Joint Probability Mass function of (X,Y) $\Rightarrow$ Unit - 3

For 2 discrete Random variables X & Y the probability x_i will take by X and y_i is the probability of Y, denoted by

$$P(X = x_i, Y = y_i) = P(x_i; y_i) = P_{ij}$$

for $i = 1, 2, 3, \ldots n$

$j = 1, 2, 3, \ldots m$

then P_{ij} is called Joint P.M.F. if it satisfies following conditions. $\Rightarrow$

1. $P_{ij} \geq 0 \; \forall \; i \,\&\, j$
2. $\sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} = 1$

Joint Probability Distribution

The JPD is the set of $(x_i, y_j, \& P_{ij})$

X/Y	y_1	y_2	$\ldots$	y_m	
x_1	P_{11}	P_{12}	$\ldots$	P_{1m}	P_{1*} represents variation in a row
x_2	P_{21}	P_{22}	$\ldots$	P_{2m}	P_{2*}
$\vdots$	$\vdots$	$\vdots$			
x_n	P_{n1}	P_{n2}	$\ldots$	P_{nm}	P_{n*}
	P_{*1}	P_{*2}	$\ldots$	P_{*m}	1

Marginal Probability distribution.

Marginal probability Distribution $\rightarrow$ Let (X, Y) be 2 dimensional discrete Random variable then marginal probability function Random variable X is defined by

$$P(X = x_i) = \sum_{j=1}^{m} P_{ij} = P_{i*}$$

Let (X, Y) be 2 dimensional Discrete RV the ~~marginal~~ marginal for Y is

$$P(Y=y_j) = \sum_{i=1}^{m} P_{ij} = P_{*j}$$

Q-1. For the Bivariate Probability distribution (X, Y) then given below. Find value of

(I) $P(X \le 1)$
(II) $P(Y \le 3)$
(III) $P(X \le 1, Y \le 3)$.
(IV) $P(X \le 1 \mid Y \le 3)$
(V) $P(X+Y \le 4)$

X|Y	1	2	3	4	5	6	
0	0	0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{3}{32}$	$\frac{8}{32}$
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{5}{8}$
2							
3	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{64}$	0	$\frac{2}{64}$	$\frac{8}{64}$
	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{11}{64}$	$\frac{13}{64}$	$\frac{6}{32}$	$\frac{16}{64}$	

(I) $P(X \le 1) = P(X=1) + P(X=0)$

$= \left(\frac{1}{16}+\frac{1}{16}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}\right) + \left(0+0+\frac{1}{32}+\frac{2}{32}+\frac{2}{32}+\frac{3}{32}\right)$

$= \frac{7}{8}$

(II) $P(Y \le 3) = P(Y=3) + P(Y=2) + P(Y=1)$

$= \left(\frac{1}{32}+\frac{1}{8}+\frac{1}{64}\right) + \left(0+\frac{1}{16}+\frac{1}{32}\right) + \left(0+\frac{1}{16}+\frac{1}{32}\right)$

$= \frac{23}{64}$

(III) $P(X \le 1, Y \le 3) = P(X=0, Y \le 3) + P(X=1, Y \le 3)$ (fixing Y)

$= \left(0+0+\frac{1}{32}\right) + \left(\frac{1}{16}+\frac{1}{16}+\frac{1}{8}\right)$

$= \frac{9}{32}$

(V) - $X+Y \le 4$

$X=0, Y=4$
$X=1, Y=3$
$X=2, Y=2$

~~$= \frac{2}{32} + \frac{1}{8} + \frac{1}{16}$~~

$= P(X=0, Y=4) + P(X=1, Y=3) + P(X=2, Y=2)$

$= \left(\frac{2}{32}+\frac{1}{32}\right) + \left(\frac{1}{16}+\frac{1}{16}+\frac{1}{8}\right) + \left(\frac{1}{32}+\frac{1}{32}\right)$

$= \frac{17}{32}$

(IV) - $P(X \le 1 \mid Y \le 3) = \dfrac{P(X \le 1, Y \le 3)}{P(Y \le 3)} = \dfrac{\frac{9}{32}}{\frac{23}{64}} = \dfrac{18}{23}$

Cumulative Distributive Function.

The CDF of a 2-D Discrete R.V. (X, Y) denoted by $F(x,y)$ and defined as $F(x,y) = P(x \le x_i, y \le y_j) = \sum_{y_j \le y} \sum_{x_i \le x} P_{ij}$

$$P(X=x_i \mid Y=y_j) = \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)}.$$

Q- The Joint P.M.F. of (X,Y) is given by $P(x,y) = K(2x+3y)$; $x=0,1,2$; $y=1,2,3$. Then find marginal and conditional distribution for.

(I) $P(X=2, Y \le 2)$
(II) $P(X \le 1, Y=3)$
(III) $P(X=2)$.
(IV) $P(X \le 2)$
(V) $P(X \le 1 \mid y \le 2)$.
(VI) $P(X=0 \mid Y=3)$

X|Y	1	2	3	
0	3K	6K	9K	18K
1	5K	8K	11K	24K
2	7K	10K	13K	30K
	15K	24K	33K	72K = 1

$K = \frac{1}{72}$

for (I) $P(X=2, Y \le 2) = P(X=2, Y=2) + P(X=2, Y=1)$
$= 10K + 7K = 17K = \frac{17}{72}$

(II) $P(X \le 1, Y=3) = P(X=0, Y=3) + P(X=1, Y=3)$
$9K + 11K = \frac{20}{72}$

(III) $P(X=2) = \frac{30}{72}$ (IV) $P(X \le 2) = 18K + 24K + 30K = 1$

(v) $P(X \le 1 \mid Y \le 2) = \dfrac{P(X \le 1, Y \le 2)}{P(Y \le 2)} = \dfrac{22}{39}$

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

$= 9k + 6k + 5k + 8k = 22k$

$P(Y \le 2) = P(Y=1) + P(Y=2)$

$= 15k + 24k = 39k$

(vi) $P(X=0 \mid Y=3) = \dfrac{P(X=0, Y=3)}{P(Y=3)} = \dfrac{9}{33} = \dfrac{3}{11}$

Q-2 From the following Table find the Marginal Distribution for X & Y.

