is calculated out of n independent trials.
4) Every trial is an independent trial which means the outcome of one trial does not affect outcome of another trial.

★ Mean of Binomial Distribution $\mu = np$

Variance of Binomial Distribution $\sigma^2 = npq$

(SD) $\sigma = \sqrt{npq}$

Poisson Distribution : $(n \geq 30)$

$$\boxed{P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!}}$$

where λ = average = np $\boxed{\lambda = np}$

e = Euler's constant = $\boxed{2.718}$

☆ Mean = variance = np

Def → Poisson distribution represents the probability of a given number of events happening in a fixed time space if these cases occurs with a known steady rate & individually of the time since the last event.

Que. At a university the probability that a member of staff is absent on any day is 0.001. If there are 800 member of staff. calculate the probability that number of staff absent on any day is 4.

Sol given $n = 800$ $p = 0.001$

$x = 4$ $\lambda = np = 800 \times \frac{0.001}{1000} = 0.8$

$$\therefore P(X=4) = \frac{\lambda^4 e^{-\lambda}}{4!} = \frac{(0.8)^4 \cdot e^{-0.8}}{4!}$$

$$P(X=4) = 0.007$$

$$P(X=4) = \frac{0.4096 \times 0.449}{24} = 0.0076$$

Camlin Page
Date / /

If the probability of a bad reaction from a medicine is 0.002 determine the chances that out of thousand persons more than 3 will suffer from bad reaction.

$e = 2.718$

$n = 1000$

$p = 0.002$

$\therefore \lambda = 1000 \times 0.002 = 2$

$P(X>3) = 1 - P(X \leq 3)$

$= 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3)]$

$= 1 - \left[\frac{\lambda^0 e^{-\lambda}}{0!} + \frac{\lambda^1 e^{-\lambda}}{1!} + \frac{\lambda^2 e^{-\lambda}}{2!} + \frac{\lambda^3 e^{-\lambda}}{3!}\right]$

$P(X>3) = 1 - \left(e^{-2} + 2e^{-2} + \frac{\cancel{4}^{2} e^{-2}}{\cancel{2}} + \frac{\cancel{8}^{4} e^{-2}}{\cancel{6}_{3}}\right)$

$= 1 - \frac{19e^{-2}}{3}$

$P(X>3) = 1 - \frac{19 \times (2.718)^{-2}}{3}$

$P(X>3) = 1 - \frac{19 \times 0.1353}{3} = 1 - \frac{2.571}{3} = 0.1428$

An average of 0.61 soildiers dead by horse kicks per year in each persian army cops. You want to calculate the probability that exactly 2 soilders dead assuming that the no. of horse kicks deaths per year follows poisson distribution

given [$\lambda = 0.61$ $x = 2$ $e = 2.718$]

$\therefore P(X=2) = \frac{\lambda^2 e^{-\lambda}}{2!} = \frac{(0.61)^2 e^{-0.61}}{2!} = \frac{0.3721 \times 0.543}{2}$

$P(X=2) = 0.101$

Que. If 1% of total screws made by a factory are defective find the probability that less than 3 screws are defective in a sample of 100 screws.

Soln

$n = 100$

$p = 1\% = \frac{1}{100}$ $\quad\quad \lambda = np = 100 \times \frac{1}{100} = 1$

x = less than 3 $\therefore x < 3$

$P(X<3) = P(X=0) + P(X=1) + P(X=2)$

$= \frac{\lambda^0 e^{-\lambda}}{0!} + \frac{\lambda^1 e^{-\lambda}}{1!} + \frac{\lambda^2 e^{-\lambda}}{2!}$

$\therefore P(X<3) = e^{-1} + e^{-1} + \frac{e^{-1}}{2}$

$P(X<3) = \frac{5}{2} e^{-1} = 0.919$

Que. If in a industry there is a chance that 5% of the employ will suffer from corona, what is the probability that in a group of 20 employees more than 3 employees will suffer from corona.

Soln

$n = 20$

$p = 5\% = 0.05$

$\lambda = np = 20 \times \frac{5}{100} = 1$

$P(X>3) = 1 - P(X \leq 3) = 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3)]$

$P(X>3) = 1 - \left[\frac{\lambda^0 e^{-\lambda}}{0!} + \frac{\lambda^1 e^{-\lambda}}{1!} + \frac{\lambda^2 e^{-\lambda}}{2!} + \frac{\lambda^3 e^{-\lambda}}{3!}\right]$

$= 1 - \left(e^{-1} + e^{-1} + \frac{e^{-1}}{2} + \frac{e^{-1}}{6}\right) = 1 - \frac{16}{6_3} e^{-1} = 1 - \frac{8}{3}(2.718)^{-1}$

