MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY
DEPARTMENT OF INFORMATION TECHNOLOGY

II B.Tech I Sem

L	T/P/D	C
4	**1 / - / -**	**3**

(R17A0024)PROBABILTY AND STATISTICS

Objectives: To learn

- Understand a random variable that describes randomness or an uncertainty in certain realistic situation. It can be either discrete or continuous type.
- In the discrete case, study of the binomial and the Poisson random variables and the normal random variable for the continuous case predominantly describe important probability distributions. Important statistical properties for these random variables provide very good insight and are essential for industrial applications.
- Most of the random situations are described as functions of many single random variables. The objective is to learn functions of many random variables, through joint distributions.
- The types of sampling, Sampling distribution of means, Sampling distribution of variance, Estimations of statistical parameters, Testing of hypothesis of few unknown statistical parameters.
- The mechanism of queuing system, The characteristics of queue, The mean arrival and service rates, The expected queue length, The waiting line, The random processes, The classification of random processes, Markov chain, Classification of states, Stochastic matrix (transition probability matrix),Limiting probabilities, Applications of Markov chains.

UNIT -1 : Random variable and Probability distributions

Random Variables

Single and multiple Random variables -Discrete and Continuous. Probability distribution function, mass function and density function of probability distribution. mathematical expectation and variance.

Probability distributions: Binomial distribution – properties, mean and variance, Poisson distribution – properties, mean and variance and Normal distribution – properties, mean and variance

UNIT -2 :Correlation and Regression

Correlation -Coefficient of correlation , Rank correlation, Regression- Regression Coefficients , Lines of Regression.

UNIT -3 : Sampling Distributions and Statistical Inferences

Sampling: Definitions of population ,sampling ,statistic ,parameter-Types of sampling – Expected values of sample mean and variance,Standard error- Sampling distribution of means and variance

Parameter Estimations : likelihood estimate , interval estimate.
Testing of hypothesis: Null and Alternative hypothesis-Type I and Type II errors, Critical region – confidence interval – Level of significance,One tailed and Two tailed test
Large sample Tests: i) Test of significance of single mean and equality of means of two samples(cases of known and unknown variance whether equal or unequal)
ii) Tests of significance difference between sample proportion and population proportion and difference between two sample proportions

UNIT -4 : Exact Sampling Distributions(Small samples)

Exact Sampling Distributions(Small samples) Student t- distribution - properties
i)Test of significant difference between sample and population mean
ii)Test of difference between means of two small samples(independent and dependent samples)
F- distribution - properties –test of equality of two population variances
Chi-square distribution -properties –i)Test of goodness of fit
ii)Test of independence of attributes

UNIT-5 : Queuing Theory and Stochastic process

Queuing Theory
Structure of a queuing system its characteristics-Arrival and service process-Pure Birth and Death process Terminology of queuing system -Queuing model and its types-M/M/1 model of infinite queue (without proofs)and M/M/1 model of finite queue (without proofs).

Stochastic Process
Introduction to stochastic process-classification and methods of description of Random process i.e,stationary and non-stationary Average values of single and two or more random process
Markov process, Markov chain, Examples of Markov chains, Stochastic matrix.

TEXT BOOKS:

1. Probability and Statistics by T.K..V Iyengar& B.Krishna Gandhi S.Ranganatham,MVSSAN Prasad. S Chand Publishers.
2. Fundamentals of Mathematical Statistics by SC Gupta and V.K. Kapoor

REFERENCES :
1.Higher Engineering Mathematics By Dr.B.S.Grewal, Khanna Publishers
2. Probability and Statistics for Engineers and Scientists by Sheldon M.Ross,Academic Press.
Outcomes:

- Students would be able to identify distribution in certain realistic situation. It is mainly useful for circuit as well as non circuit branches of engineering. Also able to differentiate among many random variables involved in the probability models. It is quite useful for all branches of engineering.
- The student would be able to calculate mean and proportions(small and large samples)and to make important decisions from few samples which are taken out of unmanageably huge populations.It is mainly useful for non-branches of engineering.

- The student would be able to find the expected queue length, the ideal time the traffic intensity and the waiting time. these are very useful tools in many engineering and data management problems in the industry. it is useful for all branches of engineering.
- The student would able to understand about the random process, markov process and markov chains which are essentially models of many time dependent processes such as signals in communications, time series analysis, queuing systems. The student would be able to find the limiting probabilities and the probabilities in n^{th} state. It is quite useful for all branches of engineering.

INDEX

NOTE:-List only main topics

1

Random Variable :

A real variable 'X' whose value is determined by the outcome of a random experiment. It is also called a stochastic variable.

eg: 1. Tossing of a coin twice

Sample Space $S = \{s_1, s_2, s_3, s_4\}$

where $s_1 = HH$, $s_2 = HT$, $s_3 = TH$, $s_4 = TT$

Define a function $X : S \to R$ by $X(s) =$ 'no. of heads' then

$X(s_1) = 2$, $X(s_2) = 1$, $X(s_3) = 1$, $X(s_4) = 0$.

$\therefore$ Range $= \{2, 1, 0\} \Rightarrow$ Random variable $X(s) = \{0, 1, 2\}$.

eg: 2 : Throw of pair of dice

Sample space $S = 6^2 = 36 = \{ (1,1), (1,2), \ldots (1,6)$
$(2,1), (2,2) \ldots (2,6)$
$\ldots$
$(6,1), (6,2), \ldots (6,6) \}$

Define a function $X : S \to R$ by

$X(s) =$ Sum on two dice, $x(i,j) = i + j \;\forall (i,j) \in S$

then $X(s) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

Types of Random variables

They are of 2 types — Discrete Random variables / Continuous Random variables.

Discrete Random variable

A r.v 'X' which can take only a finite number of discrete values in an interval of domain is called a Discrete Random Variable.

(or)

If the r.v take the values only in the set $\{0, 1, 2, \ldots, n\}$ is called discrete random variable.

eg: 1. Throw of a dice
2. Tossing of a coin
3. No. of defectives in a sample of electric bulbs etc.

Continuous Random Variable :

A random variable X which can take values continuously i.e which takes all possible values in a given interval.

eg 1. The height, age and weight of individuals
2. Temperature and time etc

Probability Function of a Discrete random variable :

If for a discrete r.v 'X', the real valued function $P(x)$ is such that $P(X=x) = p(x)$ then $p(x)$ is called Probability function or probability mass function of a disrete r.v 'X'.

Probability Distribution function.

P.D.F associated with X is the probability that the outcome of an experiment will be one of the outcomes for which $X(s) \leq x, x \in R$

ie $F_X(x) = P(X \leq x) = P\{S : X(s) \leq x\}, \ -\infty < x < \infty$

is called Distribution function.

Properties of Distribution function

1. If F distribution function of X, and if $a < b \Rightarrow P(a < X \leq b) = F(b) - F(a)$
2. $P(a \leq X \leq b) = P(X=a) + [F(b) - F(a)]$
3. $P(a < X < b) = [F(b) - F(a)] - P(X=b)$
4. $P(a \leq X < b) = [F(b) - F(a)] - P(X=b) + P(X=a)$

NOTE: $0 \leq F(x) \leq 1$.

Discrete Probability Distribution (Probability mass function)

It is the set of its possible values together with their respective probabilities.

Let X be a discrete r.v with possible outcomes $x_1, x_2, x_3, \ldots$ then their probabilities $p_i = P(X = x_i) = P(x_i)$ for $i = 1, 2, 3, \ldots$

if $P(x_i) > 0$ & $\sum_{i=1}^{n} p(x_i) = 1$ then the function '$p(x)$' is called the 'probability mass function' of r.v X & $\{x_i, p(x_i)\}$ for $i = 1, 2, \ldots$ is called Discrete Probability Distribution.

Eg: Tossing a Coin two times with random variable $X(s)$ = no. of heads = $\{0,1,2\}$

then $P(X=0) = \frac{1}{4}$; $P(X=1) = \frac{1}{2}$, $P(X=2) = \frac{1}{4}$

x_i	0	1	2
$P(x_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$\Rightarrow \sum p(x_i) = 1$ & $p(x_i) > 0$.

Thus total probability '1' is distributed into 3 parts as $\frac{1}{4}, \frac{1}{2}$ & $\frac{1}{4}$.

Note: 1) Frequency distribution tells how total frequency distributed among different values of the variable.

2) Probability distribution tells how total probability '1' is distributed among the values which the random variable can take

Cumulative Distribution Function of a Discrete R.V.

$$F(x) = P(X \le x) = \sum_{i=1}^{x} P(x_i) \text{ where } x \text{ is any integer}$$

then $F(x)$ is called cumulative Distribution function of X.

i.e $F(x) = \begin{cases} 0, & -\infty < x < x_1 \\ P(x_1), & x_1 \le x < x_2 \\ P(x_2), & x_2 \le x < x_3 \\ \ldots & \ldots \\ P(x_1) + P(x_2) + \cdots + P(x_n), & x_n \le x < \infty \end{cases}$

Expectation of Discrete Probability Distribution

Let X be a r.v with $x_1, x_2, \ldots x_n$ with probabilities $p_1, p_2, \ldots, p_n$ then

Mathematical Expectation or Expected value of X is defined as "Sum of products of different values of x & their corresponding probabilities".

i.e $\boxed{E(X) = \sum_{i=1}^{n} x_i p_i}$,

In General $\boxed{E(g(x)) = \sum_{i=1}^{n} p_i\, g(x_i)}$

Results: 1. $E(X) = \mu$ $\left\{\because \text{Mean} = \frac{\sum p_i x_i}{\sum p_i} = \sum p_i x_i \text{ as } \sum p_i = 1\right\}$

2. $E(X+k) = E(X) + k$ (k is a constant)
3. $E(kX) = k\,E(X)$
4. $E(aX \pm b) = a\,E(X) \pm b$
5. $E(X+Y) = E(X) + E(Y)$ provided $E(X), E(Y)$ exists.
6. $E(X - \bar{X}) = 0$
7. $E(XY) = E(X) \cdot E(Y)$ if X, Y are two independent r.v.s.
8. $E\left(\frac{1}{X}\right) \ne \frac{1}{E(X)}$

Mean: The mean 'μ' of the Distribution function is given by

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

Variance: Variance of the probability distribution of r.v X is defined as

$$\boxed{Var(X) = E[X - E(X)]^2}$$

Since variance of $X = V(X) = \sum_{i=1}^{n} (x_i - \mu)^2 p_i = \sum_{i=1}^{n} p_i x_i^2 - \mu^2$

$= E(X^2) - [E(X)]^2$

Also $\boxed{V(X) = E(X^2) - [E(X)]^2}$

Standard Deviation $S.D = \sigma = \sqrt{Var(x)} = \sqrt{E[(x)-(E(x))]^2}$

Results : 1. $Var(K) = 0$ where k is constant $\left\{\begin{array}{l} \because \text{mean} = \text{constant} \\ \Rightarrow x - \mu = 0 \end{array}\right\}$

2. $V(KX) = K^2 V(X)$

3. $V(X+K) = V(X)$

★ 4. $V(aX+b) = a^2 V(X)$ where a, b are constants.

Sol: Let $Y = aX+b$

$\therefore E(Y) = E(aX+b)$

$= aE(X) + b$

Consider $Y - E(Y) = aX + \cancel{b} - aE(X) \cancel{b} = a[X - E(X)]$

Squaring & taking expectation, we get

$E[Y - E(Y)]^2 = a^2 E(X - E(X))^2$

$\Rightarrow Var(Y) = a^2 Var(X)$

$\Rightarrow V(aX+b) = a^2 V(X)$.

Hence Proved.

Do Problems.

Continuous Probability Distribution : (Probability Density function)

Def^n: By considering small interval $[x - \frac{dx}{2}, x + \frac{dx}{2}]$ of length dx around the point x. Let $f(x)$ be any continuous function of x so that $f(x)dx$ represents the probability that the variable X falls in that interval. ie $P(x - \frac{dx}{2} \leq X \leq x + \frac{dx}{2}) = f(x)dx$ then $f(x)$ is called the probability density function & the curve $y = f(x)$ is known as probability density curve.

The probability for a variate value to fall in the finite interval (a,b) is $\int_a^b f(x)dx$ which represents the area b/w the curve $y = f(x)$, x-axis & the ordinated $x = a$ & $x = b$.

Properties of density function

(i) $f(x) \geq 0 \quad \forall x \in R$

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

(iii) $P(E) = \int_E f(x)dx$ is well defined for any event E.

★ Note: Since continuous variable associate with intervals, the probability at a particular point is always zero. Thus $P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = P(a \leq X \leq b)$
$= F(b) - F(a)$

Binomial Distribution.

It was discovered by James Bernoulli in the year 1700 & is a discrete probability distribution.

The conditions for the applicability of a B.D are :

(i) There are 'n' independent trials (ie trials are repeated under identical conditions)

(ii) There are only 2 possible outcoms for each trial. Sucess 'p' & failure 'q'.

(iii) The trials are independent, ie the probability of an event in any trial is not affected by the results of any other trial.

(iv) The probability of success in each trial remains constant & does not change from trial to trial.

