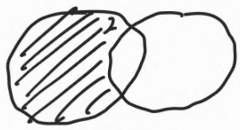
Axioms of probability

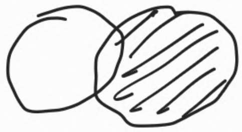
① $P(A) \geq 0$ ② $P(S) = 1$ ③ $P(A \cup B) = P(A) + P(B)$
$(\because A \cap B = \phi$
$A \cup B = S\ )$

A B

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



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



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

Conditional probability

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

↳ already occurred event

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

Multiplication theorem

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

$\rightarrow P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$

If independent,

$P(A \cap B) = P(A) \cdot P(B)$ etc, $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

$\boxed{\text{BAYE'S THEOREM}}$

mutually disjoint events

$E_1, E_2, E_3 \dots E_n \rightarrow$ mutually exclusive events

$P(E_i) \neq 0$ for $(i = 1, 2, 3, \dots n)$ then,

for arbitary event A, which is subject of $\bigcup_{i=0}^{n} E_i$

with $P(A) > 0$ ie $P\left(\frac{E_i}{A}\right) = \frac{P(E_i) \cdot P(A/E_i)}{\sum_{i=1}^{n} P(E_i) \cdot P(A/E_i)}$

Q) 2/7/18

Q In a ~~gold~~ Bolt factory, machine A, B, C manufacture 20%, 30%, 50% of the total of their output 6%, 3% and 2% of defective. A bolt is drawn at random and found to be defective. Find the probability that it is manufactured from machine A, B and C.

$P(A) = 20\%$, $P(B) = 30\%$, $P(C) = 50\%$

$P(D/A) = 6\%$, $P(D/B) = 3\%$, $P(D/C) = 2\%$

Here, D means defective

If a bolt is defective then the prob that it is from

Machine A $\quad P(A/D) = \frac{P(A) \cdot P(D/A)}{\sum P(\ ) \cdot P(D/\ )}$

$$= \frac{0.2 \cdot \frac{6}{\cancel{100}} \times \cancel{0.2}/\cancel{0.2}}{0.2 \cdot \frac{6}{\cancel{100}} \cdot \frac{\cancel{0.2}}{\cancel{0.2}} + 0.3 \cdot \frac{3}{\cancel{100}} \cdot \frac{\cancel{0.3}}{\cancel{0.3}} + 0.5 \cdot \frac{2}{\cancel{100}} \cdot \frac{\cancel{0.5}}{\cancel{0.5}}}$$

1.2 = 12

$= \frac{1.2}{1.2 + 0.9 + 1} = \frac{}{3.1} \quad \frac{}{31}$

$P(B/D) = \frac{9}{31} \qquad P(C/D) = \frac{10}{31}$

PROPERTIES OF DISTRIBUTIVE FUNCTIONS

Mean $= \mu = \Sigma x \cdot p(x)$

Variance $= \sigma^2 = \Sigma (x-\mu)^2 p(x)$

$= \Sigma x^2 p(x) - \mu^2$

Standard deviation $\sigma = \sqrt{\text{variance}} = \sqrt{\sigma^2}$

(ii)

Q. A random variable 'x' has the following probability function.

$X = x$	0	1	2	3	4	5	6	7
$P(X=x)$	0	k	2k	2k	3k	k^2	$2k^2$	$7k^2+k$

Find (i) K (ii) $p(x \geq 6)$ (iii) $p(0 < x < 5)$ and (iv) Determine the distribution function of 'x'.

(i) $p(x) = \Sigma p(x) = 1$

$0 + K + 2K + 2K + 3K + K^2 + 2K^2 + 7K^2 + K = 1$

$9K + 10K^2 = 1$

$10K^2 + 9K - 1 = 0$

$(10K-1)(K+1) = 0$

$K = \frac{1}{10}, \ -1 \rightarrow X$

(ii) $p(X \geq 6) = p(X=6) + p(X=7)$

$= 2K^2 + 7K^2 + K$

$= 9K^2 + K = 9\left(\frac{1}{10}\right)^2 + \frac{1}{10}$

$= \frac{19}{100}$

(iii) $p(0 < X < 5)$

$= p(X=1) + p(X=2) + p(X=3) + p(X=4)$

$= K + 2K + 2K + 3K = 8K = \frac{8}{10}$

(iv)

x	$p(X=x)$	$F(x) = p(X \leq x)$
0	0	$0 = 0 = 0$
1	K	$K = \frac{1}{10} = 0.1$
2	$2K$	$3K = \frac{3}{10} = 0.3$
3	$2K$	$5K = \frac{5}{10} = 0.5$
4	$3K$	$8K = 8/10 = 0.8$
5	K^2	$8K + K^2 = 8/10 + \frac{1}{100} = 0.81$
6	$2K^2$	$8K + 3K^2 = 8/10 + \frac{3}{100} = 0.83$
7	$7K^2 + K$	$9K + 10K^2 = \frac{9}{10} + \frac{10}{100} = \frac{10}{10} = 1$

Mean $\mu = \Sigma x\, p(x)$

$= 0(0) + 1\left(\frac{1}{10}\right) + 2\left(\frac{2}{10}\right) + 3\left(\frac{2}{10}\right) +$