Sol → The Marginal of X ⇒

$10/28$

$P(X=1) = P(X=1, Y=0) + P(X=1, Y=1) + P(X=1, Y=2)$

$P(X=2) = \frac{15}{28} = \frac{9}{28} + \frac{3}{14} + 0$

$P(X=3) = \frac{3}{28} = \frac{3}{28} + 0 + 0$

Y\X	1	2	3	
0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{6}{14}$
2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
	$\frac{10}{28}$	$\frac{15}{28}$	$\frac{3}{28}$	1

The Marginal of Y ⇒

$P(Y=2) = 1/28 = \frac{1}{28} + 0 + 0$

$P(Y=0) = \cancel{10/28}\ 15/28 = 3/28 + \frac{9}{28} + \frac{3}{28}$

$P(Y=1) = \cancel{15/28}\ 6/14 = \frac{3}{14} + \frac{3}{14} + 0$

~~$P(Y=3) = 3/28$~~

Independent Random Variable

combined Probability

↑ Product of their individual terms.

2 R.V. X & Y are said to be independent if $P_{ij} = P_i \times P_j$

Q- Let X & Y have following joint PMF. Show that X & Y are independent.

Sol → $P_{ij} = P_{i*} \times P_{*j}$

Y\X	2	4	
1	0.10	0.15	0.25
3	0.20	0.30	0.50
5	0.10	0.15	0.25
	0.40	0.60	1

$P_{41} = 0.15$

$P_{4*} = 0.60 = 0.15 + 0.30 + 0.15$

$P_{*1} = 0.10 + 0.15 = 0.25$

$P_{41} = P_{4*} \cdot P_{*1} = .375$

$P_{21} = P_{2*} \cdot P_{*1} = 0.40 \times 0.25 =$

$P_{23} = P_{2*} \cdot P_{*3} = 0.40 \times 0.50 = 0.20$

$P_{25} = P_{2*} \cdot P_{*5} = 0.40 \times 0.25 = 0.10$

$P_{43} = P_{4*} \cdot P_{*3} = 0.60 \times 0.50 = 0.3$

$P_{45} = P_{4*} \cdot P_{*5} = 0.60 \times 0.25 = .15$

Joint Probability density function

Let (X, Y) be a 2-D continuous random variable such that.

$P\left(x - \frac{dx}{2} \le X \le x + \frac{dx}{2},\ y - \frac{dy}{2} \le Y \le y + \frac{dy}{2}\right) = f(x,y)\,dx\,dy.$

Such that $f(x,y)$ is called Joint PDF of (X, Y) if it satisfies following conditions:

(i) $f(x,y) \ge 0 \ \forall\ x, y$

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

Properties of Distribution Function

(i). $F(-\infty, y) = F(x, -\infty) = 0 \quad = 0 \ \&\ F(-\infty, \infty) = 1$ / $\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

(ii). $F(X \le x, c \le Y \le d) = F(x, d) - F(x, c)$

(iii). $0 \le F(x,y) \le 1$ & $F(x,y)$ is non-decreasing function of x & y.

Conditional Probability Function

$F(x \mid y) = \dfrac{F(x,y)}{F_Y(y)}$ → Marginal Probability of Y.

$F(y \mid x) = \dfrac{F(x,y)}{F_X(x)}$

Q-1 Let X,Y be 2 random variables having the joint Distribution function (JPDF) $P(x,y) = \frac{(x+2y)}{27}$, where x & y can take values 0, 1 & 2. Then find the conditional probability of Y for X = x.

Sol

X\Y	0	1	2	P_{i*}
0	0	$\frac{2}{27}$	$\frac{4}{27}$	$P(X=0) = \frac{2}{9}$
1	$\frac{1}{27}$	$\frac{3}{27}$	$\frac{5}{27}$	$P(X=1) = \frac{1}{3}$
2	$\frac{2}{27}$	$\frac{4}{27}$	$\frac{6}{27}$	$P(X=2) = \frac{4}{9}$
	$P(Y=0) = \frac{1}{9}$	$P(Y=1) = \frac{1}{3}$	$P(Y=2) = \frac{5}{9}$	1

$\frac{6}{9} + \frac{2}{3}$

The marginal probability of X

$\Rightarrow$ $P(X=x_i) \Rightarrow$

X	0	1	2
$P(X=x)$	$\frac{6}{27}$	$\frac{9}{27}$	$\frac{12}{27}$

The Marginal of Y $\Rightarrow P(Y=y_j) \Rightarrow$

Y	0	1	2
$P(Y=y)$	$\frac{3}{27}$	$\frac{9}{27}$	$\frac{15}{27}$

The conditional distribution of Y for X = x

$$P(Y|X) = \frac{P(X=x_i, Y=y_j)}{P(X=x_i)} = \frac{P_{ij}}{P_{i*}}$$

X\Y	0	1	2
0	0	$\frac{2}{6}$	$\frac{4}{6}$
1	$\frac{1}{9}$	$\frac{3}{9}$	$\frac{5}{9}$
2	$\frac{2}{12}$	$\frac{4}{12}$	$\frac{6}{12}$

$\Rightarrow \frac{P_{00}}{P_{0*}} = \frac{0}{6/27} = 0$

$\Rightarrow \frac{P_{01}}{P_{0*}} = \frac{2/27}{6/27} = \frac{1}{3}$

$\Rightarrow \frac{P_{21}}{P_{2*}} = \frac{4/27}{12/27} = \frac{4}{12}$

Q-2- If the JPDF of R.V. X & Y is given by $F(x,y) = K(1-x-y)$ where $0 < x, y < \frac{1}{2}$ then find the value of K.

ASSIGNMENT

Q-1(a) The mean and standard deviation of a binomial distribution are

Solution →
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1 = \int_{x=0}^{1/2}\int_{y=0}^{1/2} K(1-x-y)\,dx\,dy = 1$$

$$= K\int_0^{1/2}\left[x - \frac{x^2}{2} - yx\right]_0^{1/2} dy = 1 \Rightarrow K\int_0^{1/2}\left(\frac{1}{2} - \frac{1}{8} - \frac{y}{2}\right) dy = 1$$

$$= K\int_0^{1/2}\left(\frac{3}{8} - \frac{y}{2}\right) dy = 1 = K\left[\frac{3y}{8} - \frac{y^2}{4}\right]_0^{1/2} = 1 \Rightarrow K = 8.$$

Q-3- Find the value of K of join PDF of Bivariate R.V. (x,y) is given by $F(x,y) = \begin{cases} K(1-x)(1-y) & \text{if } 0 < x < 4 \text{ \& } 1 < y < 5 \\ 0 & \text{; otherwise.} \end{cases}$ find K.