Def^n: A r.v. X has a B.D if it assumes only non-negative values & its probability density function is given by

$$P(X=r) = p(r) = \begin{cases} {}^nC_r\, p^r q^{n-r} & ;\ r=0,1,2,\ldots,n \ ;\ q=1-p \\ 0 & ,\ \text{otherwise} \end{cases}$$

$$= b(r;n,p)$$

Here n, p are called parameters of the distribution as they are the 2 independent constants. 'n' is sometime known as "degree of the distribution".

egs of B.D : (i) No. of defective bolts in a box containing 'n' bolts.

(ii) No. of post graduates in a group of 'n' men, etc.

The Binomial Distribution function is:

$$F_X(x) = P(X \le x) = \sum_{r=0}^{n} {}^nC_r\, p^r q^{n-r}.$$

Binomial Frequency Distribution: The possible no. of successe and their frequencies is called a Binomial frequency Distribution.

If 'n' independent trials constitute one experiment and this expt is repeated 'N' times, then the frequency of 'r' successes is

$N.\,{}^nc_r\,p^r q^{n-r}$.

$\therefore$ In 'N' sets of 'n' trials the theoretical frequencies of $0,1,2,\ldots,r,\ldots,n$ successes are given by the terms of expansion of $N(q+p)^n$.

Note ① The probabilities of $0,1,2,\ldots,r,\ldots,n$ successes in 'n' trials are given by the terms of the binomial expansion of $(q+p)^n$.

ie $(q+p)^n = q^n + {}^nc_1 q^{n-1}p + {}^nc_2 q^{n-2}p^2 + \cdots + {}^nc_r q^{n-r}p^r + \cdots + p^n$.

Here the probability of exactly 'r' successes is ${}^nc_r p^r q^{n-r}$.

② The probability of no successes in n trials is $= q^n$

" " " all successes " " " $= p^n$.

" " " atleast One success " " " $= {}^nc_1 q^{n-1}p + {}^nc_2 q^{n-2}p^2 + \cdots + p^n = 1-q^n$.

~~Do problems~~.

Constants of B.D :

(1) Mean of X : $\mu = E(X) = \sum_{r=0}^{n} r\,p(r) = \sum_{r=0}^{n} r.\,{}^nc_r p^r q^{n-r}$

$$\Rightarrow \mu = 1.{}^nc_1 p q^{n-1} + 2\,{}^nc_2 p^2 q^{n-2} + \cdots + n.{}^nc_n q^{n-n} p^n$$

$$= n p q^{n-1} + 2\frac{n(n-1)}{2!} p^2 q^{n-2} + \cdots + n p^n$$

$$= np\left[q^{n-1} + \cancel{2}\frac{(n-1)}{\cancel{2!}} p q^{n-2} + \cancel{3}\frac{(n-1)(n-2)}{\cancel{3}\lfloor 2} p^2 q^{n-3} + \cdots + p^{n-1}\right]$$

$$= np\left[q^{n-1} + (n-1)pq^{n-2} + \frac{(n-1)(n-2)}{2!}p^2 q^{n-3} + \cdots + p^{n-1}\right]$$

$$= np\left[{}^{n-1}c_0 q^{n-1}p^0 + {}^{n-1}c_1 p^1 q^{n-2} + {}^{n-1}c_2 p^2 q^{n-3} + \cdots + {}^{n-1}c_{n-1} p^{n-1}\right]$$

$$= np\,[q+p]^{n-1}$$

$$\boxed{\therefore \mu = np}$$

2. Variance of B.D:

$$V(X) = E(X^2) - [E(X)]^2 = npq.$$

Proof: $V(X) = E(X^2) - [E(X)]^2$

$$= \sum_{r=0}^{n} r^2 p(r) - \left[\sum_{r=0}^{n} r\, p(r)\right]^2$$

$$= \sum_{r=0}^{n} [r(r-1)+r]p(r) - \mu^2$$

$$= \sum_{r=0}^{n} r(r-1)p(r) + \sum_{r=0}^{n} r\, p(r) - \mu^2$$

$$= \sum_{r=0}^{n} r(r-1)p(r) + \mu - \mu^2$$

$$= 2(2-1)\,{}^nC_2\, p^2 q^{n-2} + 3(3-1)\,{}^nC_3\, p^3 q^{n-3} + \cdots + n(n-1)\,{}^nC_n\, p^n q^{n-n} + \mu - \mu^2$$

$$= 2\,{}^nC_2\, p^2 q^{n-2} + 6\,{}^nC_3\, p^3 q^{n-3} + \cdots + n(n-1)p^n + \mu - \mu^2$$

$$= \cancel{2}\,\frac{n(n-1)}{\cancel{2}} p^2 q^{n-2} + \cancel{6}\,\frac{n(n-1)(n-2)}{\cancel{3\times 2}} p^3 q^{n-3} + \cdots + n(n-1)p^n + \mu - \mu^2$$

$$= n(n-1)p^2\left[q^{n-2} + (n-2)pq^{n-3} + \cdots + p^{n-2}\right] + \mu - \mu^2$$

$$= n(n-1)p^2\left[{}^{n-2}C_0\, q^{n-2} + {}^{n-2}C_1\, p^1 q^{n-3} + \cdots + {}^{n-2}C_{n-2}\, p^{n-2}\right] + \mu - \mu^2$$

$$= n(n-1)p^2\,[q+p]^{n-2} + \mu - \mu^2$$

$$= n(n-1)p^2 \quad (\because q+p=1) + \mu - \mu^2$$

$$= \cancel{n^2p^2} - np^2 + np - \cancel{n^2p^2} \quad \{\because \mu = np\}$$

$$= np(1-p)$$

$$= npq \quad \{\because p+q=1 \Rightarrow q = 1-p\}$$

$$\boxed{\therefore \operatorname{Var}(X) = \sigma^2 = npq}$$

Poisson Distribution

S.D. Poisson (Simeon Denis Poisson) (1837) introduced Poisson Distribution as a rare distribution of rare events. i.e the events whose probability of occurrence is very small but the no. of trials which could lead to the occurrence of the Event, are very large.

Conditions of P.D : (i) The no. of trials 'n' is large.
(ii) The probability of success 'p' is very small.
(iii) $np=\lambda$ is finite.

Examples of P.D :
(1) The no. of printing mistakes per page in a large text.
(2) The no. of cars passing a certain point in 1 minute.

Definition: A r.v X is said to follow P.D if it assumes only non-negative values & its probability density function is given by

$$p(x,\lambda) = P(X=x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & x=0,1,2,\dots \\ 0, & \text{otherwise} \end{cases}$$

Here $\lambda>0$ is called parameter of the distribution.

Constants of Poisson Distribution:

(1) Mean $= \mu = E(X) = \lambda$

Proof: $E(X) = \sum x\,p(x) = \sum_{x=0}^{\infty} \dfrac{x\,e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\dfrac{x\lambda^x}{x!}$

$$= e^{-\lambda}\sum_{x=0}^{\infty}\frac{\cancel{x}\,\lambda^x}{\cancel{x}(x-1)!} = e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^x}{(x-1)!}$$

$$= e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda\cdot\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}$$

Take ~~x-1=y~~ $= \lambda e^{-\lambda}\left[1+\lambda+\dfrac{\lambda^2}{2!}+\dfrac{\lambda^3}{3!}+\cdots\right]$

$$= \lambda e^{-\lambda}[e^{\lambda}]$$

$$= \lambda.$$

$$\boxed{\therefore\ \mu=\lambda}$$

(2) Variance of P. D.

$$V(X) = \lambda \qquad E(X^2) - [E(X)]^2 = E(X^2) - \mu^2 = E(X^2) - \lambda^2 .$$

Proof: $Var(X) = E(X^2) - \lambda^2$

$$E(X^2) = \sum x^2 p(x) = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{x e^{-\lambda}\lambda^x}{(x-1)!}$$

$$= e^{-\lambda} \sum \frac{x\lambda^x}{(x-1)!} = e^{-\lambda} \sum \frac{[(x-1)+1]\lambda^x}{(x-1)!}$$

$$= e^{-\lambda} \sum \frac{(x-1)\lambda^x}{(x-1)!} + e^{-\lambda} \sum \frac{\lambda^x}{(x-1)!}$$

$$= e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2} \cdot \lambda^2}{(x-2)!} + \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda^2 e^{-\lambda} (e^{\lambda}) + \lambda e^{-\lambda} (e^{\lambda})$$

$$= \lambda^2 e^{0} + \lambda e^{0} = \lambda^2 + \lambda \Rightarrow E(X^2) = \lambda^2 + \lambda .$$

$$\therefore Var(X) = E(X^2) - \lambda^2 = \cancel{\lambda^2} + \lambda - \cancel{\lambda^2} = \lambda \Rightarrow \boxed{Var(X) = \lambda}$$

Note. S.D $\sigma = \sqrt{\lambda}$.

Recurrence Relation:

$$p(x+1) = \frac{\lambda}{x+1} p(x) .$$

Note: When n is large say greater than 30 & p is very small say less than 0.1 then B.D can be approximated by P.D.

Normal Distribution.

It is a continuous distribution. N.D was first discovered by English Mathematician De-Moivre (1667-1745) in 1733 & further refined by French Mathematician Laplace (1749-1827) in 1774 & independently by Karl Friedrich Gauss (1777-185).

It is also known as Gaussian Distribution.

It is another limiting form of Binomial Distribution

It is found so often in real life that it is called Normal Distribution, the name which is commonly used.

<u>Def</u>^{n} A r.v X is said to have a N.D if its p.d.f is given by $f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, $-\infty<x<\infty$, $\sigma>0$, $-\infty<\mu<\infty$.

μ is Mean & σ is Standard deviation of x.

μ, σ are called <u>parameters</u> of N.D.

A linear combination of independent normal variates is also a normal variate.

<u>Chief Characteristics of N.D</u>:

1. The graph of Normal Distribution in xy-plane is known as normal curve.
2. The curve is a bell-shaped curve & symmetrical w.r.t mean 'μ'. ie The two tails on right & left sides of the mean 'μ' extends to infinity.
3. Area under the normal curve represents total population.
4. Normal curve is Unimodal (ie has only one max. pt.) At $x=\mu$: mean = Median = Mode (as distribution is symmetrical).
5. x-axis is an asymptote to the curve.
6. The points of inflexion of the curve are at $x=\mu\pm\sigma$ & curve changes from concave to convex at $x=\mu+\sigma$ to $x=\mu-\sigma$.

7. Area under the normal curve is distributed as :

(i) Area of normal curve b/w $\mu-\sigma$ & $\mu+\sigma$ is 68·27%.

(ii) " " " " " $\mu-2\sigma$ & $\mu+2\sigma$ is 95·43%

(iii) " " " " " $\mu-3\sigma$ & $\mu+3\sigma$ is 99·73%.

Note : The total area bounded by the curve & x-axis is One.

ie $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

$P(a < x < b)$ = Area under normal curve b/w the vertical lines $x=a$ & $x=b$.

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

Multiple Random Variables.

In many practical problems several r.vs interact with each other. Multiple R.Vs help us to determine the joint statistical properties like mean, variance etc.

We study about 2-dimensional r.vs, which can be easily extended to multiple r.vs.

A 2-dimensional r.v is denoted by (X, Y) where the random vector (X, Y) is outcome of a trial which occurs in pairs. ie $X = x$, $Y = y$.

Defn: An 'n' dimensional random vector (Vector of r.vs) is a function from sample space S into R^n (n dimensional Euclidean Space)

Multiple r.vs are also of 2 types: ① Discrete & ② Continuous.

It is called Discrete if (X, Y) assume only finite no. of pairs.

Discrete Multiple Random Variables:

Joint Probability Mass function (Joint Probability function)

If (X, Y) is a discrete 2-dimensional r.v then the function $f(x, y)$ from R^2 into R is called JPMF & is given by $\boxed{f(x,y) = P(X = x, Y = y) = P(x_i, y_j)}$

Note: ⓐ $0 \leq P(x_i, y_j) \leq 1$. ⓑ $\sum_i \sum_j P(x_i, y_j) = 1$.

Properties: ① $P(X = x_i) = p(x_i) = \sum_j P(x_i, y_j)$

② $P(Y = y_j) = p(y_j) = \sum_i P(x_i, y_j)$

③ $P(x_i) \geq P(x_i, y_j)$ for any j

④ $P(y_j) \geq P(x_i, y_j)$ for any i

Joint Cumulative Distribution Function.

Def^n : The joint probability on cumulative distribution function of 2 r.v.s uniquely defines the probability of the joint events $\{X \le x, Y \le y\}$. It is defined by

$$\boxed{F_{XY}(x,y) = P(X \le x, Y \le y)} = \sum_{i=0}^{x_i} \sum_{j=0}^{y_j} P(x_i, y_j)$$

Properties:

① $0 \le F_{XY}(x,y) \le 1$

② $F_{XY}(-\infty, \infty) = 1$.

③ $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0$.

* ④ F_{XY} is non-decreasing. $= P(X \le x, Y \le \infty)$.

* ⑤ $F_X(x) = F_{XY}(x, \infty) = P(X \le x) = P(X \le x, Y \le \infty)$.