$4\left(\frac{3}{10}\right) + 5\left(\frac{1}{100}\right) + 6\left(\frac{2}{100}\right) +$

$7\left(\frac{7}{100} + \frac{1}{10}\right)$

$\sigma^2 = \Sigma x^2 p(x) - \mu^2$

$\sigma^2 = 0^2(0) + 1^2\left(\frac{1}{10}\right) + 2^2\left(\frac{2}{10}\right) + 3^2\left(\frac{2}{10}\right) + 4^2\left(\frac{3}{10}\right)$

$+ 5^2\left(\frac{1}{100}\right) + 6^2\left(\frac{2}{100}\right) + 7^2\left(\frac{7}{100} + \frac{1}{10}\right) - \mu^2$

$SD = \sqrt{\sigma^2}$

PROBABILITY DISTRIBUTION FUNCTIONS (PDF)

Mean: $\mu^2 = \int_{-\infty}^{x} x\, f(x)\, dx$

Median: $\int_{a}^{M} f(x)\, dx = \int_{M}^{b} F(x)\, dx = 1/2$ (divides whole series into 2 parts)

Mode: $f'(x) = 0$ & $f''(x) \leq 0$ $f(x)$ is max

Mean deviation: $\int_{-\infty}^{\infty} (x-\mu)\, f(x)\, dx$

Q. A continuous random variable x

$$f(x) = \begin{cases} \frac{1}{16}(3+x)^2, & -3 \leq x \leq -1 \\ \frac{1}{16}(6-2x^2), & -1 \leq x \leq 1 \\ \frac{1}{16}(3-x)^2; & 1 \leq x \leq 3 \end{cases}$$

Find (i) Verify whether it is pdf or not.

to verify $\rightarrow \int_{-\infty}^{\infty} f(x)\, dx = 1$

$$\int_{-3}^{3} f(x)\, dx = \int_{-3}^{-1} f(x)\, dx + \int_{-1}^{1} f(x)\, dx + \int_{1}^{3} f(x)\, dx$$

$$\int_{-3}^{-1} \frac{1}{16}(3+x)^2 dx + \int_{-1}^{1} \frac{1}{16}(6-2x^2)\, dx + \int_{1}^{3} \frac{1}{16}(3-x)^2 dx$$

$= 1$

Q. A continuous random variable x has the following Pdf $f(x) = cxe^{-x}$, $x > 0$ find (i) c (ii) $p(x<2)$ (iii) $p(2<x<3)$

(i) $\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_{0}^{\infty} f(x)\,dx = 1$

$\int_{0}^{\infty} cxe^{-x}\,dx = 1 \qquad c\int_{0}^{\infty} xe^{-x}\,dx = 1$

$c\left[-xe^{-x} - e^{-x}\right]_{0}^{\infty} = 1$

$c[0+1] = 1 \Rightarrow \boxed{c=1}$

(ii) $P(x<2) = \int_{0}^{2} f(x)\,dx = \int_{0}^{2} cxe^{-x}\,dx$

$= c\left(-xe^{-x} - e^{-x}\right)_{0}^{2}$

$= \left(-2e^{-2} - e^{-2} + 1\right) = 1 - 3e^{-2}$

similarly (iii)

Q. A continuous random variable 'x' has the following pdf $f(x) = 3x^2$, $0 \le x \le 1$. find mean & variance.

Mean $\mu = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{0}^{1} x\,3x^2\,dx$

$$= \int_0 3x^3 dx = 3\left(\frac{x^4}{4}\right)_0 = 3\left(\frac{1}{4}\right) = \frac{3}{4}$$

$$\text{Variance } (\sigma^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$

$$= \int_0^1 x^2 (3x^2)\,dx - \left(\frac{3}{4}\right)^2 = 3\left(\frac{x^5}{5}\right)_0^1 - \left(\frac{3}{4}\right)^2$$

$$= 3\left(\frac{1}{5}\right) - \frac{9}{16} = \frac{3}{5} - \frac{9}{16}$$

PROBABILITY DISTRIBUTIONS

BINOMIAL DISTRIBUTION

If x is a discrete random variable with prob func $\boxed{p(X=x) = {}^nC_x\, p^x q^{n-x}}$ where $x = 0, 1, 2, \ldots n$ is called a Binomial variant denoted by $x \sim B(n, p)$, where n, p are parameters.

NOTE $\quad (p+q)^n = \sum_{x=0}^{n} {}^nC_x\, p^x q^{n-x}$

Mean of BD $\quad E(x) = \sum_{x=0}^{n} x \cdot p(x) = np$

Variance of BD $\quad V(x) = E(x^2) - [E(x)]^2 = npq$

→ The mean of BD is always > than its variance

$$\boxed{np > npq}$$

Basic

(1) mean = 4 $\quad$ variance = $\frac{4}{3}$ $\quad$ $p(x \geq 1) = ?$

$np = 4 \qquad npq = 4/3 \qquad \boxed{q = 1/3} \qquad \boxed{n = 6} \qquad \boxed{p = 2/3}$

$P(x \geq 1) = P(x=1) + P(x=2) \ldots + P(x=6)$

or

$P(x \geq 1) = 1 - P(x < 1)$

$= 1 - P(x=0) \quad = 1 - {}^{n}C_{x} p^{x} q^{n-x}$

$= 1 - {}^{6}C_{0} (2/3)^{0} (1/3)^{6-0}$

$= 0.998$

(e) Q. The incidence of an occupational disease in an industry, the workers has 20% chance of suffering from teeth. What is the probability that out of 6 workers chosen at random, 4 or more will suffer from disease.