Solution →
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x,y)\,dx\,dy = 1 = \int_0^4\int_1^5 K(1-x)(1-y) = 1$$

$(1-x)(1-y)$
$1-x-y+xy$

$$K\int_0^4\int_1^5 1-x-y+xy = 1$$

$$K\int_0^4\left[y - xy - y^2 + xy^2\right]_1^5$$

$$K\int_0^4\left[5 - 5x - 25 + 25x - \cancel{1} + \cancel{x} + \cancel{1} - \cancel{x}\right]$$

$$K\int_0^4\left[20x - 20\right] = 1$$

$$K\left[20\frac{x^2}{2} - 20x\right]_0^4 = 1$$

$$20K\left(\frac{x^2}{2} - x\right)_0^4 = 1$$

$$20K\left[\frac{16}{2} - 4\right] = 1$$

$20K[4] = 1$

$80K = 1$

$\boxed{K = \frac{1}{80}}$

$K = \frac{1}{32}$

Q. If $f(x,y)=e^{-(x+y)}$; $x \geq 0, y \geq 0$ is joint PDF of X & Y then find $P(X<1)$

Solution- $f(x,y) = e^{-(x+y)}$

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

$f(x) = \int_0^{\infty} e^{-(x+y)}\,dy = e^{-x}\int_0^{\infty} e^{-y}\,dy = e^{-x}\left[\frac{e^{-y}}{-1}\right]_0^{\infty}$

$= -e^{-x}(e^{-\infty} - e^{-0}) = e^{-x}$ ⇒ Marginal distribution of x.

$P(X<1) = \int_{-\infty}^{1} f_x(x)\,dx$

$= \int_0^1 e^{-x}\,dx = \left[\frac{e^{-x}}{-1}\right]_0^1 = -(e^{-1} - e^{-0}) = 1 - \frac{1}{e} = 1 - \frac{1}{2.71}$

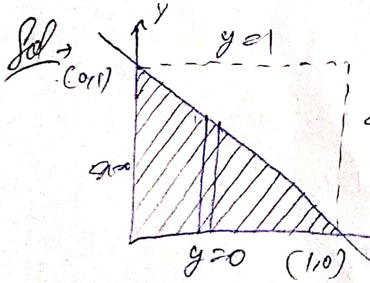
$= 0.6321$

Q2. Let X & Y are continuous random variable with joint PDF

$$f(x,y) = \begin{cases} 2xy + \frac{3y^2}{2} & : 0 \leq x < 1 ; 0 < y < 1 \\ 0 & : \text{otherwise} \end{cases}$$

then find $P(X+Y<1)$

Sol →



$= \int_{y=0}^{1}\int_{x=0}^{1-y} f(x,y)\,dx\,dy$

or

$= \int_{x=0}^{1}\int_{y=0}^{1-x} f(x,y)\,dy\,dx$

$= \int_0^1 \left[x^2 y + \frac{3y^2 x}{2}\right]_0^{1-y} dy$

$= \int_0^1 \left((1-y)^2 y + \frac{3y^2(1-y)}{2}\right) dy \quad (1-y^2-2y)$

$= \int_0^1 \left[2y - y^2 + \frac{3}{2}(y^2 - y^3)\right] dy = \left[\frac{2y^2}{2} - \frac{y^4}{4} + \frac{3}{2}\left(\frac{y^3}{3} - \frac{y^4}{4}\right)\right]_0^1$

$= \frac{1}{2} - \frac{1}{4} + \frac{3}{2}\left(\frac{1}{3} - \frac{1}{4}\right) = \frac{5}{24}$

Q-3- The joint PDF of R.V. X & Y is given by $f_{XY}(x,y)$: $0<x<2$; $0<y<2$ where C is constant, then find the value of C & $P(X<2)$ & $P(0<Y<2)$

Sol → ∵ $f_{XY}(x,y)$ is joint PDF then

$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$

$\int_0^2\int_{x=0}^{2} cxy\,dx\,dy = 1$

$\int_0^2 cy\left[\frac{x^2}{2}\right]_0^2 dy = 1$

$\int_0^2 cy\,[4/2]\,dy = 1$

$2C\int_0^2 y\,dy = 1$

$2C\left[\frac{y^2}{2}\right]_0^2 = 1$

$2C\left(\frac{4}{2}\right) = 1$

$\boxed{C = \frac{1}{4}}$

~~(i) $P(X<2) = \frac{1}{4}\int_0^2 xy = \left[\frac{xy^2}{8}\right]_0^2 = \frac{4x}{8} = \frac{x}{2}$~~

(i) $P(X<2) = \int_{-\infty}^{2} f_x(x)\,dx = \int_0^2 f_x(x)\,dx$

$f_x(x) = \int_0^2 f_{XY}(x,y)\,dy = \int_0^2 \frac{xy}{4}\,dy$

$= \frac{x}{4}\left[\frac{y^2}{2}\right]_0^2 = \frac{x}{4}\cdot\frac{4}{2} = \boxed{\frac{x}{2}}$

This is marginal

$f_X(x) = \frac{x}{2}$

$P(X<2) = \int_0^2 f_x(x)\,dx = \int_0^2 \frac{x}{2}\,dx = \left[\frac{x^2}{4}\right]_0^2$

$= \frac{1}{4}(4) = 1$

(ii) $P(0<Y<2) = \int_0^2 f_Y(y)\,dy$

$f_Y(y) = \int_0^2 f_{X,Y}(x,y)\cdot dx = \int_0^2 \frac{xy}{4}\cdot dx$

$= \frac{y}{4}\left[\frac{x^2}{2}\right]_0^2 = \frac{y}{2}$ $\quad f_Y(y) = \frac{y}{2}$

$P(0<Y<2) = \int_0^2 f_Y(y)\,dy = \int_0^2 \frac{y}{2}\,dy \to 1$

$\left[\frac{y^2}{4}\right]_0^2 = \frac{4}{4} = 1$

Q) Let X & Y are continous random variable with join PDF

$f(x,y) = \frac{3}{2}(x^2+y^2) : 0<x<1 : 0<y<1$

find $f(X|y)$

Sol $f(x|y) = \frac{f(x,y)}{f(y)}$ $\to$ Joint PDF, $\to$ Marginal of Y

Marginal distribution for Y

$f(y) = \int_0^1 f(x,y)\,dx = \int_0^1 \frac{3}{2}(x^2+y^2)\,dx$

$f(y) = \frac{3}{2}\left[\frac{x^3}{3} + y^2 x\right]_0^1 = \frac{3}{2}\left(\frac{1}{3}+y^2\right)$

$f(y) = \frac{3}{2}\left(\frac{1}{3}+y^2\right)$

$f(x|y) = \frac{\frac{3}{2}(x^2+y^2)}{\frac{3}{2}\left(\frac{1}{3}+y^2\right)} = \frac{x^2+y^2}{\frac{1}{3}+y^2}$

$f(y|x) = \frac{f(y,x)}{f(x)}$

Marginal distribution of x

$f(x) = \int_0^1 \frac{3}{2}(x^2+y^2)\,dy = \frac{3}{2}\left(\frac{1}{3}+x^2\right)$

$f(Y|X) = \frac{x^2+y^2}{\frac{1}{3}+x^2}$

Q-4 If 2 R.V. X & Y have PDF $f(x,y) = ke^{-(2x+y)}$ for $x,y>0$ then find K.