⑥ $F_Y(x) = F_{XY}(\infty, y) = P(Y \le y) = P(X \le \infty, Y \le y)$.

Marginal Probability Distribution Function.

$\boxed{F_{XY}(x,\infty), F_{XY}(\infty, y)}$ are called marginal PDF of X & Y respectively. It is given by

$$\boxed{\begin{aligned} F_{XY}(x, \infty) &= \sum_j P(x, y_j) \\ F_{XY}(\infty, y) &= \sum_i P(x_i, y). \end{aligned}}$$

eg: Spse 2 dice are rolled simultaneous, then probability of getting 3 on first die is ?

$P(X=3) = P_{XY}(3, \infty) = P(x=3, Y=1) + P(x=3, y=2) + P(x=3, y=3) + P(x=3, y=4) + P(x=3, y=5) + P(x=3, y=6)$

$= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36}$

$= \frac{1}{6}$

Continuous Multiple Random Variables:

Joint Probability density function : It is denoted by $\underline{f_{xy}(x,y)}$. Also $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{xy}(x,y)\,dx\,dy = 1$

$$f_{xy}(x,y)\,dx\,dy = P(x \le X \le (x+dx),\ y \le Y \le (y+dy))$$

Note : $\boxed{f_{xy}(x,y) = \dfrac{\partial^2 F_{xy}(x,y)}{\partial x \partial y}}$ Relation b/w JPF & JPDF.

Properties:

① $P(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f_{xy}(x,y)\,dx\,dy$

② $F_{xy}(x_1,y_1) = \int_{-\infty}^{x_1} \int_{-\infty}^{y_1} f_{xy}(x,y)\,dx\,dy$

Marginal Probability density function (MPDF)

The MPDF of one of the r.vs is the integral of joint p.d.f over the other r.v.

$f_x(x) = \int_{-\infty}^{\infty} f_{xy}(x,y)\,dy$ is marginal p.d.f of X.

$f_y(y) = \int_{-\infty}^{\infty} f_{xy}(x,y)\,dx$ is " " " Y.

Statistical propeties of Jointly Distributed Random Variables.

Def^n: 2 jointly distributed r.vs X, Y are said to be statistically independent of each other iff

$$f_{XY}(x,y) = f_X(x) f_Y(y) \quad \forall (x,y) \text{ in the given range}$$

ie joint p.d.f = product of 2 marginal p.d.fs.

Also if $F_{X_1X_2}(x_1,x_2) = F_X(x_1)\cdot F_X(x_2) \ \forall \ x_1, x_2$ then also the 2 r.vs X_1, X_2 are said to be statistically independent.

Conditional Probability ~~distribution~~ Density function.

Def^n: $f_Y(Y/X) = \dfrac{f_{XY}(x,y)}{f_X(x)}$ is conditional p.d.f of Y.

$$f_X(X/Y) = \frac{f_{XY}(x,y)}{f_Y(y)} \text{ is conditional p.d.f of } X.$$

Note: If X, Y are statistically independent Then

$f_X(X|Y) = f_X(x)$

& $f_Y(Y|X) = f_Y(y)$

$\because f_X(X/Y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$ (by def^n)

$= \dfrac{f_X(x)\cdot \cancel{f_Y(y)}}{\cancel{f_Y(y)}}$ $\{\because$ Statistical independent

$= f_X(x)$

Conditional Probability Distribution Function.

Def^n: Let X, Y be 2 r.vs discrete or continuous.

$F_Y(Y/X) = \dfrac{F_{XY}(x,y)}{F_X(x)}$, $F_X(x) > 0$ is conditional distribution of r.vs Y given $X = x$.

Also $F_X(X/Y) = \dfrac{F_{XY}(x,y)}{F_Y(y)}$, $F_Y(y) > 0$ is conditional distribution of r.vs X given $Y = y$.

Problems related to Binomial Distribution.

① Pg 125 Pb 3. A die is thrown 6 times. If getting an even no. is a success, find the probabilities of (i) atleast one success (ii) ≤ 3 successes (iii) 4 successes.

w.k.t $P(x=r) = {}^nc_r p^r q^{n-r}$.

Sol: Given $n=6$.

p = getting an even no. is success

we have 2,4,6 as even no.s on die.

$\therefore p = \frac{3}{6} = \frac{1}{2} \Rightarrow q = \frac{1}{2}$ ($\because$ Total faces of dice = 6, Even " " " = 3)

(i) $P(X \geq 1) = 1 - P(x=0)$

$= 1 - {}^6c_0 p^0 q^{6-0} = 1 - q^6 = 1 - \left(\frac{1}{2}\right)^6 = \frac{63}{64}$

(ii) $P(X \leq 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3)$

$= {}^6c_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^6 + {}^6c_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^5 + {}^6c_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^4 + {}^6c_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^3$

$= \left(\frac{1}{2}\right)^6 \left[1 + 6 + \frac{\cancel{6}^3 \times 5}{\cancel{2}} + \frac{\cancel{6} \times 5 \times 4}{\cancel{3 \times 2}}\right]$

$= \frac{1}{64}[1+6+15+20] = \frac{\cancel{42}^{21}}{\cancel{64}_{32}} = \frac{21}{32}$.

(iii) $P(x=4) = {}^6c_4 p^4 q^{6-4} = \frac{\cancel{6}^3 \times 5}{\cancel{2}} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2 = \frac{15}{64}$

② Pg 126 Pb 4: 10 Coins are thrown simultaneously. Find the probability of getting atleast (i) 7 heads (ii) 6 heads.

Sol. Here $n = 10$, p = getting head $= \frac{1}{2} \Rightarrow q = \frac{1}{2}$.

(i) $P(X \geq 7) = P(x=7) + P(x=8) + P(x=9) + P(x=10)$

$= {}^{10}c_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 + {}^{10}c_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^2 + {}^{10}c_9 \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right) + {}^{10}c_{10} \left(\frac{1}{2}\right)^{10}$.

$= \left(\frac{1}{2}\right)^{10} \left[\frac{10 \times \cancel{9}^3 \times \cancel{8}^4}{\cancel{3} \times \cancel{2}} + \frac{\cancel{10}^5 \times 9}{\cancel{2}} + 10 + 1\right]$.

$= \frac{(120+45+11)}{1024} = \frac{\cancel{176}^{\cancel{88}\,\cancel{22}\,11}}{\cancel{1024}_{\cancel{512}\,\cancel{128}\,64}} = \frac{11}{64}$.

(ii) $P(X \geq 6) = P(x=6) + P(x \geq 7)$

$= {}^{10}c_6 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^4 + \frac{11}{64} = \frac{10 \times \cancel{9}^3 \times \cancel{8} \times 7}{\cancel{4} \times \cancel{3} \times \cancel{2}} \cdot \frac{1}{1024} + \frac{176}{1024}$

$= \frac{210+176}{1024} = \frac{\cancel{386}^{193}}{\cancel{1024}_{512}} = \frac{193}{512}$.

(3) If 3 of 20 tyres are defective & 4 of them are randomly chosen for inspection, what is the probability that only one of the defective tyre will be included? (Pg 127 Pb 7)
*.

Sol: Given = 20 tyres; 3 are defective ⇒ 17 are good ⇒ $p = \frac{3}{20}$ is defective tyres probability.

To chose 4 at random ⇒ $n = 4$.

$P(x=1) = ?$ $\quad {}^{n}c_1 \, p^1 q^{n-1}$

$$= {}^{4}c_1 \left(\frac{3}{20}\right)\left(\frac{17}{20}\right)^3 = \frac{4 \times 3 \times 17^3}{(20)^4} = 0.3685.$$

→ Pg 127 Pb7.

(4) Determine B.D for which mean is 4 & variance 3 (Pg 128 Pb 9)

B.D = $b(x; n, p)$.

To find n; p. Give $\mu = 4$, $\sigma^2 = 3$

w.k.t $\mu = np \Rightarrow np = 4$ Also $\sigma^2 = npq = 3 \Rightarrow npq = 3$

$\therefore npq = 3 \Rightarrow 4q = 3 \Rightarrow q = \frac{3}{4} \Rightarrow p = 1 - q = \frac{1}{4}$

$\therefore n\left(\frac{1}{4}\right) = 4 \Rightarrow n = 16.$

Thus $b(x; n, p) = b(x; 16, \frac{1}{4})$

(5) In 256 sets of 12 tosses of a coin, in how many cases one can expect 8 heads & 4 tails. (Pg 128 Pb11)

Sol: $N = 256$, $n = 12$, p = getting head is success $= \frac{1}{2} \Rightarrow q = \frac{1}{2}$.

$$P(x=8) = ? = {}^{12}c_8 \, p^8 q^4 = {}^{12}c_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^4 = \frac{\cancel{12} \times 11 \times \cancel{10}^{5} \times 9}{\cancel{1 \times 2} \times 3 \times \cancel{4}} \left(\frac{1}{2}\right)^{12}$$

$$= \frac{495}{(2)^{12}}$$

$\therefore$ Expected no. of such cases in 256 sets

$$= N \cdot P(x=8) = \frac{(256)(495)}{(2)^{12}} = \frac{2^8 (495)}{2^{12}} = \frac{495}{2^4}$$

$$= \frac{495}{16}$$

$$= 30.9375$$

$$\simeq 31.$$

6. Six dice are thrown 729 times How many times do you expect atleast three dice to show a 5 or 6? (Pg 129 Pb 13)

Sol: $N = 729$, $n = 6$, $p =$ to show 5 or 6 $= \frac{2}{6} = \frac{1}{3} \Rightarrow q = \frac{2}{3}$.

$$P(x \geq 3) = ? = {}^nc_3 p^3 q^{n-3} + {}^nc_4 p^4 q^{n-4} + {}^nc_5 p^5 q^{n-5} + {}^nc_6 p^6 q^{n-6}$$

$= P(x=3) + P(x=4) + P(x=5) + P(x=6) =$

$$= {}^6c_3 \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^3 + {}^6c_4 \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^2 + {}^6c_5 \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right) + {}^6c_6 \left(\frac{1}{3}\right)^6.$$

$$= \frac{\cancel{6} \times 5 \times 4}{\cancel{3 \times 2}} \times \frac{8}{3^6} + \frac{6 \times 5}{\cancel{2}} \times \frac{\overset{2}{\cancel{4}}}{3^6} + 6 \times \frac{2}{3^6} + \frac{1}{3^6}$$

$$= \frac{160 + 60 + 12 + 1}{3^6} = \frac{233}{3^6} = \frac{233}{729}$$

$\therefore$ Expected no. of such cases in 729 times $= N \cdot P(x \geq 3)$

$$= \cancel{729}\left(\frac{233}{\cancel{729}}\right) = \underline{233}.$$

★ 7 (i) Out of 800 families with 5 children each, how many would you expect to have (a) 3 boys (b) 5 girls (c) either 2 or 3 boys (d) atleast one boy? Assume equal probabilities for boys & girls.

(ii) Out of 800 families with 4 children each, how many families would be expected to have (a) 2 Boys & 2 girls (b) atleast one boy (c) no girl (d) atmost one girls? Assume equal probabilities for boys & girls. (Pg 131 Pb 19)

Sol: (i) $N = 800$, $n = 5$, $p =$ probability of child to be boy $= \frac{1}{2} \Rightarrow q = \frac{1}{2}$.

(a) $P(X=3) = ? = {}^5c_3 p^3 q^2 = {}^5c_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = \frac{5 \times \cancel{4}}{2} \frac{1}{\cancel{4} \times 8} = \frac{5}{16}$

(b) $P(X=0) = {}^5c_0 p^0 q^5 = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$.

(c) $P(X=2) + P(X=3) = {}^5c_2 p^2 q^3 + {}^5c_3 p^3 q^2 = \frac{5 \times \overset{2}{\cancel{4}}}{\cancel{2}} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 + \frac{5 \times \overset{2}{\cancel{4}}}{\cancel{2}} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2$

$$= \frac{\cancel{20}5}{\cancel{4} \times 8} = \frac{5}{8}.$$

(d) $P(X \geq 1) = 1 - P(X=0) = 1 - \frac{1}{32} = \frac{31}{32}$.

Thus (a) For 800 families Expectation of 3 boys $= 800 \cdot P(X=3)$

$$\Rightarrow \overset{50}{\cancel{800}} \times \frac{5}{\cancel{16}} = \underline{250} \text{ famlies}$$

(b) Expected no. of families to have 5 girls $= N \cdot P(x=0)$

$= \cancel{800}^{\cancel{100}\,25} \times \frac{1}{\cancel{32}_{4}} = \underline{25}$ families.

(c) Expected no. of families to have either 2 or 3 boys

$= N \cdot [P(x=2) + P(x=3)] = \cancel{800}^{100} \times \frac{5}{\cancel{8}} = \underline{500}$ families.

(d) Expected no. of families with atleast one boy

$= N \cdot P(x \geq 1) = \cancel{800}^{25} \times \frac{31}{\cancel{32}} = \underline{775}$ families.

(ii) $N = 800$, $n = 4$, $p =$ probability of child to be boy $= \frac{1}{2} \Rightarrow q = \frac{1}{2}$.