$n = 6$

$p = \frac{20}{100} = 1/5 \qquad p + q = 1$

$q = 4/5$

$P(x \geq 4) = P(x=4) + P(x=5) + P(x=6)$

$= {}^{6}C_{4} (1/5)^{4} (4/5)^{2} + {}^{6}C_{5} (1/5)^{5} (4/5)^{1} + {}^{6}C_{6} (1/5)^{6} (4/5)^{0}$

$= 0.01696$

FITTING OF BINOMIAL DISTRIBUTION

(f) Fitting of binomial distribution :-

Q. Fit a binomial distribution to the following frequency distribution.

x	0	1	2	3	4	5	6
$f(x)$	13	25	52	58	32	16	4.

$n = 6$

Total frequency $N = \sum_{x=0}^{n} f(x) = 13 + 25 + 52 + 58 + 22 + 16 + 4$

$= 200$

Mean $\mu = \dfrac{\sum f_i \cdot x_i}{N}$

$\sum_{x=0}^{n} f_i x_i = 0 \times 13 + 1 \times 25 + 2 \times 52 + \ldots 6 \times 4$

$= 535$

$\mu = \dfrac{535}{200} = 2.675$

Mean for BD $= np$

$np = 2.675$

$p = \dfrac{2.675}{6} = 0.4458$

$q = 1 - 0.4458 = 0.5541 \quad (\because p + q = 1)$

x	$P(X=x) = {}^nC_x\, p^x q^{n-x}$	$N \cdot P(X=x)$	
0	${}^6C_0 (0.4458)^0 (0.5541)^6$	200×0.0289	$= 5.78 \approx 6$
1	${}^6C_1 (0.4458)^1 (0.5541)^5$	200×0.1397	$= 27.9427 \approx 28$
2	${}^6C_2 (0.4458)^2 (0.5541)^4$	200×0.2807	$= 56.1526 \approx 56$
3	${}^6C_3 (0.4458)^3 (0.5541)^3$	200×0.30138	$= 60.2766 \approx 60$
4	${}^6C_4 (0.4458)^4 (0.5541)^2$	200×0.1814	$= 36.2854 \approx 36$
5	${}^6C_5 (0.4458)^5 (0.5541)^1$	200×0.0585	$= 11.7076 \approx 12$
6	${}^6C_6 (0.4458)^6 (0.5541)^0$	200×0.0078	$= 1.5695 \approx 2$
			$= 199.7162$
			$\simeq 200$

Modified BD

x	0	1	2	3	4	5	6
$f(x)$	13	25	52	58	32	16	4
Modified values	6	28	56	60	36	12	2

POISSON DISTRIBUTION

It is a limiting case of a binomial distribution under the conditions that

(i) n is very large $(n \to \infty)$

(ii) p is very small $(p \to 0)$

(iii) $\lambda = np$

$$p(x) = \begin{cases} \dfrac{e^{-\lambda} \cdot \lambda^{x}}{x!} & ; (\text{if } x \geq 0) \\ 0 & ; \text{otherwise} \end{cases}$$

Mean of PD

$\mu = \Sigma x p(x) \Rightarrow \boxed{\mu = \lambda}$

Variance of PD

$\sigma^2 = \Sigma x^2 p(x) - \mu^2 \Rightarrow \boxed{\sigma^2 = \lambda}$

SD $= \boxed{\sigma = \sqrt{\lambda}}$

20.218, 4.111, 0.626.

Q. Fit a poisson distribution for the following data and also calculate expected frequencies

x	0	1	2	3	4	5
$f(x)$	125	95	49	20	8	3.

$N = \sum f(x) = 125 + 95 + 49 + 20 + 8 + 3 = 300$

$\sum x f(x) = 0 \times 125 + 1 \times 95 + 2 \times 49 + \dots 5 \times 3 = 300$

$$\lambda = \frac{300}{300} = 1$$

$$N\, p(x) = N \cdot \frac{e^{-\lambda} \cdot \lambda^x}{x!} = 300 \cdot \frac{e^{-1}(1)^x}{x!}$$

$$P(x=0) = 300 \times \frac{e^{-1}(1)^0}{0!} = 110.3638$$

$$P(x=1) = 300 \times \frac{e^{-1}(1)^1}{1!} = 110.3635$$

$$P(x=2) = 300 \times \frac{e^{-1}(1)^2}{2!} = 59.18$$

$$P(x=3) = 300 \times \frac{e^{-1}(1)^3}{3!} = 18.393$$

$$P(x=4) = 300 \times \frac{e^{-1}(1)^4}{4!} = 4.598$$

$$P(x=5) = 300 \times \frac{e^{-1}(1)^5}{5!} = 0.9196$$

$$\simeq 299.815$$

The expected frequencies are 110.3635, 110.3638, 55.18, 18.39, 4.598, 0.9196

NORMAL DISTRIBUTION

A random variable 'X' is said to having normal distribution if its density function is given by

$$\boxed{f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}}, \quad -\infty \le x \le \infty \text{ where } -\infty \le \mu \le \infty$$