Solution $\int_0^\infty\int_0^\infty ke^{-(2x+y)}\,dx\,dy$

$k\left[\int_0^\infty\left[\frac{e^{-(2x+y)}}{-2}\right]_0^\infty dy\right] = k\left[\int_0^\infty \frac{0-1}{2}\,dy\right] = \boxed{k=2}$

Expectation of 2-D Random Variable

$E(X) = \sum x_i P(X=x_i)$

$E(Y) = \sum y_i P(Y=y_i)$

Q-1 - The Joint Distribution of R.V. X & Y is given by the formulae,

$f(x,y) = \frac{x+y}{21}$ where $x = 1,2,3$; $y = 1,2$ then find the value of marginal distribution for X & y.

and also find the mean of x & y.

Solution

X\Y	1	2	
1	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{5}{21}$
2	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{7}{21}$
3	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{9}{21}$
	$\frac{9}{21}$	$\frac{12}{21}$	

Marginal Distribution of X ⇒

x_i	1	2	3
$P(X=x_i)$	$\frac{5}{21}$	$\frac{7}{21}$	$\frac{9}{21}$

Marginal Distribution of Y ⇒

y_i	1	2
$P(Y=y_i)$	$\frac{9}{21}$	$\frac{12}{21}$

$E(X) = x_i P(x_i)$

Mean of $X = E(X) = \sum x_i P(X=x_i)$

$E(X) = 1 \times \frac{5}{21} + 2 \times \frac{7}{21} + 3 \times \frac{9}{21} = \frac{46}{21}$

Mean of $Y = E(Y) = \sum y_i P(Y=y_i)$

$E(Y) = 1 \times \frac{9}{21} + 2 \times \frac{12}{21} = \frac{33}{21}$

Variance of 2-D Random Variable

$V(X) = E(X^2) - (E(X))^2$

Q-2 - If the joint PDF of 2-D R.V. is given by $P(x,y) = \frac{k(x+y)}{2}$, $0 \le x, y \le 1$ then find the marginal for x & y, Expectation of X & Y & Variance, mean.

Solution →

$x \backslash y$	0	1	
0	0	1/2	1/2
1	1/2	1	3/2
	1/2	3/2	4

→ NOT PDF

1/2 can't come. Not valid PDF

If $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$ then only our PDF is valid

$F(x) = K(x+y)$

$\int_0^1\int_0^1 k(x+y)\,dx\,dy = 1$

$k\int_0^1 \left[\frac{x^2}{2} + yx\right]_0^1 dy = 1 = \frac{1}{2} + y$

$k\int_0^1 \left[\frac{1}{2} + y\right]_0^1 = 1 = \left[\frac{x}{2} + yx\right]_0^1 = \left[\frac{1}{2} + y\right]$

$= k\left[\frac{y}{2} + y^2\right]_0^1 = 1 = \frac{k}{2} + k = 1 \Rightarrow \frac{3k}{2} = 1$

$2k = 1$ $\cancel{k = \frac{1}{2}}$ $\boxed{k = \frac{2}{3}}$

$F(x,y) = \frac{2}{3}(x+y)$

$x \backslash y$	0	1	
0	0	$\frac{2}{3}$	$\frac{2}{3}$
1	$\frac{2}{3}$	$\frac{4}{3}$	$\frac{6}{3}$
	$\frac{2}{3}$	$\frac{6}{3}$	

Wrong Question!

LOL!

Q 3 > Let X & Y are 2 R.V. Each having values $-1, 0, 1$ and having a following joint probability distribution. Then find the value of $E(X)$, $E(Y)$, $V(X)$, $V(Y)$.

$X \backslash Y$	-1	0	1	
-1	0	0.1	0.1	0.2
0	0.2	0.2	0.2	0.6
1	0	0.1	0.1	0.2
	0.2	0.4	0.4	1

Marginal of X

Y	-1	0	1
	0.2	0.6	0.2

X	-1	0	1
	0.2	0.4	0.4

$E(Y) = 0.2(-1) + 0 + 0.2$

$= 0$

$E(Y^2) =$

$E(X) = -0.2 + 0.4$

$= 0.2$

$E(X)^2 = 0.6$

$V(X) = E(X^2) - (E(X))^2$

$= 0.6 - 0.04 = 0.56$

$V(Y) = E(Y^2) - (E(X))^2$

$= 0.4 - 0 = 0.4$

Pair all element 1 by 1 by making 1 constant.

$E(XY) = \sum xy\, (P\, X = x, Y = y)$

$\cancel{\text{at}(1,1)} = (-1)(-1)(0) = 0 + (-1)(0)(0.2) + (-1)(1)(0)$

$\text{at}\cancel{(-1,0) = (-1)(0)(0.2)} = 0 + (1)(-1)\,0 + (1)(1)(0.1) = 0$

$E(XY) = 0$

Covariance $\Rightarrow \text{cov}(X,Y) = -E(X)\cdot E(Y) + E(XY)$

$= -0.2\times 0 + 0$

$= 0.$

Correlation

It is the degree of relationship b/w 2 variable.