(a) $P(X=2) = {}^{n}C_2 p^2 q^2 = {}^{4}C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{\cancel{4} \times 3}{2} \times \frac{1}{\cancel{4} \times 4} = \frac{3}{8}$

$\therefore$ Expected no. of families to have 2 Boys & 2 girls

$= N \cdot P(x=2) = \cancel{800}^{100} \times \frac{3}{\cancel{8}} = \underline{300}$ families.

(b) $P(x \geq 1) = 1 - P(x=0) = 1 - {}^{4}C_0 p^0 q^4 = 1 - \left(\frac{1}{2}\right)^4 = \frac{15}{16}$

$\therefore$ Expected no. of families to have atleast 1 boy

$= N \cdot P(x \geq 1) = \cancel{800}^{50} \times \frac{15}{\cancel{16}} = \underline{750}$ families.

(c) $P(x=4) = {}^{4}C_4 \left(\frac{1}{2}\right)^4 = \frac{1}{16}$ $\{\because$ No girl $\Rightarrow$ all 4 are boys$\}$

$\therefore$ Expected no. of families to have ~~atleast~~ ~~all~~ NO girls

$= N \cdot P(x=4) = \cancel{800}^{50} \times \frac{1}{\cancel{16}} = \underline{50}$ families.

(d) $P(x=4) + P(x=3)$ ($\because$ Atmost 1 girl $\Rightarrow$ No girl or 1 girl $\Rightarrow$ 4 Boys or 3 boys)

$= {}^{4}C_4 \left(\frac{1}{2}\right)^4 + {}^{4}C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)$

$= \frac{1}{16} + 4 \times \frac{1}{16} = \frac{5}{16}$

$\therefore$ Expected no. of families to have atmost 1 girl

$= N \cdot [P(x=3) + P(x=4)]$

$= \cancel{800}^{50} \times \frac{5}{\cancel{16}} =$

$= \underline{250}$ families

⑧ (Pg 135 Pb25) In a B.D consisting of 5 independent trials, probabilities of 1 & 2 success are 0.4096 & 0.2048 respectively. Find the parameter 'p' of the distribution.

Sol: $n=5$, $P(x=1)=0.4096$; $P(x=2)=0.2048$ $p=?$

$P(x) = {}^nC_r\, p^r q^{n-r}$

$\therefore P(x=1) = {}^5C_1\, p\, q^{5-1} = 0.4096 \Rightarrow 5pq^4 = 0.4096$ — ①

$P(x=2) = {}^5C_2\, p^2 q^{5-2} = 0.2048 \Rightarrow \frac{5\times 4}{2} p^2 q^3 = 10p^2q^3 = 0.2048$ — ②

$\frac{①}{②} \Rightarrow \frac{P(x=1)}{P(x=2)} = \frac{\cancel{5}pq^{\cancel{4}}}{\cancel{10}_2 p^{\cancel{2}} \cancel{q^3}} = \frac{\cancel{0.4096}}{\cancel{0.2048}}{}^2$

$\Rightarrow \frac{q}{2p} = 2 \Rightarrow q = 4p$ $\{\because p = 1-q\}$ $\Rightarrow 1-p = 4p$
$\Rightarrow 1 = 5p$
$\Rightarrow p = \frac{1}{5}$
$\Rightarrow \boxed{p = 0.2}$

OR

By recurrence relation $P(r+1) = \frac{(n-r)p}{(r+1)q} P(r)$.

Thus $P(x=2) = \frac{(n-r)p}{(r+1)q} P(x=1)$

$= \frac{(5-1)p}{2q} P(x=1)$

$\Rightarrow \frac{\cancel{0.2048}}{1} = \frac{4p}{\cancel{2}q}(\cancel{0.4096})$

$\Rightarrow q = 4p \Rightarrow 1-p = 4p \Rightarrow 1 = 5p \Rightarrow p = \frac{1}{5} = 0.2$

⑨ (Pg 138 Pb32) A coin is biased in a way that a head is twice as likely to occur as likely to occur as a tail. If the coin is tossed 3 times, find the probability of getting 2 tail & 1 head.

Sol: Let p = probability of getting tail as success, $n=3$.

Given $P(H) = 2P(T)$ w.k.t $P(H)+P(T)=1 \Rightarrow 3P(T)=1$
$\Rightarrow P(T)=\frac{1}{3} \Rightarrow p=\frac{1}{3}$.
$\boxed{\therefore P(H)=\frac{2}{3}} \Rightarrow q = \frac{2}{3}$

To find $P(2T) = ?$ $P(x=2) = {}^nC_2\, p^2 q^{n-2}$

$= {}^3C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^{3-2}$

$= \cancel{3}\left(\frac{1}{3}\right)^2 \times \frac{2}{\cancel{3}} = \frac{2}{9}$

(10) Pg139 Pb 33. Fit a B·D to the following frequency distribution.

x	0	1	2	3	4	5	6
f	13	25	52	58	32	16	4

Fit a B·D $\Rightarrow$ Obtain $E(x)$ where $E(x) = N \cdot p(x)$; $N = \sum_i f_i$ & $p(x) = f(x)$.

From data $n = 6$ ($\because x = 0, 1, \dots 6$) ; $N = \sum_i f_i = 13+25+52+58+32+6+4 = 200$

$$\text{Mean} = \frac{\sum x_i f_i}{\sum_i f_i} = np \Rightarrow \frac{(0\times13)+(1\times25)+(2\times52)+(3\times58)+(4\times32)+(5\times16)+(6\times4)}{200}$$

$$\Rightarrow np = \frac{535}{200} = 2.675$$

$\because n=6 \Rightarrow 6p = 2.675 \Rightarrow p = 0.446 \Rightarrow q = 0.554$.

$\therefore p(x=0) = {}^nC_0 p^0 q^{n-0} = {}^6C_0 (0.554)^6 = (0.554)^6 \cdot 1 = 0.02891$

$P(x=1) = {}^6C_1 (0.554)^5 (0.446)^1 = 6(0.446)(0.554)^5 = 0.1396$

$P(x=2) = {}^6C_2 (0.446)^2 (0.554)^4 = \frac{\cancel{6}^3\times5}{\cancel{2}}(0.446)^2(0.554)^4 = 0.2809$

$P(x=3) = {}^6C_3 (0.446)^3 (0.554)^3 = \frac{\cancel{6}\times5\times4}{\cancel{3\times2}}(0.446)^3(0.554)^3 = 0.3016$

$P(x=4) = {}^6C_4 (0.446)^4 (0.554)^2 = \frac{6\times5}{2}(0.446)^4(0.554)^2 = 0.1821$

$P(x=5) = {}^6C_5 (0.446)^5 (0.554)^1 = 6(0.446)^5(0.554) = 0.05864$

$P(x=6) = {}^6C_6 (0.446)^6 (0.554)^0 = (0.446)^6 = 0.007866$

$\therefore$ B·D is

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

where $E(X)$ values are rounded off to nearest integer.

(11) (Pg 147 Pb45) Find the probability of getting an even number 3 or 4 or 5 times in throwing 10 dice, using B·D.

Sol: Let p = prob. of getting even no. We have 2, 4, 6 as possible even no·s on a dice $\therefore p = \frac{3}{6} = \frac{1}{2} \Rightarrow q = \frac{1}{2}$.

Given $n = 10$. To find $P(x=3) = ?$ $P(x=4) = ?$; $P(x=5) = ?$

$\therefore P(X=3) = {}^{10}C_3 \left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^7 = \frac{10\times\cancel{9}^3\times\cancel{8}^4}{\cancel{3}\times\cancel{2}\times1}\times\frac{1}{2^{10}} = \frac{120}{1024} = 0.112$

$P(X=4) = {}^{10}C_4 \left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^6 = \frac{10\times\cancel{9}\times\cancel{8}\times7}{\cancel{4}\times\cancel{3}\times\cancel{2}}\left(\frac{1}{2^{10}}\right) = \frac{210}{1024} = 0.2$

$P(X=5) = {}^{10}C_5 \left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5 = \frac{\cancel{10}^2\times9\times\cancel{8}\times7\times\cancel{6}}{\cancel{5}\times\cancel{4}\times\cancel{3\times2}}\left(\frac{1}{2^{10}}\right) = \frac{252}{1024} = 0.246$

$\Big\}\ .558$

⑥ Find the probability of getting an even number 3 or 4 or 5 times ~~with~~ in throwing 10 dice, using B.D.

Sol: Given $n = 10$

We know no. of even no.s on dice $= 3$ ($\because 2,4,6$ are the no.s)

$\therefore p =$ prob. of getting an even number $= \frac{3}{6} = \frac{1}{2}$

$\Rightarrow q = 1-p = \frac{1}{2}$.

$P(X=3) + P(X=4) + P(X=5) = {}^{10}C_3\, p^3 q^7 + {}^{10}C_4\, p^4 q^6 + {}^{10}C_5\, p^5 q^5$

$= \frac{10\times 9\times 8}{3\times 2}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^7 + \frac{10\times 9\times 8\times 7}{4\times 3\times 2}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^6 + \frac{10\times 9\times 8\times 7\times 6}{5\times 4\times 3\times 2}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5$

$= \left(\frac{1}{2}\right)^{10}\left[120 + 210 + 252\right] = 0.112 + 0.2 + 0.246$

$= \frac{582}{1024} =$

⑦ A coin is biased in a way that a head is twice as likely to occur as a tail. If the coin is tossed 3 times, find the probability of getting 2 tail and 1 head.

Given $n=3$, To find $P(X = 2\text{ Tail } \& \text{ 1 Head}) = ?$

Sol: $P(H) = 2P(T)$; $\Rightarrow 2P(T) + P(T) = 1 \Rightarrow P(T) = \frac{1}{3}$

w.k.t $P(H) + P(T) = 1$ $\therefore P(H) = \frac{2}{3}$.

$P(x=2) = ? = {}^3C_2\, p^2 q^1 = \frac{3\times 2}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right) = \frac{2}{9}$ Ans.

⑧ Out of 800 families with 5 children each, how many could you expect to have ⓐ 3 boys ⓑ 5 girls ⓒ either 2 or 3 boys ⓓ atleast 1 boy? Assume equal probabilities for boys & girls.

Let p – prob. of getting boy is success.

Sol: Given $N = 800$, $n = 5$

$\therefore p = \frac{1}{2}$; $q = \frac{1}{2}$ {$\because$ equal probabilities for boys & girls}

ⓐ $P(X=3) = {}^5C_3\, p^3 q^2 = \frac{5\times 4}{2}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{5}{16}$ per family.

$\therefore$ Total no. of families with 3 boys $= N \cdot P(x) = \frac{800}{16}\times 5 = 250$ families

ⓑ $P(X=0)$ ($\because$ all are girls ie 5 girls) $= {}^5C_0\, p^0 q^5 = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$

$\therefore$ Total no. of families with 5 girls $= 800\times\frac{1}{32} = 25$ families.

(c) $P(x=2)+P(x=3) = {}^5c_2\, p^2 q^3 + {}^5c_3\, p^3 q^2 = \frac{5\times\cancel{4}^{2}}{\cancel{2}}\left(\frac{1}{2}\right)^5 + \frac{5\times\cancel{4}^{2}}{\cancel{2}}\left(\frac{1}{2}\right)^5 = \frac{\cancel{20}^{5}}{\cancel{32}_{8}} = \frac{5}{8}$

$\therefore$ total no. of families with either 2 or 3 boys $= \cancel{800}^{100} \times \frac{5}{\cancel{8}} = 500$ families.

(d)

Problems on Poisson Distribution.

① (Pg 159 Pb 5) If a bank received on the average 6 bad cheques per day, find the probability that it will receive 4 bad cheques on any day Sol, $\lambda=6$ $P(X=4)=0.1339$..

Sol: $P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$; Given $\lambda=6$ To find $P(X=4)=?$

$\therefore P(X=4) = \frac{e^{-6}6^4}{4!} = \frac{54}{e^6} = 0.1339$.

② A car-hire firm has 2 cars which it hires out day by day. The no. of demands for a car on each day is distributed as a Poisson distribution with mean 1.5. Calculate the proportion of day (i) On which there is no demand (ii) on which demand is refused. (Pg 158 Pb 2)

Sol. Given $\lambda=1.5$; Total cars $=2$; $N=365$ days in a year

No demand $\Rightarrow$ cars were not hired $\Rightarrow P(X=0)=?$

No. of days when there is no demand in a year. $= E(X=0) = ?$ $= N \cdot P(X=0)$

(i) $P(X=0) = \frac{e^{-1.5}(1.5)^0}{0!} = e^{-1.5} = 0.2231$.

$\therefore$ No. of days in a year when there is no demand of car $= N \cdot P(X=0)$
$= 365 \times 0.2231$
$= 81$ days.

(ii) First we find probability on which demand is refused. w.r.t refusal takes place only when demand is for more than 2 cars. ie To find $P(X>2)$

To calculate proportion of days when there is refusal of demand $= E(X>2) = ?$

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

$= 1-e^{-1.5}[0+(1.5)+\frac{(1.5)^2}{2}] = 0.1913$

$E(X>2) = NP(X>2) = 365(0.1913) = 69.82 \simeq 70$ days

(3) (Pg 160 PbP) It has been found that 2% of the tools produced by a certain machine are defective. What is the probability that in a shipment of 400 such tools (a) 3% or more (b) 2% or ~~more~~ less will prove defective.