$\mu \to$ mean of ND $\sigma \to$ SD of ND $\mu, \sigma \to$ parameters of ND

Mean of ND $\mu = \int_{-\infty}^{\infty} x f(x)\, dx = \mu$

Variance $= \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2 = \sigma^2$

Mode: $f'(x) = 0$ $f''(x) < 0$

Median $= \int_{-\infty}^{M} f(x)\, dx = 1/2$ $\boxed{M = \mu}$

STANDARD NORMAL DISTRIBUTION

If x is a normal variant with mean (μ) and SD (σ) then $Z = \frac{x-\mu}{\sigma}$ is a standard normal variate with mean (0) & standard deviation (1) then the probability density function standard normal variate (Z)

$$f(Z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-Z^2/2} \quad ; -\infty \leq Z \leq \infty$$

Normal distribution curve

The graph of ND depends on μ and σ

μ determines location of graph.

σ depends on the height & width of graph.

If σ is ↑ then curve is ↓ height and ↑ width.

If σ is $\downarrow$ then curve is $\uparrow$ height and $\downarrow$ width

<u>Characteristics</u> of normal distribution :-

→ For ND, mean = mode = median

→ curve is bell-shaped.

→ Normal curve is symmetrical about the line $x = \mu$

→ As $x \uparrow$, $f(x) \downarrow$ rapidly.

→ Max prob occurs at the point $x = \mu$.

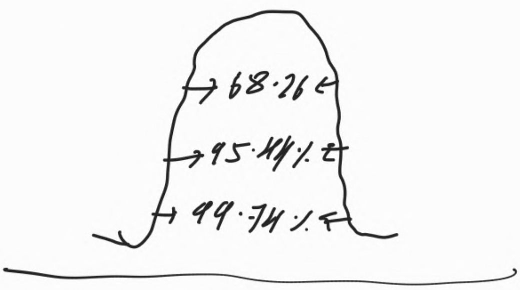
and given by $\boxed{[F(x)]_{max} = \frac{1}{\sigma\sqrt{2\pi}}}$

→ Total area under curve $\int_{-\infty}^{\infty} f(x)\,dx = 1$ is distributed as follows

$\mu - \sigma \le x \le \mu + \sigma$ covers 68.26 % of area

$\mu - 2\sigma \le x \le \mu + 2\sigma$ covers 95.44 % of area

$\mu - 3\sigma \le x \le \mu + 3\sigma$ covers 99.74 % of area



NOTE $z = \frac{x - np}{\sqrt{npq}}$ where $q = 1 - p$

standard normal variate

07/11

(7) Q. The S.Deviation and mean of the of the normal distribution are 70 and 160. Find (i) $p(38 < x < 16)$ (ii) $p(x > 152)$ (iii) $p(x < 140)$

$\sigma = 70 \quad \mu = 160$

$z = \frac{x-\mu}{\sigma} = \frac{x-160}{70}$

(i) $p(38 < x < 46)$

$\Rightarrow p\left(\frac{38-160}{70} < Z < \frac{46-160}{70}\right)$

$\Rightarrow p(-1.74 < Z < -1.62)$

$= p(Z < -1.62) - p(Z < 1.74)$

$= 0.0526 - 0.0409$

(ii) $p(x > 152) = p\left(Z > \frac{152-160}{70}\right)$

$= p\left(Z > \frac{-8}{70}\right) = p(Z > -0.11) = 1 - p(Z < -0.11)$

$= 1 - 0.4562 = 0.5438$

(iii) $p(x < 140) = p\left(Z < \frac{140-160}{70}\right)$

$= p(Z < -0.28) = 0.3897$

NOTE $QD : MD : SD$ QD → Quartile Diviation

$\frac{2}{3}\sigma : \frac{4}{5}\sigma : \sigma$ MD → mean deviation

$10 : 12 : 15$

- CORRELATION

If the change in one variable effect in the

If the change in the ... then the 2
change in the other variable, then the 2 variables are said to be correlated (or)
the degree of relationship is known as Correlation.

Karl pearson's coefficient of correlation

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\ \sqrt{n\sum y^2 - (\sum y)^2}}$$

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\ \sqrt{\sum (y-\bar{y})^2}}$$ r (or) r' (or) r_{xy}

$x-\bar{x}$ means deviation of mean from $\bar{x}$

$n \to$ no. of observations.

NOTE : $-1 \le r \le 1$

$r = +1 \to$ perfectly positively correlated.

$r = -1 \to$ " negatively "

$r > 0 \to$ positively correlated

$r < 0 \to$ negatively "

$r = 0 \to$ uncorrelated

correlated.