Let X & Y are 2 R.V. Then correlation coffecient $\Rightarrow \rho_{x,y}$ or r_{xy}

$$r_{xy} = \rho_{xy} = \rho(x,y) = \frac{\text{cov}(X,Y)}{\sigma_x \cdot \sigma_y} = \frac{E(XY) - E(X)\cdot E(Y)}{\sqrt{V(X)\cdot V(Y)}}$$

Remarks $\Rightarrow$

(I) $\rho^2 \le 1 \Rightarrow -1 \le \rho \le 1$

(II) If X & Y are independent R.V. the $\text{cov}(X,Y) = 0$ but converse not true

Q-1 $\Rightarrow$ 2 R.V. X & Y having the following Joint PDF

$$F(x,y) = \begin{cases} \frac{xy}{96} & ; 0<x<4 \ \& \ 1<y<5 \\ 0 & ; \text{otherwise} \end{cases}$$

then find the value of $E(X)$, $E(Y)$, $E(XY)$, $E(2X+3Y)$, $V(X)$, $V(Y)$, $\text{cov}(X,Y)$, ρ

Sol $\Rightarrow$

Marginal of $X \Rightarrow f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_1^5 \frac{xy}{96}\,dy$

$$\cancel{f(x)} = \frac{x}{96}\left[\frac{y^2}{2}\right]_1^5 = \frac{x}{96}\left(\frac{5^2-1^2}{2}\right) = \frac{x}{8}$$

$$\boxed{f(x) = \frac{x}{8} \ ; \ 0<x<4}$$

Marginal of $Y \Rightarrow f(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^4 \frac{xy}{96}\,dx = \frac{y}{96}\left[\frac{x^2}{2}\right]_0^4 = \frac{y}{12}$

$$f(y) = \frac{y}{12} \ ; \ 1<y<5$$

(I) $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^4 \frac{x^2}{8}\,dx = \left[\frac{x^3}{24}\right]_0^4 = \frac{64}{24} = \frac{8}{3}$

(II) $E(Y) = \int_{-\infty}^{\infty} y f(y)\,dy = \int_1^5 \frac{y^2}{12} = \left[\frac{y^3}{3}\right]_1^5 \cdot \frac{1}{12} = \frac{31}{9}$ Ans.

(III) $E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\cdot f(x,y)\,dx\,dy = \int_1^5\int_0^4 xy\cdot\frac{xy}{96}\,dx\,dy$

$$= \frac{1}{96}\int_1^5\int_0^4 (x^2y^2)\,dx\,dy$$

$$= \frac{1}{96}\int_1^5\left[\frac{x^3y^2}{3}\,dy\right]_0^4 = \frac{y^2}{96\times 3}(4)^3 = \frac{64y^2}{288}$$

$$\frac{64}{288}\int_1^5 y^2 = \frac{64}{288}\left[\frac{y^3}{3}\right]_1^5 = \frac{64}{288}\left[\frac{125-1}{3}\right] = \frac{248}{27}$$

(IV) $E(2X+3Y) = 2E(X) + 3E(Y)$

$$= 2\times\frac{8}{3} + 3\times\frac{31}{9} = \frac{47}{3}$$

(V) $V(X) = E(X^2) - (E(X))^2$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^4 \frac{x^3}{8}\,dx = \frac{1}{8}\left[\frac{x^4}{4}\right]_0^4 = 8$$

(VI) $E(Y^2) = \int_{-\infty}^{\infty} y^2 f(y)\,dy = \int_1^5 \frac{y^3}{12}\,dy = 13.$

(v) $V(X) = E(X^2) - (E(X))^2$
$= 8 - \left(\frac{8}{3}\right)^2 = \frac{8}{9}$

(vi) $V(Y) = E(Y^2) - (E(Y))^2$
$= 13 - \left(\frac{31}{9}\right)^2 = \frac{92}{81}$

(vii) $Cov(X,Y) = E(XY) - E(X).E(Y)$
$= \frac{248}{27} - \frac{8}{3} \times \frac{31}{9} = 0$

(viii) $\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X) \times V(Y)}} = \frac{0}{\sqrt{\frac{8}{9} \times \frac{92}{81}}}$

∵ coff. of correlation = 0 thus X & Y are not correlated.

Regression

It is a mathematical measure of average relationship b/w 2 or more variables in form of original units of data.

Line of Regression :- Equation of regression line

(i) Regression line of y on x :-
$y - \bar{y} = \rho \frac{\sigma y}{\sigma x} (x - \bar{x})$

(ii) Regression line of x on y.
$x - \bar{x} = \rho \frac{\sigma x}{\sigma y} (y - \bar{y})$

Regression coefficient of x on y → $b_{xy} = \frac{\rho\, \sigma x}{\sigma y}$

(ii) of y on x : $b_{yx} = \frac{\rho\, \sigma y}{\sigma x}$

$\rho = \pm \sqrt{b_{xy} \cdot b_{yx}}$

Q-2 → Calculate the correlation coefficient lines of regression and angle b/w them for following data. (10 marks)

x	y	xy	x^2	y^2
4	2.5	10	16	6.25
7	6	42	49	36
5	4.5	22.5	25	20.25
6	5	30	36	25
8	4.5	36	64	20.25
5	2	10	25	4
6	3	18	36	9
6	4.5	27	36	20.25
4	3	12	16	9
9	5.5	49.5	81	30.25
$\Sigma x = 60$	$\Sigma y = 40.5$	257		

$\rho = \frac{cov(X,Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{E(XY) - E(X) \cdot E(Y)}{\sqrt{V(X) \cdot V(Y)}}$

$E(X) = \bar{x} = \frac{\Sigma x}{n}$

$\bar{x} = E(X) = \frac{60}{10} = 6$

$E(Y) = \bar{y} = \frac{\Sigma y}{n} = \frac{40.5}{10} = 4.05$

$E(XY) = \frac{\Sigma xy}{n} = \frac{257}{10} = 25.7$

$V(X) = E(X^2) - (E(X))^2$

$E(X^2) = \frac{384}{10} = 38.4$

$E(Y^2) = \frac{180.25}{10} = 18.025$

$V(X) = E(X^2) - (E(X))^2 = 38.4 - 6^2 \Rightarrow V(X) = 2.4$ Ans

$V(Y) = E(Y^2) - (E(Y))^2 = 18.02 - (4.05)^2 = 1.6225$ Ans

$\rho = \dfrac{E(XY) - E(X)\cdot E(Y)}{\sqrt{V(X)\cdot V(Y)}} = \dfrac{25.7 - 6\times 4.05}{\sqrt{2.4\times 1.6225}} = 0.709.$

Regression line of x on $y \Rightarrow x - \bar{x} = \rho \dfrac{\sigma_x}{\sigma_y}(y - \bar{y})$

$x - 6 = 0.709 \times \sqrt{\dfrac{2.4}{1.6225}}\,(y - 4.05)$

$\boxed{x = 2.505 + 0.86y}$

Regression line of y on $x \Rightarrow y - \bar{y} = \rho \dfrac{\sigma_y}{\sigma_x}(x - \bar{x})$

$y - 4.05 = 0.709\left(\sqrt{\dfrac{1.6}{2.4}}\right)(x - 6)$

$\boxed{y = 0.58x + 0.55}$

Angle b/w lines

The angle b/w regression lines

$\tan\theta = \dfrac{1-\rho^2}{\rho}\;\dfrac{\sigma_x\cdot\sigma_y}{\sigma_x^2 + \sigma_y^2} = \dfrac{1-(0.709)^2}{0.709}\left(\dfrac{\sqrt{2.4}\times\sqrt{1.6225}}{(2.4)^2 + (1.6225)^2}\right)$

$= 0.325 \Rightarrow \theta = 5.67\times 10^{-3}$

Q-3 → If $\bar{x} = 970$, $\bar{y} = 18$, $\sigma_x = 38$, $\sigma_y = 2$, $\rho = 0.6$ then find the regression line of x on y.