Sol: No. of tools = 400.

% of defective tools = 2% = P

λ = No. of defective tools for 400 = $nP = (2\%)(400)$

$= \cancel{400} \times \frac{2}{\cancel{100}}$

$\Rightarrow \boxed{\lambda = 8}$

(i) $P(x \geq 3\%) = ?$

$3\% \Rightarrow \frac{3}{\cancel{100}} \times 4\cancel{00} = 12$ tools

$\therefore P(x \geq 3\%) = P(x \geq 12) = 1 - P(x \leq 11)$

$= 1 - e^{-\lambda}\left[\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \cdots + \frac{\lambda^{11}}{11!}\right]$

$= 1 - e^{-8}\left[1 + 8 + \frac{8^2}{2} + \frac{8^3}{3!} + \cdots + \frac{8^{11}}{11!}\right]$

$= 1 - e^{-8}\left[2647.29\right]$

$= 0.1119.$

(ii) $P(x \leq 2\%)$

$2\% \Rightarrow \frac{2}{100} \times 400 = 8$ tools

$\therefore P(x \leq 2\%) = P(x \leq 8) = P(x=0) + P(x=1) + P(x=2) + P(x=3) + P(x=4) + P(x=5) + P(x=6) + P(x=7) + P(x=8)$

$= e^{-\lambda}\left[\lambda^0 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} + \frac{\lambda^5}{5!} + \frac{\lambda^6}{6!} + \frac{\lambda^7}{7!} + \frac{\lambda^8}{8!}\right]$

$= e^{-8}\left[1 + 8 + \frac{\overset{32}{\cancel{64}}}{\cancel{2}} + \frac{8^3}{3!} + \frac{8^4}{4!} + \frac{8^5}{5!} + \frac{8^6}{6!} + \frac{8^7}{7!} + \frac{8^8}{8!}\right]$ $\frac{16777216}{40320}$

$= e^{-8}\left[9 + 32 + 85.3333 + 170.6667 + 273.0667 + 364.0889 + 416.1016 + 416.1016\right]$

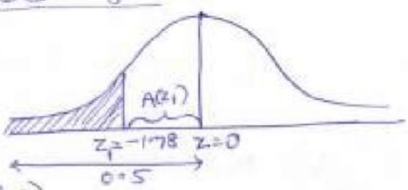
$= e^{-8}\left[\cancel{1756.42\ 3888}\right] = \underline{0.5925}$

Problems related to Normal Distribution.

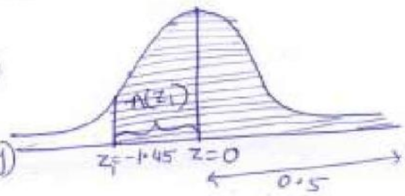
1. (Pg 205 pb10) If X is a normal variate, find the area A (i) to the left of $z = -1.78$ (ii) to the right of $z = -1.45$ (iii) corresponding to $-0.8 \leq z \leq 1.53$ (iv) to the left of $z = -2.52$ and to the right of $z = 1.83$.

Sol: (i) Required Area 'A' is 'shaded region'.

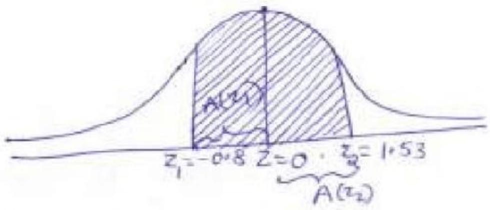
$$\text{Area} = 0.5 - A(z_1)$$
$$= 0.5 - A(-1.78)$$
$$= 0.5 - A(1.78)$$
$$= 0.5 - 0.4625 \text{ (from tables)}$$
$$= 0.0375$$



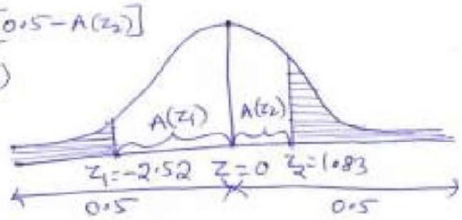
(ii) Required Area $= 0.5 + A(z_1)$
$$= 0.5 + A(-1.45)$$
$$= 0.5 + A(1.45) \text{ (By symmetry)}$$
$$= 0.5 + 0.4265$$
$$= 0.9265$$



(iii) Area $= A(z_1) + A(z_2)$
$$= A(-0.8) + A(1.53)$$
$$= A(0.8) + A(1.53)$$
$$= 0.2881 + 0.437$$
$$= 0.7251$$



(iv) Required Area $= [0.5 - A(z_1)] + [0.5 - A(z_2)]$
$$= 0.5 - A(-2.52) + 0.5 - A(1.83)$$
$$= 1 - A(2.52) - A(1.83)$$
$$= 1 - 0.4941 - 0.4664$$
$$= 1 - 0.9605$$
$$= 0.0395$$



2. (Pg 208 Pb13) If the masses of 300 students are normally distributed with mean 68 kgs & standard deviation 3 kgs, how many students have masses

(i) Greater than 72 kg

(ii) Less than or equal to 64 kg

(iii) Between 65 & 71 kg inclusive.

Sol. No. of students = 300 = N.

$\mu = 68$; $\sigma = 3$

(i) No. of students with mass greater than 72 kg = $N \cdot P(X)$ where $x = 72$.

$z = \frac{x-\mu}{\sigma}$ $\Rightarrow z = \frac{72-68}{3} = \frac{4}{3} = 1.33$

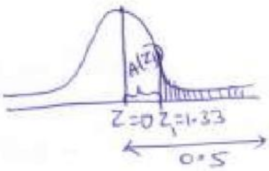
$\therefore P(X > 72) = P(Z > 1.33)$

$= 0.5 - A(z_1)$

$= 0.5 - A(1.33)$

$= 0.5 - 0.4082$

$= 0.0918$



$\therefore$ No. of students $= 300 \times 0.0918 = \boxed{28}$ students.

(ii) less than or equal to 64 kg = $P(X \le 64) = ?$

$Z = \frac{x-\mu}{\sigma} = \frac{64-68}{3} = -\frac{4}{3} = -1.33$

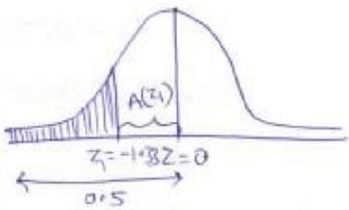
$\therefore P(X \le 64) = P(Z \le -1.33)$

$= 0.5 - A(z_1)$

$= 0.5 - A(-1.33)$

$= 0.5 - A(1.33)$

$= 0.5 - 0.4082 = 0.0918$.



$\therefore$ No. of students $= 300 \times 0.0918 = \boxed{28}$ students.

(iii) $P(65 \le X \le 71)$ $\quad z_1 = \frac{65-68}{3} = -\frac{3}{3} = -1$; $z_2 = \frac{71-68}{3} = -1$

$= P(-1 \le Z \le 1)$

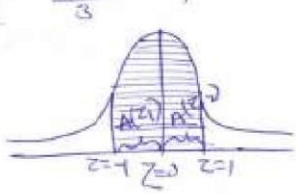
$= A(-1) + A(1)$

$= 2A(1)$

$= 2(0.3413)$

$= 0.6828$

$\therefore$ No. of students $= 300 \times 0.6828 = \boxed{205}$ students



3. (Pg 213 Pb 20) If X is normally distributed with mean 2 & S.D 0.1, then find $P(|X-2| \geq 0.01)$?

Sol: $\mu = 2$, $\sigma = 0.1$

$P(|X-2| \geq 0.01) = 1 - P(|X-2| < 0.01)$

$|X-2| < 0.01 \Rightarrow x = 2+0.01$ & $x = +2-0.01$

$= 2.01$ & $+1.99$

Let $x_1 = 1.99$, $x_2 = 2.01$

$z = \frac{x-\mu}{\sigma} \Rightarrow z_1 = \frac{1.99-2}{0.1} = \frac{-0.01}{0.1} = -0.1$

Also $z_2 = \frac{2.01-2}{0.1} = \frac{0.01}{0.1} = 0.1$

$\therefore P(|X-2| \geq 0.01) = 1 - P(z_1 < Z < z_2)$

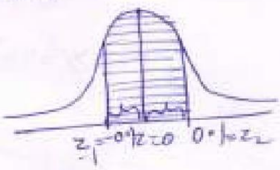
$= 1 - P(-0.1 < Z < 0.1)$

$= 1 - [A(z_1) + A(z_2)]$

$= 1 - 2A(0.1)$

$= 1 - 2(0.0398)$

$= 1 - 0.0796 = 0.9204$



4. (Pg 222 Pb 20) Suppose the weights of 800 male students are normally distributed with mean 28.8 kg & S.D 2.06 kg. Find the no. of students whose weights are b/w (i) 28.4 kg & 30.4 kg (ii) more than 31.3 kg.

$N = 800$, mean $\mu = 28.8$; $\sigma = 2.06$

(i) $P(28.4 \leq x \leq 30.4) = ?$

No. of students $= N \times P(28.4 \leq X \leq 30.4)$; $z = \frac{x-\mu}{\sigma}$

$x_1 = 28.4$, $x_2 = 30.4$

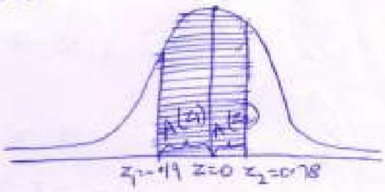
$z_1 = \frac{28.4-28.8}{2.06} = -\frac{0.2}{1.03} = \frac{-20}{103} = -0.1942 \simeq 0$

$= -0.19$

$z_2 = \frac{30.4-28.8}{2.06} = \frac{0.8}{1.03} = \frac{80}{103} = 0.7767$

$= 0.78$

$$P(28.4 \le x \le 30.4) = P(z_1 \le Z \le z_2)$$
$$= P(-0.19 \le Z \le 0.78)$$
$$= A(z_1) + A(z_2)$$
$$= A(-0.19) + A(0.78)$$
$$= 0.0753 + 0.2823$$
$$= 0.3576$$

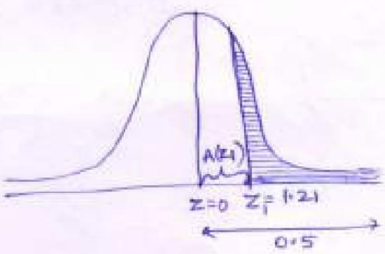


$\therefore$ No. of students $= N\,P(28.4 \le x \le 30.4)$
$= 800 \times 0.3576$
$= 286$

(ii) more than 31.3kg.

$P(x \ge 31.3)$

$$z = \frac{31.3 - 28.8}{2.06} = \frac{2.5}{2.06} = 1.21$$



$$P(x \ge 31.3) = P(Z \ge 1.21)$$
$$= 0.5 - A(z_1)$$
$$= 0.5 - A(1.21)$$
$$= 0.5 - 0.3869$$
$$= 0.1131$$

$\therefore$ No. of students $= N\,P(x \ge 31.3) = 800 \times 0.1131 = 90$

(Pg 200 Pb3)

✗ (5) In a Normal distribution, 7% of the items are under 35 and 89% are under 63. Determine mean & variance of the distribution.

$P(x < 35) = 7\%$ $\quad P(x < 63) = 89\%$ $\quad \mu = ?$ $\sigma^2 = ?$

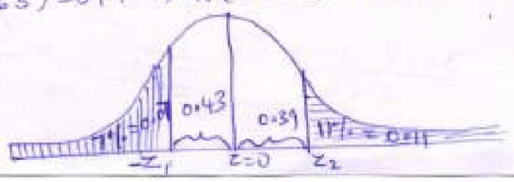
$x_1 = 35$, $x_2 = 63$ $\Rightarrow z_1 = \frac{35-\mu}{\sigma}$; $z_2 = \frac{63-\mu}{\sigma}$.

$\because$ Total $= 100$ {%} we consider values of $x < 50$ to be left of $z=0$ {ie $x=\mu$} & values of $x > 50$ to be rt - of $z=0$. $\Rightarrow z_1 < 0$ & $z_2 > 0$ $\Rightarrow \frac{35-\mu}{\sigma} = -z_1$ (say)

$\frac{63-\mu}{\sigma} = z_2$ (say). $P(x<35) = 7\% \Rightarrow A(z < z_1) = 7\% \Rightarrow A(z_1) = 0.5 - 0.07 = 0.43$

$P(x<63) = 89\% \Rightarrow A(z < z_2) = 89\% \Rightarrow A(z \ge z_2) = 11\%$. The fig is

$\Rightarrow A(z_2) = 0.5 - 0.11$
$= .39$



Next page.