(1) Calculate the correlation co-efficient for the following heights in inches of fathers and their sons

x	65	66	67	67	68	69	70	72	= 5.4.4.
y	67	68	65	68	72	72	69	71	

x	y	x^2	y^2	xy
65	67	4225	4484	4355
66	68	4356	4624	4488
67	65	4489	4225	4355
67	68	4489	4624	4556
68	72	4624	5184	4896
69	72	4761	5184	4968
70	69	4900	4761	4830
72	71	5184	5041	5112
544	552	37028	38132	37560

$\Sigma x = 544$ $\quad \Sigma y = 552$ $\quad \Sigma x^2 = 37028$ $\quad \Sigma y^2 = 38132$

$\Sigma xy = 37560$

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{8 \times 37560 - 544 \times 552}{\sqrt{8 \times 37028 - (544)^2}\sqrt{8 \times 38132 - (552)^2}}$$

$$= 0.603$$

$\therefore$ x & y are positively correlated.

For the following data compute the correlation b/w x and y. No. of items is equal to 15 ; Sum of Squares of deviation from the mean x and y are 136 and 138. Summation of product of deviation of x and y from the respective arithmetic is 122.

$n = 15$

$\Sigma(x-\bar{x})^2 = 136$

$\Sigma(y-\bar{y})^2 = 138$

$\Sigma(x-\bar{x})(y-\bar{y}) = 122$

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}} = \frac{122}{\sqrt{136}\sqrt{135}} = 0.89$$

$\therefore$ x & y are positively correlated.

SPEARMAN RANK CORRELATION (Rank Correlation)

$$\boxed{r = 1 - \frac{6\Sigma d^2}{n(n^2-1)}}$$

$d \rightarrow$ diff btw two ranks

$n \rightarrow$ no. of observations.

properties: same as previous correlation.

calculate the rank correlation from the following data.

x	y	R_1	R_2	$d = R_1 - R_2$	d^2
10	30	9	9.	0	0
15	42	5	3	2	4
12	45	8	2	6	36
17	46	3	1	2	4
13	33	7	8	-1	1
16	34	4	7	-3	9
24	40	1	4	-3	9
14	35	5	6	0	0
22	39.	2	5	-3	9
					72.

$\Sigma d^2 = 72$

$n = 9.$

$$r = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{\cancel{6}(\cancel{72})\,\cancel{24}\,12}{\cancel{9}(\cancel{80})\,\cancel{40}\,20}$$

$$= \frac{8}{20} = 0.40$$

$$= 0.4$$

$\therefore$ x and y are positively correlated

Calculate rank correlation of 6 students from the following data.

Marks in statistics	40	42	45	35	36	35
Marks in maths	46	43	44	39	40	43

x	R_1	y	R_2	$d = R_1 - R_2$	d^2
40	3	46	1	2	4
42	2	43	3.5	-1.5	2.25
45	1	44	2	-1	1
35	5.5	39	6	-0.5	0.25
36	4	40	5	-1	1
35	5.5	43	3.5	+2	4
					12.50

$n = 6$

$\Sigma d^2 = 12.50$

$$\rho = 1 - \frac{6\left[\Sigma d^2 + 1/12\,(T_x + T_y)\right]}{n(n^2-1)}$$

$T_x = m\,(m^2-1)$ ⟨m=2⟩

$= 2(2^2-1)$

$= 6$

$T_y = m\,(m^2-1)$, ⟨m=2⟩

$= 2\,(2^2-1)$

$= 6.$

$$\rho = 1 - \frac{6\left[12.50 + \frac{1}{12}(12)\right]}{6(36-1)}$$

$$= 1 - \frac{\cancel{6}\,(13.50)}{\cancel{6} \times 35}$$

$$= 1 - \frac{\cancel{13.5}\ 2.7}{\cancel{35}\ 7}$$

$$= \frac{4.3}{7} = 0.614$$

$\begin{array}{r} 7.0 \\ -2.7 \\ \hline 4.3 \end{array}$

∴ marks in statistics and marks in maths are +vel correlated

If there are repeated ranks of x & y.

rank correlation $\boxed{r = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}(T_x + T_y)\right]}{n(n^2-1)}}$

where $T_x = \sum m_i (m_i^2 - 1)$

$T_y = \sum m_j (m_j^2 - 1)$

x	R_1	y	R_2	$d = R_1 - R_2$	d^2
65	8	67	7	1	1
66	7	68	5	2	4
67	5.5	65	8	-2.5	6.25
67	5.5	68	5	0.5	0.25
68	3.5	72	1.5	2	4
68	3.5	72	1.5	2	4
70	2	68	5	-3	9
72	1	71	3	-2	4
					32.5

$n = 8$

$\sum d^2 = 32.5$

$$r = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}(T_x + T_y)\right]}{n(n^2-1)}$$

$T_x = m_1(m_1^2 - 1) + m_2(m_2^2 - 1)$

$m_1 = 2, \; m_2 = 2$

$= 2(2^2-1) + 2(2^2-1) = 12$

$T_y = m_1(m_1^2 - 1) + m_2(m_2^2 - 1)$

$m_1 = 2, \; m_2 = 3$

$= 2(2^2-1) + 3(3^2-1) = 30$

$$r = 1 - \frac{6\left[32.5 + \frac{1}{12}(12+30)\right]}{8(8^2-1)}$$

$= 0.57$

$\therefore$ x and y are positively correlated

... test are ranked by the

REGRESSION

- Regression equation y on x.
- Regression equation x on y

Regression is defined as the method that estimates the value of one variable when that of another variable known.