Sol → $x - \bar{x} = \rho \dfrac{\sigma_x}{\sigma_y}(y - \bar{y}) = x - 970 = \cancel{0.6}^{0.3} \times \dfrac{38}{\cancel{2}}(y - 18)$

$\Rightarrow x \cdot \cancel{970}\; 11.4y + 764.83$

Q-4 → Let x denotes ~~age~~ a sour husband. Y denotes age of there using the equation.

x : Age of husband
y : ... wife

$x - y = 6$

$0.64x + 4.08 = 0$

Then using following regression lines find the value of mean of x & y.

(ii) correlation coefficient b/w x & y and ~~coff. and y and regression coff. ... of~~ σ_y if $\sigma_x = 4$.

$\boxed{\begin{array}{c}\rho^2 \le 1 \\ -1 \le \rho \le 1\end{array}}$

Sol → $\bar{x} - \bar{y} = 6$

$0.64\bar{x} + 4.08 = 0$

$\bar{x} = \dfrac{-4.08}{0.64} = -6.39$

$\bar{y} = -12.37$

$\rho = \sqrt{b_{xy} + b_{yx}}$

$b_{xy} = \cancel{coff}\cdot\rho\dfrac{\sigma_x}{\sigma_y} =$ coeffecient of y. $= 0$

$b_{yx} = \rho\dfrac{\sigma_y}{\sigma_x} =$ coffecient of x. $= 1$

(-1)

(0.69) -0.64 ?

not possible

ρ must be +ve

$\rho = \sqrt{-0.64}$

$\rho^2 = -0.64$ X

$\cancel{\sigma_y}\;\sigma_y = 0$

$\rho\dfrac{\sigma_y}{\sigma_x} = 1$

$\rho = \sqrt{b_{xy}\cdot b_{yx}} = 0$

Q-5, 2 regression lines $4x-5y+33=0$ & $20x-9y=107$.
and $V(X)=25$

(i) ~~Find~~ Mean of x & y

(ii) value of ρ & σ_x

(iii) Angle b/w regression lines.

for mean - $x=\bar{x}$, $y=\bar{y}$

$\rho=\sqrt{b_{xy}\cdot b_{yx}}$

$\rho^2 = +ve$ & ≤ 1

UNIT-IV/V

Sampling Theory ⇒

Population ⇒ A population includes all the elements from ~~the~~ a set of data.

Sample ⇒ A sample consist 1 or more observations ~~to~~ drawn from a population.
→ subset of Population

Parameter ⇒ It is measurable characteristics of population is called Parameter.
(Mean, S.D.)

Statistic ⇒ Measurable characteristic of sample is called Statistic.

* Formula for S.D. of a population is different ~~for~~ from formula of S.D. of Sample.

Sampling Distribution

If we draw sample of size n from a given finite population of size N, then the total no. of possible outcomes is

$$N_{C_n} = \frac{N!}{n!(N-n)!} = K \text{ (let)}$$

For each of these 'K' samples we can form some statistic $t = t(x_1, x_2 \ldots x_n)$
as diff. value of K form n.

Standard Error → The standard deviation of a sampling distribution of a statistic is known as standard error.

$$S.E(t) = \sqrt{Var(t)} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(t_i - \bar{t})^2}$$

Standard Error for some Statistic ⇒

S.N.	Statistic	Standard Error
1	Sample Mean ($\bar{x}$)	$\frac{\sigma}{\sqrt{n}}$
2	Sample ~~population~~ Proportion (p)	$\sqrt{\frac{PQ}{n}}$
3	Sample Standard Deviation (s)	$\sqrt{\frac{\sigma^2}{2n}}$
4	Sample variance (s^2)	$\sigma^2\sqrt{\frac{2}{n}}$
5	Sample Correlation coefficient (r)	$\frac{1-\rho^2}{\sqrt{n}}$ (ρ correlation coefficient)
6	Different of 2 independent sample ~~mean~~ Mean ($\bar{x}_1 - \bar{x}_2$)	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
7	Difference of 2 independent Sample SD ($s_1 - s_2$)	$\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$

Correlation coff. regression line Angle b/w – Test wednesday

Q-1 ⇒ Calculate the correlation coefficient tho following data:

x	y	xy	x^2	y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49
$\Sigma x = 30$	$\Sigma y = 40$	$\Sigma = 214$	$\Sigma = 220$	$\Sigma = 340$

$E(X) = \frac{\Sigma x_i}{n} = \frac{30}{5} = 6$

$E(Y) = \frac{\Sigma y}{n} = \frac{40}{5} = 8$

$E(XY) = \frac{\Sigma xy}{n} = \frac{214}{5} = 42.8$

$V(X) = E(X^2) - (E(X))^2 = 8$

$V(Y) = 4$

$$\rho = \frac{Cov(X,Y)}{\sqrt{V(X)\cdot V(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)\cdot V(Y)}} = \frac{42.8 - 48}{\sqrt{32}} = \frac{-5.2}{\sqrt{32}}$$

$\rho = -0.92$ (approx)

Regression line of x on y ⇒

$\bar{x} = 6$, $\bar{y} = 8$, $\rho = -0.92$, $\sigma_x = \sqrt{8}$, $\sigma_y = \sqrt{4}$

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

$$x - 6 = -0.92 \cdot \frac{\sqrt{8}}{\sqrt{4}}(y - 8)$$

$x - 6 = -1.3y + 10.4$

$x + 1.3y = 16.4$ — (I)

Regression line of y on x ⇒

$$(y - 8) = -0.92\sqrt{\frac{8}{4}}(x - 6) =$$

$y - 8 = -1.3x + 7.8$

$1.3x + y = 15.8$ — (II)

$x = 6$
$y = 8$

Statistical Hypothesis

A statistical Hypothesis is some assumption or statements which may or may not be true ~~is~~. ~~It may be simple hypothesis~~

Simple / Composite.

Simple → If the hypothesis completely specifies the population, then it is known as simple Hypothesis.