$P(X>3) = 1 - 0.981 = 0.0188.$

Difference between Binomial & Poisson distribution

Binomial Distribution	Poisson Distribution
$n \ngtr 30$	$n > 30$
→ 'n' : no of trials	→ n is no of trials
$P(X=x) = {}^nC_x \, p^x q^{n-x}$	$P(X=x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$
n = no. of trials p: probability of success q: probability of failure	λ = average $\lambda = np$ e = Euler's constant $e = 2.718$
Mean $= \mu = np$ Var $= \sigma^2 = npq$ sd $= \sigma = \sqrt{npq}$	Mean $= \lambda = np$ Var $= \sigma^2 = np$ sd $= \sigma = \sqrt{np}$

Uniform Distribution

numericals

→ Also known as rectangular distribution

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

$$P(x) = \int_c^d f(x)\,dx$$

$$P(c < x < d) = \int_c^d f(x)\,dx = \frac{d-c}{b-a}$$

Mean $= \frac{a+b}{2}$ Median $= \frac{a+b}{2}$ sd $= \sqrt{\frac{(b-a)^2}{12}}$

Que. If x is uniformly distributed in $(-1,4)$ then

1) Its mean $\mu = \frac{-1+4}{2} = 1.5$
2) Var $\frac{(4+1)^2}{12} = \frac{25}{12} = 2.083$
3) sd $\sqrt{2.083} = 1.443$
4) Median 1.5

Que. Using uniform distribution PDF for a random variable x in $(0,20)$ find $P(3<x<16)$.

Soln $f(x) = \frac{1}{b-a} = \frac{1}{20-0} = \frac{1}{20}$

$\therefore P(3<x<16) = \int_3^{16} \frac{1}{20}\, dx$

$= \frac{1}{20}\int_3^{16} dx$

$P(3<x<16) = \frac{1}{20}[x]_3^{16} = \frac{16-3}{20} = \frac{13}{20} = 0.65$

or $P(3<x<16) = \frac{d-c}{b-a} = \frac{16-3}{20-0} = \frac{13}{20} = 0.65.$

Exponential distribution

$\boxed{\text{mean} = \frac{1}{\lambda} \quad \text{var} = \frac{1}{\lambda^2}}$

It is a continous probability distribution in which events occurs independently at a constant rate. The PDF of exponential distribution with rate 'λ' is given by

$\boxed{p = \int f(x)\,dx}$ $\boxed{f(x) = \lambda e^{-\lambda x}}$ for $x \geq 0$

Suppose that we have a light bulb that has a constant failure rate & time until failure follows an exponential distribution with the rate of $\lambda = 0.02$ What is the probability that the light bulb will last more than 50 hrs without failing. $0.02 e^{-0.02x}$

0.1353

given $\lambda = 0.02$

since $f(x) = \lambda e^{-\lambda x}$

$f(x) = \lambda e^{-\lambda x}$

$\lambda \int (-e^{-x})$

$-\lambda [e^{-\infty} - e^{-50\lambda}]$

Normal Distribution

It is also known as gaussian distribution & it is a continous probability distribution which is widely used in statistics.
It is characterised by its bell shaped curve & is defined as follows.
A random variable 'x' is said to have a normal distribution with parameters μ & σ^2 (mean & var) respectively. Its density fncn is given by the probability function.

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

where $-\infty < x < \infty$
$-\infty < \mu < \infty$
$\sigma > 0$

Que. ~~A random variable x with mean μ & var σ^2.~~

Q. Regression - num. defi. define type. prob.
Q. Probab. Poi, Bino, Unif - num

Que. ~~If $x \simeq$~~

Note :-

1) Random variable x with mean μ & variance σ^2 which follows the normal law is expressed by
$$x \simeq N(\mu, \sigma^2)$$

2) If $x \sim N(\mu, \sigma^2)$ then $z = \frac{x-\mu}{\sigma}$ is a standard normal varient with $\mu = 0$ & $\sigma^2 = 1$.

* The distribution of normal random variable x with $\mu=0$ & $\sigma^2=1$ is called <u>standard normal distribution</u>

* Short Note for all probability distributions —

Discrete

① Bernoulli distribution
probability either 0 or 1

② Binomial distribution

$P(X=x) = {}^nC_x \, p^x q^{n-x}$

$p+q=1$

mean $= \mu = np$

var $= \sigma^2 = npq$

sd $= \sigma = \sqrt{npq}$

③ Poisson Distribution

$P(X=x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$ $\quad e = 2.718$

mean $= \lambda = np$

var $= \sigma^2 = np$

Continous

① Uniform Distribution

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a<x<b \\ 0 & \text{otherwise} \end{cases}$$

$$P(x) = \int_c^d f(x)\,dx = \frac{d-c}{b-a}$$

- mean $= \dfrac{a+b}{2} =$ median
- sd $= \sqrt{\dfrac{(b-a)^2}{12}}$ $\quad$ var $= \dfrac{(b-a)^2}{12}$

② Exponential distribution

$f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$

$P(x) = \int f(x)\,dx$

- mean $= \dfrac{1}{\lambda}$ • var $= \dfrac{1}{\lambda^2}$

③ Normal Distribution

$$f(x,\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$