Direct method of regression eq y on x :-

$$\boxed{y = a + b_{yx} x}$$

$x \to$ independent variable
$y \to$ dependent variable
$a \to$ intercept
$b_{yx} \to$ estimated by applying least square method

$$a = \bar{y} - b_{yx}\bar{x}$$

$$b_{yx} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

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

where $\bar{x}$ & $\bar{y}$ be the mean of x & y respectively.

$$b_{yx} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

x on y

$$\boxed{x = a + b_{xy} y}$$

$$a = \bar{x} - b_{xy}\bar{y}$$

$$b_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum y^2 - \frac{(\sum y)^2}{n}}$$

Props:

→ Correlation coefficient is the geometric mean of the 2 regression coefficient.

→ Arithmetic μ of regre coeff > correla coeff

NOTE

μ → point of intersection of 2 lines.

if $r=0$ → 2 variables are uncorrelated
lines of regression are $\perp$ to each other.

if $r=\pm 1$ → 2 lines of regression either coincide / parallel to each other.

correlation coefficient btw x & y is given by

$\boxed{r = \pm\sqrt{b_{yx} b_{xy}}}$; $b_{yx} = \frac{r\sigma_y}{\sigma_x}$ $b_{xy} = \frac{r\sigma_x}{\sigma_y}$

r → correla coeff σ_x, σ_y → SD of x & y.

Formulas for two regression equations by using deviation method.

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

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

0/8/18
16. Production (x) 55 56 58 59 60 60 62
Export (y) 35 38 38 39 44 43 45.
Find (i) Karl pearson's correlation co-efficient
(ii) regression line x on y
(iii) regression line y on x.
(iv) Estimate export when production is 50

Sol

x	y	x^2	y^2	xy
55	35	3025	1225	1925
56	38	3136	1444	2128
58	38	3364	1444	2204
59	39	3481	1521	2301
60	44	3600	1936	2640
60	43	3600	1849	2580
62	45	3844	2025	2790
410	282	24050	11444	16568

$n=7$.
$\Sigma x = 410$
$\Sigma y = 282$
$\Sigma x^2 = 24050$
$\Sigma y^2 = 11444$
$\Sigma xy = 16568$

$\bar{x} = \frac{410}{7} = 58.57$

$\bar{y} = \frac{282}{7} = 40.28$

(ii) x on y

$x = a + b_{xy} y$

$b_{xy} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}} = \frac{16568 - \frac{410\times 282}{7}}{11444 - \frac{(282)^2}{7}} = 0.609$

$a = \bar{x} - b_{xy}\bar{y} = 58.57 - (0.609)(40.28)$
$= 34.03$

$x = 34.03 + (0.609)y$

(iii) y on x

$y = a + b_{yx} x$

$b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} = \frac{16568 - \frac{410\times 282}{7}}{24050 - \frac{(410)^2}{7}} = 1.4238$

$a = \bar{y} - b_{yx}\bar{x}$

$= 40.28 - (1.42)(58.5)$

$= -42.79.$

$y = (-42.79) + 1.42x.$

$r = \pm\sqrt{b_{yx}\, b_{xy}}$

$= \pm\sqrt{1.42 \times 0.609}$

$= \pm 0.929 = 0.929$

Both regression co-efficients are positive therefore Correlation co-efficient ~~are~~ is positive.

(iv) Given that $x = 50$

$y = -42.79 + (1.42)(50)$

$= 28.21$

---

2 lines of regressive are given by $x+2y-5=0$, $2x+3y-8=0$ and variance of x is 12. Calucculate

(i) mean values of x & y

(ii) correlation co-efficient (iii) variance

$x+2y-5=0 \rightarrow$ ①

$2x+3y-8=0 \rightarrow$ ②

On solving ① & ②

$y-2=0$

$y=2$

$x=1.$

$x = 5-2y$

$y = \frac{8}{3} - \frac{2}{3}x$

$b_{xy} = -2$, $b_{yx} = \frac{-2}{3}$

$r = \pm\sqrt{b_{yx}\, b_{xy}} = \pm\sqrt{(-2)(-2/3)}$

$= 1.15$ ✗

$x = 4 - 3/2\, y$

$y = \frac{5}{2} - \frac{1}{2}x$

$b_{xy} = \frac{-3}{2}$, $b_{yx} = \frac{-1}{2}$

$r = \pm\sqrt{b_{yx}\, b_{xy}}$

$= \pm\sqrt{\frac{-3}{2} \times \frac{1}{2}} = \sqrt{0.75}$

$= \pm 0.86$ $r = -0.86$ ✓

iii) $\sigma_x^2 = 12$

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

$$b_{yx}^2 = r^2 \frac{\sigma_y^2}{\sigma_x^2}$$

$$\left(\frac{-1}{2}\right)^2 = (-0.86)^2 \frac{\sigma_y^2}{12}$$

$$\sigma_y^2 = 4.056$$

→ Angle between 2 regression lines $y - \bar{y} = b_{yx}(x - \bar{x})$
and $x - \bar{x} = b_{xy}(y - \bar{y})$

$$\boxed{\theta = \tan^{-1} \left| \frac{\left(\frac{r^2 - 1}{r}\right) \sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \right|}$$

LEAST SQUARE

Sum of squares of residuals is minimum.

straight line $y = a + bx$

residuals or derivations denoted by 'ε' and is given by $\boxed{\varepsilon = y - a - bx}$

- Used for obtaining the eqs of curve which fits best to a given set of observations.