$H: \mu = \mu_0$ and $\sigma^2 = \sigma_0^2$ → Simple Hypothesis

Composite → A hypothesis which is not simple. Eg. $H: \mu = \mu_0, \sigma^2 > \sigma_0^2$

$H: \mu = \mu_0$ (Nothing about σ^2) $H: \mu > \mu_0, \sigma^2 = \sigma_0^2$

Test of Significance →

Difference b/w 2 independent statistics is known as Test of Significance.

Null Hypothesis → A statistical hypothesis which is under test ~~what~~ is usually a hypothesis of no difference & called null hypothesis. In other words, null hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true.

Usually null hypothesis is expressed as an equality

$H_0 : \theta = \theta_0$ $\quad H_1 : \theta > \theta_0$

Mean (μ)
Proportion (P)
Variance (σ^2)

~~Not Hypothesis~~
Not null

Alternative Hypothesis

A hypothesis which is complementary to the null hypothesis is called Alternative Hypothesis.

$H_1 : \theta > \theta_0$ → Right tailed Alternative
$H_1 : \theta < \theta_0$ → Left " "
$H_1 : \theta \neq \theta_0$

~~Errors of Hypo~~ Testing of Errors

A decision to accept or reject a null hypothesis H_0 after inspecting / taking only 1 sample from it which involves risk of taking wrong decision then the following 4 possible mutually exclusive and exhaustive events of decision are:

1. Reject H_0 when actually it is not true ⇒ H_0 is false.
2. Accept H_0 when it is true.
3. Reject ~~H_0~~ H_0 when it is true.
4. Accept H_0 when it is false. In which

Decisions ① & ② are correct and ③ & ④ are wrong decision.

		Decision from Sample	
		Reject H_0	Accept H_0
True Statements	H_0 (True	Wrong (Type I error)	correct
	H_0 : False	correct	Wrong (Type II error)

Type I error → The error of rejecting H_0 when H_0 is true is called Type I error.

Type II error → Error of accepting H_0 when H_0 is false.

P [Reject H_0 when it is true] = P [Type I occurs] $= \alpha$

⇒ Prob. of Type I error.

⇒ Probability of type II error

P [Accept H_0 when it is wrong] = P [Type II error]

Power of test ⇒ $P = 1 - \beta$

Level of Significance ⇒ The maximum size of type I error is known as level of significance.

$\alpha = P$ [Reject H_0 when it is true.

5% → level of significance
1%

Generally 5% & 1% level of Significance are taken for easier calculation.

If we had taken 5% level of Significance it implies that in 5 samples out of 100 we are likely to reject the correct H_0

OR

We are 95% confident that our decision to reject H_0 is correct.

Critical Region / Rejection Region

~~Let~~ The statistics which lead to the rejection of H_0 gives us a region called rejection Region.

While those which lead to the acceptance ~~to~~ of H_0 called ~~accept~~ accepted region.

Their intersection is always ϕ

$\Rightarrow$ If $t \in \underline{C}$: H_0 is rejected

$t \in \underline{A}$: H_0 is accepted

$C \cap A = \phi$

$C \cup A = S$ (Sample Space)

Q:- In order to test whether a coin is perfect it is tossed 5 times, the null hypothesis of perfectness is rejected iff more than 4 heads are obtained. Then find the value of

(1) Critical region (2) Probability of type 1 error (3) Probability of Type II error.

~~Also~~ ~~if~~ when the corresponding probability of getting head is 0.2.

Sol Let X be the No. of heads obtained in 5 tosses of a coin

$\because$ coin is perfect then probability of coming head $\boxed{H_0 : p = \frac{1}{2}}$ Null Hypothesis

under $H_0 \sim B(5, \frac{1}{2})$

$P(X = x | H_0) = {}^nC_x \, p^x q^{n-x} = {}^5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x} = {}^5C_x \times \frac{1}{32}$

$X = x$ under H_0

$P(X = x | H_0) = \frac{{}^5C_x}{32}$; $x = 0, 1, 2, \dots, 5$

(i) Critical Region $\rightarrow$ Reject H_0 / Null Hypothesis is Rejected if more than 4 Heads are obtained.

where Null Hypothesis is rejected

$\{X > 4\} = \{X = 5\}$

(ii) Probability of Type 1 error $\Rightarrow$ $\alpha = P[\text{Reject } H_0 | H_0]$

$= P[X = 5 | H_0] = \frac{{}^5C_5}{32} = 0.03125$

(3) Probability of Type II error $\Rightarrow$ $\beta = P[\text{accept } H_0 | H_1] = 1 - P[\text{reject } H_0 | H_1]$

accept H_0 such that H_1 is true

$= 1 - P[X = 5 | H_1]$

$= 1 - \frac{\cancel{5C_5}\,1}{32} \times 0.2$

$= 0.9938$

Q-2. In order to test whether a coin is perfect it is tossed 5 times. The null hypothesis is accepted if at most 3 heads are obtained. Then find the power of test corresponding to the alternative hypothesis when probability of coming heads is 0.4.

Sol → Let X be the no. of heads in 5 tosses of coin.

$p = \frac{1}{2}$, Null Hypothesis $= H_0 : p = \frac{1}{2}$

alternative hypothesis $H_1 : p = 0.4$

Critical Region ⇒ $\{X > 3\}$

Power of Test $= 1 - \beta = P$ (reject H_0 whenever H_1 is true).

$= P(X>3) = P(X>3 \mid p=0.4]$

$= P(X=4 \mid p=0.4) + P(X=5 \mid p=0.4)$

$= P(X=4 \mid p=0.4) + P(X=5 \mid p=0.4)$

$= {}^5C_4 (0.4)^4 (0.6)^1 + {}^5C_5 (0.4)^5 (0.6)^0$

$= 0.0256 \times \frac{5}{4} \times 0.6 + \frac{1}{32}$

$= \frac{272}{3125}$

Q-3. A sample of 100 objects give a mean of 7.4 kg & standard deviation of 1.2 kg. Find 95% ~~of~~ confidence ~~interval~~ limits for the population mean.

Solution ⇒ $n = 100$, $\bar{x} = 7.4$, $s = 1.2$

Confidence limits ⇒ $\bar{x} \pm 1.96\, \underbrace{S.E(\bar{x})}_{\text{Standard error of mean} = \frac{\sigma}{\sqrt{n}}}$

$= 7.4 \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$

$= 7.4 \pm \frac{1.2}{\sqrt{100}} \times 1.96 = 7.4 \pm 0.12 \times 1.96$

$= 7.1698, 7.6352$

$7.1648 \le \mu \le 7.6352$

Q-4 A random sample for 700 units from a large consignment showed that 200 were damaged. Find ① 95% & 99% confidence limits for the proportion of damaged units in the consignment.