Given $P(x<35) = 7\% = 0.07$; $P(x<63) = 89\% = 0.89$

when $x = 35$, $z = \frac{x-\mu}{\sigma} = \frac{35-\mu}{\sigma} = -z_1$ say — (1)

when $x = 63$, $z = \frac{x-\mu}{\sigma} = \frac{63-\mu}{\sigma} = z_2$ say — (2)

From the fig: $P(0<z<z_2) = 0.39 \Rightarrow A(z_2) = 0.39$

& $P(0<z<z_1) = 0.43 \Rightarrow A(z_1) = 0.43$.

Using tables we find $z_1 = 1.48$, $z_2 = 1.23$.

$\Rightarrow z_1 = -1.48$, $z_2 = 1.23$.

∴ (1) & (2) are $\frac{35-\mu}{\sigma} = -1.48 \Rightarrow 35-\mu = -1.48\sigma$.

$\Rightarrow \mu - 1.48\sigma = 35$ — (3)

$\frac{63-\mu}{\sigma} = 1.23 \Rightarrow 63 = \mu + 1.23\sigma \Rightarrow \mu + 1.23\sigma = 63$ — (4)

Solving (3) & (4) $\sigma = \frac{28}{2.71} = 10.332$

$\Rightarrow \boxed{\sigma^2 = 106.75}$

From (3) $35 = \mu - 1.48(10.332) \Rightarrow \mu = 35 + 15.3 = 50.3$

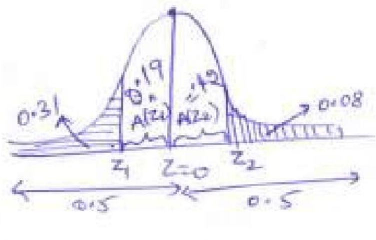
$\boxed{\therefore \mu = 50.3}$

*. (6) In a N.D 31% of the items are under 45 & 8% are over 64. Find the mean & variance of the distribution. (Pg 202 P64)

$P(x<45) = 31\% = 0.31$; $P(x>64) = 8\% = 0.08$

$z = \frac{x-\mu}{\sigma} \Rightarrow \underset{(say)}{-z_1} = \frac{45-\mu}{\sigma}$; $z_2 = \frac{64-\mu}{\sigma}$.

The fig.:



From table

Area $= 0.19$ when $z = 0.5$

$\Rightarrow \boxed{z_1 = -0.5}$

Area $= 0.42$ when $z = 1.4 \Rightarrow$

$\Rightarrow \boxed{z_2 = 1.4}$

$\Rightarrow \frac{45-\mu}{\sigma} = -0.5 \Rightarrow 45-\mu = -0.5\sigma \Rightarrow \mu - 0.5\sigma = 45$ — ①

$\frac{64-\mu}{\sigma} = 1.4 \Rightarrow 64 = \mu + 1.4\sigma$ — ②

from ① & ② we get $\sigma = \frac{19}{1.9} = 10 \Rightarrow \boxed{\sigma = 10}$

From ① : $\mu = 45 + 0.5\sigma = 45 + 0.5(10) = 50 \Rightarrow \boxed{\mu = 50}$

⑦ (Pg 203 P67) The marks obtained in mathematics by 1000 students is normally distributed with mean 78% & S.D 11%. Determine

(i) How many students got marks above 90%

(ii) What was the highest mark obtained by the lowest 10% of the students

(iii) Within what limits did the middle of 90% of the students lie.

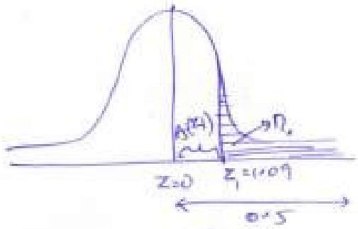
Sol: No. of students = 1000 ; $\mu = 78\% = 0.78$; $\sigma = 11\% = 0.11$

(i) $P(X > 90\%) = ?$ $\quad z = \frac{x-\mu}{\sigma} = \frac{0.9 - 0.78}{0.11} = 1.09 = Z_1$

$P(X > 90\%) = P(Z > Z_1) = P(Z > 1.09) = 0.5 - A(Z_1 = 1.09)$

$= 0.5 - 0.3621$

$= .1379$

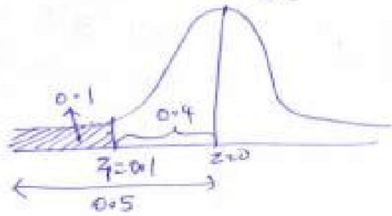


∴ No. of students who got marks above 90% $= 1000\, P(X > 90\%)$

$= 1000 \times 0.1379$

$= 137.9$ ~~stud~~

$= 138$ students.



(ii) $P(X < x_1) = 10\% = 0.1$

$\Rightarrow P(z < z_1) = 0.1$.

$\Rightarrow A(z_1) + A(z < z_1) = 0.5$

$\Rightarrow A(z_1) = 0.5 - 0.1 = 0.4$

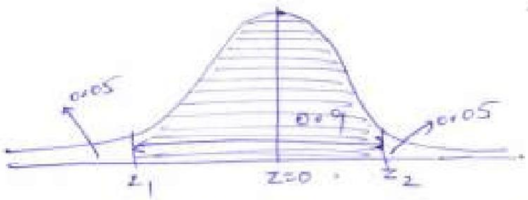
From Tables $Z_1 = 1.28 \Rightarrow Z_1 = -1.28$ { ∵ left side of z }.

$-1.28 = \frac{x-\mu}{\sigma} = \frac{x - 0.78}{0.11} \Rightarrow \underline{x = 0.6392}$

Hence the highest mark obtained by the lowest 10% of students $= 0.6392 \times 100 = 64\%$.

Next page.

7(iii)



To find $z_1 = ?$ $z_2 = ?$ such that $A(z_1) + A(z_2) = 0.9$

Middle 90% ⇒ Area in the middle = 0.9.
⇒ Area on both sides left over = 0.05.

From table : value of z where area = 0.45 is $z = 1.64$

∵ Middle area = 0.9 ⇒ $A(z_1) = A(z_2) = \frac{0.9}{2} = 0.45$

From table Area = 0.45 for $z = 1.64$.

$\therefore z_1 = -1.64$; $z_2 = 1.64$.

$\Rightarrow \frac{x_1 - \mu}{\sigma} = -1.64$; $\frac{x_2 - \mu}{\sigma} = 1.64$.

$\Rightarrow x_1 = \mu - 1.64\sigma$; $x_2 = \mu + 1.64\sigma$.

Given $\mu = .78$; $\sigma = 0.11$.

$\therefore x_1 = 0.5996$; $x_2 = 0.9604$.

∴ Limits of 90% are $0.5996 \times 100 = 59.96 = 60\%$
& $0.9604 \times 100 = 96.04 = 96\%$.

⇒ B/W 60 & 96 ~~students~~ marks.

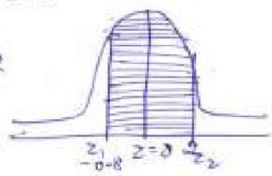
8. If X is a normal variate with mean 30 & standard deviation 5, find the probabilities that (i) $26 \le X \le 40$ (ii) $X \ge 45$.

Sol: $\mu = 30$, $\sigma = 5$; $z = \frac{x - \mu}{\sigma}$

(i) $P(26 \le X \le 40) = ?$

$x_1 = 26 \rightarrow z_1 = \frac{26 - 30}{5} = -\frac{4}{5} = -0.8$

$x_2 = 40 \Rightarrow z_2 = \frac{40 - 30}{5} = \frac{10}{5} = 2$



Thus $P(26 \le X \le 40) = P(-0.8 \le Z \le 2)$

$= A(0.8) + A(2)$

$= 0.2881 + 0.4772$

$= 0.7653$ (using Normal distribution tables)

Problems on Joint Probability [ie Multiple Random Variables]

(1) Pg 237 Pb 2. For following bivariate p.d. of X & Y find
(i) $P(X \le 2, Y=2)$ (ii) $P(X<3, Y \le 4)$ (iii) $P(Y=3)$
(iv) $F_X(2)$ (v) $F_Y(3)$.

X/Y	1	2	3	4
1	0.1	0	0.2	0.1
2	0.05	0.12	0.08	0.01
3	0.1	0.05	0.1	0.09

(i) $P(X \le 2, Y=2) = P(X=1, Y=2) + P(X=2, Y=2)$
$= 0 + 0.12$
$= 0.12$

(ii) $P(X<3, Y \le 4) = P(X=1, Y=1) + P(X=1, Y=2) + P(X=1, Y=3) + P(X=1, Y=4) + P(X=2, Y=1) + P(X=2, Y=2) + P(X=2, Y=3) + P(X=2, Y=4)$
$= 0.1 + 0 + 0.2 + 0.1 + 0.05 + 0.12 + 0.08 + 0.01$
$= 0.66$

(iii) $P(Y=3) = P(X=1, Y=3) + P(X=2, Y=3) + P(X=3, Y=3)$
$= 0.2 + 0.08 + 0.1$
$= 0.38$.

(iv) $F_X(2) = P(X=2, Y=1) + P(X=2, Y=2) + P(X=2, Y=3) + P(X=2, Y=4) + P(X=1, Y=1) + P(X=1, Y=2) + P(X=1, Y=3) + P(X=1, Y=4)$
$\{\because F_X(2) = P(X \le 2)\}$ $= 0.05 + 0.12 + 0.08 + 0.01 + 0.1 + 0 + 0.2 + 0.1$
$= 0.66$

(v) $F_Y(3) = P(Y \le 3) = P(X=1, Y=1) + P(X=1, Y=2) + P(X=1, Y=3) + P(X=2, Y=1) + P(X=2, Y=2) + P(X=2, Y=3) + P(X=3, Y=1) + P(X=3, Y=2) + P(X=3, Y=3)$
$= 0.1 + 0 + 0.2 + 0.05 + 0.12 + 0.08 + 0.1 + 0.05 + 0.1$
$= 0.70$

Unit II

Correlation

INTRODUCTION

In a bivariate distribution and multivariate distributions we may be interested to find if there is any relationship between the two variables under study.

The Correlation is a statistical tool which studies the relationship between two variables and Correlation analysis involves various methods and techniques used for studying and measuring the extent of the relationship between them.

→ Two variables are said to be Correlated if the change in one variable results in a corresponding change in the other variable.

Types of Correlation

1) Positive and Negative Correlation

If the values of the two variables deviate in the same direction i;e If the increase in the values of one variable results, on an average, in a corresponding increase in the values of the other variable (or) If the decrease in the values of one variable results, on an average, in a corresponding decrease in the value of the other variable, Correlation is said to be positive Correlation.

Eg:- 1. Heights and weights of the individuals.
2. The family income and expenditure on luxury items
3. Amount of Rainfall and yield of the Crop.

→ If the increase (decrease) in one Variable results on an average, in a Corresponding decrease (Increase) in the values of the other variable, Correlation is Said to be <u>Negative Correlation</u>

Eg: 1. Price and demand of a Commodity.
2. Sale of wollen garments and the day temperature
3. Volume and pressure of a Perfect gas.

2. <u>Linear and NonLinear Correlation</u>

→ The Correlation between two Variables is Said to be <u>linear</u> if Corresponding to a unit change in one Variable, there is a Constant Change in the other variable over the entire range of the values. (or)

Two Variables x and y are said to be linearly related if there exists a relationship of the form $y=a+bx$.

→ Two Variables are Said to be non linear or Curvelinear if Corresponding to a unit change in one Variable, the other Variable doesnot change at a Constant rate but at fluctuating rate.

Methods of studying Correlation

1) Scatter diagram method.
2) Karl pearson's coefficient of correlation.
3) Rank method.

Scatter Diagram Method

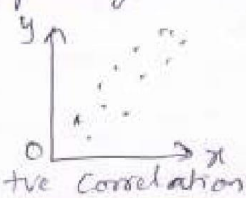
If 'n' pair of values for two variables x and y are given as $(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)$. Then these points are dotted on the x-axis and y-axis in the xy plane. The diagram of dots so obtained is known as scatter diagram.

Note:- This method gives a fairly good, though rough, idea about the relationship between the two variables.

→ If the points are very dense (close to each other), then we say a fairly good amount of correlation between two variables.

→ If the points are widely scattered, a poor correlation between those variables.

→ If there is an upward trend rising from ~~left~~ lower left hand corner and going upward to the upper right hand corner, the correlation is positive



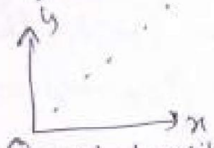
+ve Correlation

Negative Correlation

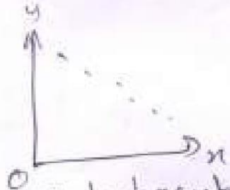
→ If the points depict a downward trend from the upper left hand corner to the lower right hand corner, the correlation is <u>negative</u>.