FITTING OF STRAIGHT LINE (or) LINEAR CURVE FIT

$y = a + bx$

residual → $\varepsilon = y - a - bx$

Acc to princi of least sq method, sum of sq of residuals is min.

$$E = \sum (y - a - bx)^2$$

$\boxed{\sum y = na + \sum bx}$ —— $\boxed{\sum xy = a\sum x + b\sum x^2}$

$\boxed{\sum x = na + \sum by}$ —— $\boxed{\sum xy = a\sum y + b\sum y^2}$

predict y at $x=5$ by fitting a least square st. line to the following data;

x	2	4	6	8	10	12
y	18	15	14	11	11	9

x	y	xy	x^2
2	18	36	4
4	15	60	16
6	14	84	36
8	11	88	64
10	11	110	100
12	9	108	144
42	78	486	364

$n=6$

$\sum x = 42$

$\sum y = 78$

$\sum x^2 = 364$

$\sum xy = 486$

we predict y value that is enough to fit a st. line equation of form $y = a + bx$ i.e the normal eq. are $\sum y = na + b\sum x$

& $\sum xy = a\sum x + b\sum x^2$

$\sum y = na + b\sum x$

$78 = 6a + b(42) \rightarrow ①$

$\sum xy = a\sum x + b\sum x^2$

$486 = 42a + 364b \rightarrow ②$

solving ① & ②

$6a + 42b = 78$

$42a + 364b = 486$

$a = 19, \; b = -0.8571$

∴ Required straight line is,

$y = 19 - 0.571 \times 5$

$y = 14.7145$

Fit a straight line for the following data;

x	1	2	3	4	6	8
y	2.4	3	3.6	4	5	6

x	y	xy	x^2
1	2.4	2.4	1
2	3	6	4
3	3.6	10.8	9
4	4	16	16
6	5	30	36
8	6	48	64
24	24	113.2	130

$n = 6$

$\Sigma x = 24$

$\Sigma y = 24$

$\Sigma x^2 = 130$

$\Sigma xy = 113.2$

$\Sigma y = na + b\Sigma x$

$24 = 6a + b(24)$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

$113.2 = a(24) + b(130)$

$6a + 24b = 24$

$24a + 130b = 113.2$

$a = 1.97$, $b = 0.505$

$\therefore$ Required straight line is,

$y = 1.97 + 0.5x$

Estimate the production for the year 2010 by fitting a straight line to the following data,

Year	2003	2004	2005	2006	2007
Production (in lakhs)	5	8	14	12	13

x	y	xy	x^2
2003	5	10015	4012009
2004	8	16032	4016016
2005	14	28070	4020025
2006	12	24072	4024036
2007	13	26091	4028049
10025	52	104280	20100135

$n = 5$

$\Sigma x = 10025$

$\Sigma y = 52$

$\Sigma xy = 104280$

$\Sigma x^2 = 20100135$

$y = a + bx$

$\Sigma y = an + b\Sigma x$

$52 = 5a + b(10025)$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

$104280 = a(10025) + b(20100135)$

$a = -3999.6$

$b = 2$

$y = a + bx$

at $x = 2010$,

$y = -3999.6 + 2(2010)$

$y = 20.4$

$y = a + bx$

$\Sigma y = na + b\Sigma x$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

FITTING OF A PARABOLA / 2^{nd} DEGREE POLYNOMIAL NON-LINEAR CURVE

non-linear curve $\rightarrow$ $\boxed{y = a + bx + cx^2}$

$$\Sigma y = na + b\Sigma x + c\Sigma x^2$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$
$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

Fit a 2^{nd} degree parabola for a following data:-

x	0	1	2	3	4
y	1	5	10	22	38

x	y	xy	x^2	x^3	x^4	x^2y
0	1	0	0	0	0	0
1	5	5	1	1	1	5
2	10	20	4	8	16	40
3	22	66	9	27	81	198
4	38	152	16	64	256	608
10	76	243	30	354	100	851

$\Sigma x = 10$, $\Sigma y = 76$, $\Sigma xy = 243$, $\Sigma x^2 = 30$, $\Sigma x^3 = 100$,
$\Sigma x^4 = 354$, $\Sigma x^2 y = 851$, $n = 5$

$\Sigma y = na + b\Sigma x + c\Sigma x^2$
$76 = 5a + 10b + 30c$

$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$
$243 = 10a + 30b + 100c$

$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$
$851 = 30a + 100b + 354$

$a = 1.42$, $b = 0.24$, $c = 2.21$.