→ If all the points, lie on straight line from the left bottom and ~~getting~~ going up towards the right top, the correlation is said to be <u>perfect positive</u>

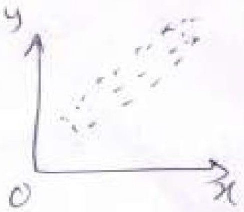
→ If all the points, lie on straight line from the left top and coming down toward right bottom, the correlation is said to be perfect Negative



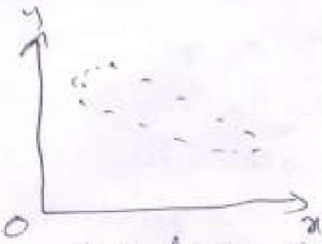
perfect positive



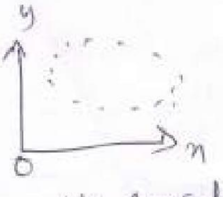
Perfect negative



High degree of +ve correlation



High degree of −ve correlation



No correlation

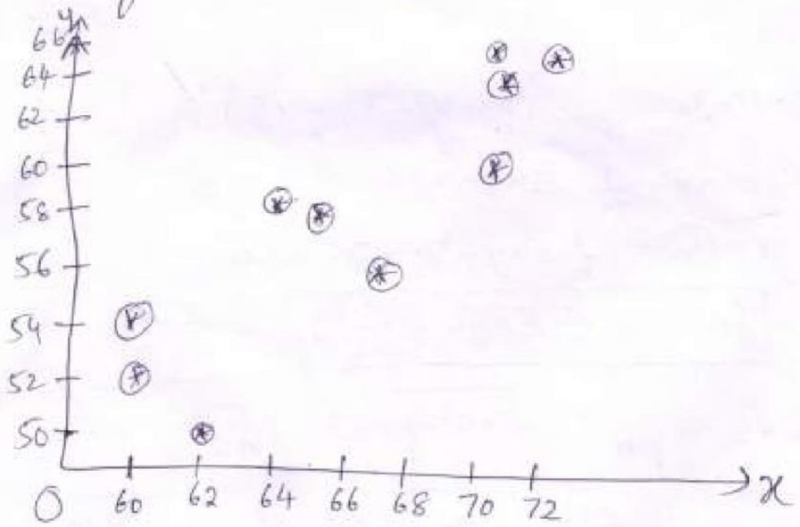
(3)

Prob1 Draw a scatter diagram from the following data

Height (inches) : 62 72 70 60 67 70 64 65 60 70
Weight (lbs) : 50 65 63 52 56 60 59 58 54 65

Also indicate whether Correlation is positive or negative.

Sol.^n Let us take Heights as random variable 'x' and taking those values on X-axis.
Let us take weights as variable 'y' and taking its values on Y-axis then we get scatter diagram as follows.



Since the points are close to each other we can say High degree of Correlation (positive) between the Series of heights and weights.

Karl Pearson's Coefficient of Correlation.

Pearson's coefficient of correlation between two variables (series) X and Y is denoted by 'r' is a measure of linear relationship between them and is defined as

$$r = \frac{Cov(x,y)}{\sigma_x \sigma_y} \longrightarrow (1)$$

where $Cov(x,y) = \frac{1}{n} \Sigma (x-\bar{x})(y-\bar{y})$

$$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} \text{ and } \sigma_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}}$$

* Note 1:

$$Cov(x,y) = \frac{1}{n} \Sigma (xy - x\bar{y} - \bar{x}y + \bar{x}\bar{y})$$

$$Cov(x,y) = \frac{1}{n} \Sigma xy - \frac{\Sigma x}{n}\bar{y} - \bar{x}\frac{\Sigma y}{n} + \frac{\bar{x}\bar{y}}{n}\Sigma 1$$

$$Cov(x,y) = \frac{1}{n}\Sigma xy - \bar{x}\bar{y} - \cancel{\bar{x}\bar{y}} + \cancel{\bar{x}\bar{y}}$$

$$\boxed{Cov(x,y) = \frac{1}{n}\Sigma xy - \frac{\Sigma x}{n}\left(\frac{\Sigma y}{n}\right)} \longrightarrow (2)$$

Sly

$$\sigma_x = \sqrt{\frac{1}{n}\Sigma(x^2 + \bar{x}^2 - 2x\bar{x})} = \sqrt{\frac{1}{n}\Sigma x^2 + \bar{x}^2 - 2\bar{x}^2}$$

$$\therefore \sigma_x = \sqrt{\frac{1}{n}\Sigma x^2 - \left(\frac{\Sigma x}{n}\right)^2} = \frac{1}{n}\sqrt{n\Sigma x^2 - (\Sigma x)^2}$$

$$\therefore \sigma_x = \frac{1}{n}\sqrt{n\Sigma x^2 - (\Sigma x)^2} \longrightarrow (3)$$

$$\text{and } \sigma_y = \frac{1}{n}\sqrt{n\Sigma y^2 - (\Sigma y)^2} \longrightarrow (4)$$

Sub (2), (3), (4) in (1) we get

$$\boxed{r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}} \longrightarrow (5)$$

(4)

Note 2:- $r = \dfrac{\Sigma dx\, dy}{\sqrt{\Sigma dx^2\, \Sigma dy^2}} \rightarrow (6)$ where $dx = x - \bar{x}$, $dy = y - \bar{y}$.

→ This formula is used when $\bar{x}$ and $\bar{y}$ are not fractionals (i.e not decimals or when natural numbers)

→ If the values of x and y are large, $\bar{x}, \bar{y}$ are fractionals then using this formula becomes tedious. So, we use step deviation method

→ If the values of x and y are small but $\bar{x}, \bar{y}$ are fractionals then we use formula (5)

Properties of Correlation coefficient

1. The correlation coefficient lies between -1 and 1 i.e $-1 \le r \le 1$.

Proof:- we know that $\Sigma \left[\dfrac{x-\bar{x}}{\sigma_x} \pm \dfrac{y-\bar{y}}{\sigma_y} \right]^2 \ge 0$

$$\Sigma \left\{ \left[\frac{x-\bar{x}}{\sigma_x} \right]^2 + \left[\frac{y-\bar{y}}{\sigma_y} \right]^2 \pm \frac{2(x-\bar{x})(y-\bar{y})}{\sigma_x \sigma_y} \right\} \ge 0$$

$$\Rightarrow \frac{\Sigma (x-\bar{x})^2}{\sigma_x^2} + \frac{\Sigma (y-\bar{y})^2}{\sigma_y^2} \pm \frac{2}{\sigma_x \sigma_y} \Sigma (x-\bar{x})(y-\bar{y}) \ge 0$$

Dividing with 'n' on both sides, we get

$$\frac{\sigma_x^2}{\sigma_x^2} + \frac{\sigma_y^2}{\sigma_y^2} \pm \frac{2\, Cov(x,y)}{\sigma_x \sigma_y} \ge 0 \quad \left[\because \frac{\Sigma (x-\bar{x})^2}{n} = \sigma_x^2 \right]$$

$$2 \pm 2r \ge 0$$

$2 - 2r \ge 0$	$2 + 2r \ge 0$
$2 \ge 2r \Rightarrow 1 \ge r$	$2 \ge -2r$
$\Rightarrow r \le 1$	$1 \ge -r$
	$-1 \le r$

$\boxed{\therefore -1 \le r \le 1}$

$\rightarrow$ If $r=1 \Rightarrow$ +ve correlation (perfect)
$r=-1 \Rightarrow$ −ve Correlation (perfect)
$r=0 \Rightarrow$ Uncorrelated (No Correlation)

2. The coefficient of Correlation is independent of the change of origin and scale.

Proof:- Let x and y are transformed into new variables u and v by the change of origin and scale

i.e $\boxed{u=\dfrac{x-A}{h}}$, $\boxed{v=\dfrac{y-B}{K}}$

where $h,K>0$ and A,B,h,k are Constants

then $x=A+uh$ and $y=B+vK$

$$\left.\begin{aligned}\Sigma x &= \Sigma(A+uh)\\ \frac{\Sigma x}{n} &= \frac{nA}{n}+\frac{h\Sigma u}{n}\\ \Rightarrow \bar{x} &= A+\bar{u}h\end{aligned}\right\} \text{ slly we get } \bar{y}=B+\bar{v}K$$

Now $x-\bar{x}=h(u-\bar{u})$ and $y-\bar{y}=K(v-\bar{v})$

$$\therefore r_{xy}=\frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\,\sigma_x\sigma_y}=\frac{\Sigma hK(u-\bar{u})(v-\bar{v})}{n\sqrt{\dfrac{\Sigma(u-\bar{u})^2h^2}{n}\,\dfrac{\Sigma(v-\bar{v})^2K^2}{n}}}$$

$$\therefore r_{xy}=\frac{\Sigma(u-\bar{u})(v-\bar{v})}{\sqrt{\Sigma(u-\bar{u})^2\Sigma(v-\bar{v})^2}}=\frac{Cov(u,v)}{\sigma_u\sigma_v}=r_{uv}$$

$$\boxed{\therefore r_{xy}=r_{uv}}$$

Note:- $r=\dfrac{\Sigma du\,dv}{\sqrt{\Sigma du^2\Sigma dv^2}}=\dfrac{n\Sigma uv-\Sigma u\Sigma v}{\sqrt{[n\Sigma u^2-(\Sigma u)^2][n\Sigma v^2-(\Sigma v)^2]}}$ is used when $\bar{x}$ or/and $\bar{y}$ are fractionals or if x,y values are large.

Problem 1. Calculate coefficient of correlation for the following data

x	9	8	7	6	5	4	3	2	1
y	15	16	14	13	11	12	10	8	9

Solution Given $n = 9$.

$\therefore \bar{x} = \frac{\Sigma x}{n} = \frac{45}{9} = 5$ and $\bar{y} = \frac{\Sigma y}{n} = \frac{108}{9} = 12$.

Since the given values are small and $\bar{x}, \bar{y}$ are not fractionals,

correlation coefficient is given by

$$r = \frac{\Sigma dx\,dy}{\sqrt{\Sigma dx^2 \Sigma dy^2}} \quad \text{where} \quad dx = x - \bar{x}, \quad dy = y - \bar{y}$$

Now the table is

x	y	$dx = x - \bar{x}$ ($x-5$)	$dy = y - \bar{y}$ ($y-12$)	dx^2	dy^2	$dx\,dy$
9	15	4	3	16	9	12
8	16	3	4	9	16	12
7	14	2	2	4	4	4
6	13	1	1	1	1	1
5	11	0	-1	0	1	0
4	12	-1	0	1	0	0
3	10	-2	-2	4	4	4
2	8	-3	-4	9	16	12
1	9	-4	-3	16	9	12
$\Sigma x = 45$	$\Sigma y = 108$			$\Sigma dx^2 = 60$	$\Sigma dy^2 = 60$	$\Sigma dx\,dy = 57$

$$\therefore r = \frac{57}{\sqrt{60(60)}} = \frac{57}{60} = 0.95$$

$\therefore$ There is a high degree of positive correlation between x and y series.

2. Find correlation coeff from

x	10	12	18	24	23	27
y	13	18	12	25	30	10

3. Find Karl pearson's Coefficient of Correlation from

wages	100	101	102	102	100	99	97	98	96	95
Cost of living	98	99	99	97	95	92	95	94	90	91

Sol. Given $n = 10$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{2(100)+101+2(102)+99+97+98+96+95}{10} = \frac{990}{10} = 99$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{950}{10} = 95$$

The table is

x	y	$dx = x-\bar{x}$ $x-99$	$dy = y-\bar{y}$ $y-95$	dx^2	dy^2	$dxdy$
100	98	1	3	1	9	3
101	99	2	4	4	16	8
102	99	3	4	9	16	12
102	97	3	2	9	4	6
100	95	1	0	1	0	0
99	92	0	-3	0	9	0
97	95	-2	0	4	0	0
98	94	-1	-1	1	1	1
96	90	-3	-5	9	25	15
95	91	-4	-4	16	16	16
990 Σx	950 Σy			$\Sigma dx^2=54$	$\Sigma dy^2=96$	$\Sigma dxdy = 61$

$$\therefore r = \frac{\Sigma dxdy}{\sqrt{\Sigma dx^2 \Sigma dy^2}} = \frac{61}{\sqrt{54(96)}} = \frac{61}{705.453} = \cancel{0.080}\ 0.847$$

$$\boxed{\therefore r = 0.847}$$

(6)

4. Calculate the coeff of correlation between age of cars and annual maintenance cost and comment

Age of Cars (yrs)	2	4	6	7	8	10	12
Annual Maintenance cost (Rs)	1600	1500	1800	1900	1700	2100	2000

Sol. Since 'y' values are large, we use step deviation method.

Now, $n=7$, $\bar{x}=\frac{\Sigma x}{n}=\frac{49}{7}=7$

and $\bar{y}=\frac{\Sigma y}{n}=\frac{12600}{7}=1800$

let $h=1$ and $K=100$ and $A=\bar{x}$; $B=\bar{y}$

then $u=\frac{x-\bar{x}}{h}=x-7$ & $v=\frac{y-1800}{100}$

The table is

x	y	$u=x-7$	$v=\frac{y-1800}{100}$	uv	u^2	v^2
2	1600	-5	-2	10	25	4
4	1500	-3	-3	9	9	9
6	1800	-1	0	0	1	0
7	1900	0	1	0	0	1
8	1700	1	-1	-1	1	1
10	2100	3	3	9	9	9
12	2000	5	2	10	25	4
49	12600	Σu 0	Σv 0	37 Σuv	70 Σu^2	28 Σv^2

$$\therefore r=\frac{n\Sigma uv-\Sigma u\Sigma v}{\sqrt{[n\Sigma u^2-(\Sigma u)^2][n\Sigma v^2-(\Sigma v)^2]}}=\frac{7(37)}{\sqrt{7^2(70)(28)}}=\frac{37}{\sqrt{1960}}$$

$=0.8357$

$\simeq 0.836$

5. Find out the coeff of correlation from

Height of father (inches)	65	66	67	68	69	71	73
Height of son (inches)	67	68	64	72	70	69	70

Sol. Given $n=7$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{65+66+67+68+69+71+73}{7} = 68.428$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{67+68+64+72+70+69+70}{7} = 68.57$$

Since $\bar{x}, \bar{y}$ are fractionals,

let $u = \frac{x-68}{1}$ and $v = \frac{y-69}{1}$ ($\because h=k=1$)

The table is

x	y	$u=x-68$	$v=y-69$	u^2	v^2	uv
65	67	-3	-2	9	4	6
66	68	-2	-1	4	1	2
67	64	-1	-5	1	25	5
68	72	0	3	0	9	0
69	70	1	1	1	1	1
71	69	3	0	9	0	0
73	70	5	1	25	1	5
		$\Sigma u=3$	$\Sigma v=-3$	$\Sigma u^2=49$	$\Sigma v^2=41$	$\Sigma uv=19$

$$\therefore r = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{[n\Sigma u^2-(\Sigma u)^2][n\Sigma v^2-(\Sigma v)^2]}} = \frac{7(19)+9}{\sqrt{[7(49)-9][7(41)-9]}}$$

$$= \frac{142}{\sqrt{334(278)}} = \frac{142}{\sqrt{92852}} = \frac{142}{304.72} = 0.466$$

$$\boxed{\therefore r = 0.466}$$

7

7. Given $n = 10$, $\sigma_x = 5.4$, $\sigma_y = 6.2$ and sum of the product of deviation from the mean of x and y is 66. find the correlation coeff.

Sol^n Given $n=10$, $\sigma_x = 5.4$, $\sigma_y = 6.2$ and $\Sigma(x-\bar{x})(y-\bar{y}) = 66$

$$\therefore \quad r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y} = \frac{66}{10(5.4)(6.2)} = \frac{6.6}{(5.4)(6.2)} = 0.197$$

Rank Correlation Coefficient

Spearman's rank correlation is given by

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

where d_i = difference of corresponding ranks of x and y.

n = no of the terms in the series

Properties

1. $-1 \le \rho \le 1$
2. If $\rho = 1$, there is complete agreement in the order of the ranks and they are in same direction.
3. If $\rho = -1$, there is complete disagreement in the order of the ranks and they are in opposite direction.

Problems when ranks are given

1. Following are the ranks obtained by 10 students in two subjects, statistics and Maths. To what extent the knowledge of the students in two subjects is related?

Statistics	1	2	3	4	5	6	7	8	9	10
Maths	2	4	1	5	3	9	7	10	6	8

Sol. Given $n=10$; let x = Ranks in Statistics
y = Ranks in Maths

The table is

x	y	$d_i = x-y$	d_i^2
1	2	-1	1
2	4	-2	4
3	1	2	4
4	5	-1	1
5	3	2	4
6	9	-3	9
7	7	0	0
8	10	-2	4
9	6	3	9
10	8	2	4
			$\Sigma d^2 = 40$

Rank Correlation Coefficient is given by

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

$$= 1 - \frac{6(40)^4}{10(10^2-1)}$$

$$= 1 - \frac{24}{99}$$

$$= \frac{75}{99} = 0.7575$$

$$\boxed{\therefore \rho = 0.76}$$

2. A random sample of 5 students are selected and their grades in Maths and statistics are found to be

	1	2	3	4	5
Maths	85	60	73	40	90
stats	93	75	65	50	80

8

Sol. Let x = Ranks in Maths.
y = Ranks in Stats $n = 5$

[Here give ranks based on their marks in decending order]

Marks in Maths	Rank x	Marks in Stats	Rank y	$d_i = x-y$	d_i^2
85	2	93	1	1	1
60	4	75	3	1	1
73	3	65	4	-1	1
40	5	50	5	0	0
90	1	80	2	-1	1
					4

$$\therefore \rho = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)} = 1 - \frac{6(4)}{5(5^2-1)} = 1 - \frac{6(4)}{5(24)}$$

$$\therefore \rho = 1 - \frac{1}{5} = \frac{4}{5} = 0.8$$

$$\boxed{\therefore \rho = 0.8}$$

Problems on equal or Repeated Ranks

Formula

$$\rho = 1 - 6\left\{\frac{\Sigma d^2 + \frac{1}{12}(m)(m^2-1) + \frac{1}{12}m(m^2-1) + \cdots}{n(n^2-1)}\right\}$$

Here $m \rightarrow$ no of times rank is repeated.

Procedure

→ If the items are repeated in the Series then Common ranks are assigned such that the mean of the ranks which these items would have got if they were different from each other and the next item will get the rank next to the rank used in computing the common Rank.

Eg: 1. An item is repeated at rank 4 twice then

Common rank $= \frac{4+5}{2} = \frac{9}{2} = 4.5$

and next rank will be '6'.

2. An item is repeated thrice at rank 3 then

Common Rank $= \frac{3+4+5}{3} = \frac{12}{3} = 4$

and next rank will be '6'.

Prob1. The ranks of the 15 students in two subjects A and B are given below, the two numbers with in the brackets denoting the ranks of the same student in A and B resply.

(1,10), (2,7), (3,2), (4,6), (5,4), (6,8), (7,3), (8,1), (9,11), (10,15), (11,9), (12,5), (13,14), (14,12) and (15,13).

Use Spearman's formula to find the rank correlation coefficient.

Sol^n [Since ranks not repeated use

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

Ans:- 0.514

where $n = 15$,

2. The following table gives the score obtained by 11 students in English and Telugu. Find the rank correlation coeff.

Scores of English	40	46	54	60	70	80	82	85	85	90	95
Scores of Telugu	45	45	50	43	40	75	55	72	65	42	70

Sol. Given $n = 11$

Marks in English	Scores in Telugu	Rank x	Rank y	d_i	d_i^2
40	45	11	7.5	3.5	12.25
46	45	10	7.5 (m=2)	2.5	6.25
54	50	9	6	3	9
60	43	8	9	-1	1
70	40	7	11	-4	16
80	75	6	1	5	20
82	55	5	5	0	0
85	72	3.5	2	1.5	2.25
85	65	3.5 (m=2)	4	-0.4	0.16
90	42	2	10	-8	64
95	70	1	3	-2	4
					$\Sigma d_i^2 = 134.91$

∴ Spearman's rank correlation is given by

$$\rho = 1 - 6\left\{\frac{\Sigma d^2 + \frac{1}{12}m(m^2-1) + \frac{1}{12}m(m^2-1)}{n(n^2-1)}\right\}$$

$$= 1 - 6\left\{\frac{134.91 + \frac{1}{\cancel{6}}2(3)}{11(120)}\right\}$$

$$= 1 - \cancel{6}\left\{\frac{135.91}{\cancel{1320}\ 220}\right\} = 1 - 0.618 = 0.382$$

$\boxed{\therefore \rho = 0.382}$

2. A Sample of 12 fathers and their elder sons gave the following data about their elder sons. Calculate the Coeff of rank Correlation.

Fathers	65	63	67	64	68	62	70	66	68	67	69	71
Sons	68	66	68	65	69	66	68	65	71	67	68	70

Sol^n Here $n = 12$. The table is

Father X	Sons Y	Rank x	Rank y	d_i	d_i^2
65	68	9	5.5	3.5	12.25
63	66	10	9.5	1.5	2.25
67	68	6.5	5.5	1	1
64	65	10	11.5	-1.5	2.25
68	69	4.5	3	1.5	2.25
62	66	12	9.5	2.5	6.25
70	68	2	5.5	-3.5	12.25
66	65	8	11.5	-3.5	12.25
68	71	4.5	1	3.5	12.25
67	67	6.5	8	-1.5	2.25
69	68	3	5.5	-2.5	6.25
71	70	1	2	-1	1
					72.5

In X series 68 repeated 2 times, 67 repeated two times

So, $m = 2, 2$.

In Y series 68 repeated 4 times, 65 repeated 2 times, 66 repeated 2 times

So, $m = 4, 2, 2$

$$\therefore \rho = 1 - 6\left\{\frac{\sum d^2 + \frac{1}{12}2(2^2-1)(2) + \frac{1}{12}4(4^2-1)}{12(12^2-1)}\right\} = 1 - 6\left\{\frac{72.5 + 2 + 5}{12(143)}\right\}$$

$$\therefore \rho = 1 - 6\left\{\frac{79.5}{1716}\right\} = 1 - \frac{79.5}{286} = 1 - 0.278 = \underline{0.722}$$

$$\boxed{\therefore \rho = 0.722}$$

Regression

The statistical method which helps us to estimate the unknown value of one variable from the known value of the realated variable is called Regression.

The lines described in the average relationship between two variables is known as lines of Regression

Comparision between Correlation and Regression

Correlation	Regression
1. It is a measure of degree of covariability between two variables.	1. Regression establishes the functional relationship between dependent & indepndt variables
2. In correlation, both the variables are random variables	2. In Regression, one variable is dependent variable and other one is independent variable
3. The coefficient of correlation is a relative measure	3. Regression coefficient is an absolute measure

Linear Regression

→ If the Regression equation is a straight line then we say it is a Linear Regression. Otherwise we call it as Non-linear Regression

Lines of Regression

→ Line of Regression of 'y' on 'x' is the line which gives the best estimate for the value of y for any specified value of 'x'.

→ Line of Regression of 'x' on 'y' is the line which gives the best estimate for the value of x for any specified value of 'y'.

Regression Equation of y on x ⟹ $y = a + bx$ → (1)

$$na + b\Sigma x = \Sigma y$$
$$a\Sigma x + b\Sigma x^2 = \Sigma xy$$
(Normal equations)

Solving we get a, b values

Substituting in (1) we get Required

Regression eq^n of y on x.

Regression Equation of x on y is $x = a + by$ → (2)

Normal equations are

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

Solving we get a, b values

Sub in (2) to get Required

Regression Equation of x on y.

→ Regression Equations by using deviations from arithmetic mean of x and y.

Regression Equation of 'y' on 'x' is

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

where $\bar{x}$ = Mean of 'x' series $= \frac{\Sigma x}{n}$

$\bar{y}$ = Mean of 'y' series $= \frac{\Sigma y}{n}$

r = Correlation coefficient of x & y.

$\sigma_x, \sigma_y \rightarrow$ S.D of x, y series

Regression Coefficient of y on x $= r\frac{\sigma_y}{\sigma_x}$

$$= \frac{cov(x,y)}{\sigma_x \cancel{\sigma_y}} \cdot \frac{\cancel{\sigma_y}}{\sigma_x}$$

$$= \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\,\sigma_x^2}$$

$$= \frac{\Sigma XY}{\cancel{n} \cdot \frac{1}{\cancel{n}}\Sigma(x-\bar{x})^2}$$

$\Rightarrow$ Regression Coeff of y on x $= \frac{\Sigma XY}{\Sigma X^2}$

$\therefore$ Regression equation of 'y' on x is

$$\boxed{y - \bar{y} = \frac{\Sigma XY}{\Sigma X^2}(x - \bar{x})}$$

Slly Regression Equation of 'x' on y is

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

$$\Rightarrow \boxed{x - \bar{x} = \frac{\Sigma XY}{\Sigma Y^2}(y - \bar{y})}$$

Here if Regression coeff of y on 'x' is b_{yx} and regression coeff of x on 'y' is b_{xy} then

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} \quad \text{and} \quad b_{xy} = r\frac{\sigma_x}{\sigma_y}.$$

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

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

<u>Note 1</u> Regression line always passes through $(\bar{x}, \bar{y})$

<u>Prob</u> If two regression lines of y on x and x on y are $a_1x + b_1y + c_1 = 0$ and $a_2x + b_2y + c_2 = 0$ then prove that $a_1b_2 < a_2b_1$

<u>Sol</u>. Given lines can also be written as

Regression eqn of y on x is

$$y = -\frac{a_1}{b_1}x - \frac{c_1}{b_1} = -\frac{c_1}{b_1} + \left(\frac{-a_1}{b_1}\right)x.$$

$$\Rightarrow y = a + bx$$

Regression coeff of y on x $= b = \left(\frac{-a_1}{b_1}\right) = b_{yx}$

Sly we can find.

Regression coeff of x on $y = \left(\frac{-b_2}{a_2}\right) = b_{xy}$

But $r^2 = b_{yx} \cdot b_{xy} = \left(-\frac{a_1}{b_1}\right)\left(-\frac{b_2}{a_2}\right) = \frac{a_1b_2}{a_2b_1}$

Since correlation coeff $-1 \le r \le 1$

$$\Rightarrow \frac{a_1b_2}{a_2b_1} < 1 \Rightarrow \boxed{a_1b_2 < a_2b_1}$$