The required equation is $y = 1.42 + 0.24x + 2.21x^2$

FITTING OF AN EXPONENTIAL CURVE

form of exp curve → $\boxed{y = ae^{bx}}$ (or) $\boxed{y = ab^{x}}$

$Y = A + BX$

$\boxed{\Sigma Y = nA + B\Sigma X}$

$\boxed{\Sigma XY = \Sigma XA + B\Sigma X^2}$

$y = ae^{bx}$

$\log y = \log ae^{bx}$

$= \log a + \log e^{bx}$

$\log y = \log a + bx$

let $Y = \log y$, $A = \log a$

$B = b$, $X = x$

$Y = A + BX$

$\boxed{\Sigma Y = An + B\Sigma X}$

$\boxed{\Sigma XY = \Sigma XA + B\Sigma X^2}$

$y = ab^{x}$

$\log y = \log ab^{x}$

$\log y = \log a + \log b^{x}$

$\log y = \log a + x\log b$

let $Y = \log y$, $A = \log a$,

$B = \log b$, $X = x$

Find the curve for the best fit of curve to data by method least squares.

x	1	5	7	9	12
y	10	15	12	15	21

$y = ae^{bx}$

$\log y = \log ae^{bx}$

$\log y = \log a + bx\log e$

$\log y = \log a + bx$

$Y = \log y$

$A = \log a$

$B = b$

$X = x$

$Y = A + BX$

$\Sigma Y = \Sigma A + B\Sigma X$

$\Sigma XY = \Sigma AX + B\Sigma X^2$

$x = X$	y	$Y = \log y$	X^2	XY
1	10	1	1	1
5	15	1.17	25	5.85
7	12	1.079	49	7.49
9	15	1.17	81	10.53
12	21	1.322	144	15.86
34	73	5.741	300	40.75

$n = 5$, $\Sigma X = 34$, $\Sigma Y = 5.741$, $\Sigma XY = 40.75$, $\Sigma X^2 = 300$

$\Sigma Y = nA + B\Sigma X$

$5.741 = 5A + B(34)$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$40.75 = 34A + 300B$

$A = \log a$

$a = e^{A}$

$A = 0.978$, $B = 0.024$

$b = B = 0.024$

$a = 10^{A} = 10^{0.978}$ [illegible]

$y = ae^{bx}$

$y = 10^{0.978}(e^{0.024x})$

$y = 218 \cdot e^{-0.0009x}$

Fit an exponential curve $y = ab^x$ for the following data.

$X = x$	y	$Y = \log y$	X^2	XY
40	30	1.47	1600	58.8
65	20	1.30	4225	84.5
90	10	1	8100	90
5	80	1.90	25	9.5
30	40	1.60	900	48
10	65	1.81	100	18.1
80	15	1.17	6400	93.6
85	15	1.17	7225	99.45
70	25	1.38	4900	91.3
25	50	1.69	625	42.25
500		14.41	34100	635.2

$\Sigma n = 10$

$\Sigma X = 500$

$\Sigma Y = 14.41$

$\Sigma X^2 = 34100.$

$\Sigma XY = 635.2$

$y = ab^x$

$\log y = \log a + x \log b$

$\log y = Y, \quad \log a = A$

$\log b = B, \quad x = X$

$Y = A + BX$

$\Sigma Y = An + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$14.41 = 10A + 500B$

$636.901 = 500A + 34100B$

$A = 1.90$

$B = -0.009.$

$\left\{ \begin{array}{l} A = \log a \\ \Rightarrow \underline{a = 10^A} \end{array} \right.$

$\left\{ \begin{array}{l} B = \log b \\ \underline{b = 10^B} \end{array} \right.$

$a = 10^A = 10^{1.90} = 79.4$

$b = 10^B = 10^{-0.009} = 0.979$

$y = ab^x$

$y = 79.4\,(0.979)^x$

^{x} for the following

FITTING OF A POWER CURVE (SQUARE CURVE)

$y = ax^b$

$\log y = \log ax^b$

$\log y = \log a + \log x^b$

$\log y = \log a + b \log x$

$\log y = Y$, $\log a = A$, $b = B$, $\log x = X$

$Y = A + BX$

$\Sigma Y = An + B \Sigma X$

$\Sigma XY = A \Sigma X + B \Sigma X^2$

Q. Fit a power curve $y = ax^b$ for the following data and estimate y when $x = 4.6$

x	y	$X = \log x$	$Y = \log y$	X^2	XY
1	20	0	1.301	0	0
2	30	0.301	1.477	0.090	0.444
3	52	0.477	1.716	0.227	0.818
4	77	0.602	1.886	0.362	1.135
5	135	0.698	2.130	0.487	1.486
6	211	0.778	2.324	0.605	1.808
7	326	0.845	2.513	0.714	2.110
8	550	0.903	2.740	0.815	2.474
9	1052	0.954	3.022	0.910	2.882
		5.558	9.109	4.15	13.107

$n = 9$

$\Sigma X = 5.558$

$\Sigma Y = 9.109$

$\Sigma X^2 = 4.15$

$\Sigma XY = 13.107$

$Y = A + BX$

$\Sigma Y = AN + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$y = ax^b$

$\log y = \log a + \log x^b$

$= \log a + b \log x$

$19.109 = 9A + 5558 B$

$13.17 = 5.558 A + 4.15 B$

$A = 0.94$

$B = 1.907$

$a = 10^A = 10^{0.94} = 8.70$

$b = B = 1.907$

$y = 8.70 x^{1.907}$

$x = 4.6 \Rightarrow y = 8.70 (4.6)^{1.907}$

$= 159.73$