Sol → $n = 700$

p = Proportion of damaged units $= \frac{200}{700} = \frac{2}{7}$

$q = 1 - p = 1 - \frac{2}{7} = \frac{5}{7}$

$S.E(p) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{\frac{2}{7} \times \frac{5}{7}}{700}} = 0.017$

Standard Error for Proportion

Confidence limits for 95% ⇒ $p \pm 1.96 \sqrt{\frac{pq}{n}} = (0.253, 0.319)$ Ans

" " " 99% ⇒ $p \pm 2.58 \sqrt{\frac{pq}{n}}$

⇒ $p \pm 2.58 \sqrt{\frac{pq}{n}}$ ⇒ $\frac{2}{7} \pm 2.58 \times 0.017$

0.729

$=$

$0.285 \le \mu \le 0.729$ Ans

Large Sample Test

Steps → ① Set up Null Hypothesis H_0

② Set up Alternative Hypothesis H_1

③ Set level of Significance (α)

④ Calculate test statistic $Z = \frac{t - E(t)}{S.E.(t)}$ → Expectation / Estimation

Standard error.

⑤ Obtain the value of Z.

⑥ Conclude. (if $|Z| > 3$ then null hypothesis is rejected)

Critical value of Z ⇒	level of significance 1%	5%
2 tailed Test	$\lvert Z_{\alpha/2}\rvert = 2.58$	$\lvert Z_{\alpha/2}\rvert = 1.96$
Right tailed	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$
Left Tailed	$-Z_\alpha = -2.33$	$-Z_\alpha = -1.645$

Q-1. A die was thrown 9000 times out of which 3220 yield 3 or 4. Can the die be regarded as unbiased?

Sol. $n = 9000$

$p = \frac{3220}{9000} = 0.3578$

We will set the null hypothesis H_0 that die is unbiased.

P = getting population proportion success.

$P = [\text{getting } 3 \text{ or } 4] = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3} = 0.33$

$Q = 1 - P = 1 - \frac{1}{3} = \frac{2}{3}$

Alternative Hypothesis.

$H_1 : P \neq \frac{1}{3}$

$\neq$ →	Two tailed Test
$>$ →	Right tailed Test
$<$ →	Left Tailed Test

$Z = \frac{p - E(p)}{S.E(p)} = \frac{p - P}{\sqrt{PQ/n}}$

$= \frac{0.3578 - 0.333}{\sqrt{\frac{0.333 \times 0.667}{9000}}} = 4.94$

$Z = 4.94$

$|Z| = 4.94 > 0$

∵ Calculated value ($|Z|$) > 3 then null hypothesis is rejected.

⇒ die is biased.

Z cal	< Z tab →	Null Hypothesis is accepted
Z cal	> Z tab	✗

§ If $|Z| \leq 3$ and Z calculated > Z tabulated then null hypothesis is rejected.

§ If Z cal < Z tab then Null hypothesis is accepted.

Q-2. It is claimed that a random sample of 100 tires with a mean life of 15269 Km. is drawn from a population of tires which has a mean life of 15200 km & SD is 1248 km. Test the validity of claim at

ⓐ 5% level of significance ⓑ 1% level of significance.

Solution: $n = 100$ $\bar{x} = 15269$ km

Null Hypothesis ⇒ $H_0 : \mu = 15200$

$\sigma = 1248$

Alternative Hypothesis ⇒ $H_1 : \mu \neq 15269$

(2 tailed Test)

$Z = \frac{\bar{x} - \mu}{S.D(\bar{x})} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$= \frac{15269 - 15200}{1248/\sqrt{100}}$

$Z(a) = 0.5529$

$|Z| = 0.5529$

at 5% level of significance

$Z_{tab} = 1.96$

$Z_{cal} < Z_{tab}$ => Null Hypothesis is accepted.

Test Statistic for

Mean $\quad Z = \dfrac{\bar{x} - \mu}{S.D(\bar{x})} = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Proportion $\quad Z = \dfrac{p - q}{\sqrt{pQ/n}}$

Q-3 → A radio shop sells 200 radios per day with a SD 50 radios. After an advertisement the management will compute the average sale for next 25 days to see that increase has occurred or not. Then find the following:

(I) - Null / Alternative Hypothesis

(II) - Test the Hypothesis at 5% level of significance if $\bar{x} = 216$.

Sol → (I) Null Hypothesis => $H_0: \mu = 200$

Alternative Hypothesis $H_1: \mu > 200$ (Right tailed test)

(II) $\bar{x} = 216 \quad \sigma = 50 \quad n = 25$

$Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{216 - 200}{50/\sqrt{25}} = 1.6 = Z_{cal}$

$Z_{tab} = 1.645 \qquad$ 5, 6 => Right

$Z_{tab} > Z_{cal}$ => Null Hypothesis is accepted.

Chi - Square Test

$x^2 = \left(\dfrac{x - \mu}{\sigma}\right)^2$ → chi - square variate

Q-1 → The following figure shows the distribution of digits in numbers chosen at random from a telephone directory. Test whether the digits may be taken to occur equally frequently in the directory. The table value of X^2 for 9° of freedom at 5% level of significance is 16.92)

Chi - square test →

$(O-E)$	Expected freq.	Digit	Observed frequency	$(O-E)^2/E$
26	1000	0	1026	0.676
107	1000	1	1107	11.449
-3	1000	2	997	0.009
-34	1000	3	966	1.156
75	1000	4	1075	5.625
-67	1000	5	933	4.489
107	1000	6	1107	11.479
-28	1000	7	972	0.784
-36	1000	8	964	1.296
-147	1000	9	853	21.809

Sol → Let Null Hypothesis says that the digits 0, 1, 2 ...... 9 in the telephone directory are uniformly distributed.

∵ every digit is equally probable then Expected frequency of each digit will be $\dfrac{10000}{10} = 1000$

$$\boxed{X^2_{cal} = \sum \frac{(O-E)^2}{E} \; ; \; O = \text{Observed freq} \; ; \; E = \text{Expected freq}}$$

$X^2_{tab} = 16.92$

$X^2_{cal} = 58.542$

$X^2_{cal} > X^2_{tab}$

$\cancel{X^2_{cal} = \dfrac{\sum O}{1000}} = \boxed{O = X^2_{cal}}$

=> Null hypothesis is rejected at 5% level of significance. Hence digits are not uniformly distributed.

Q-2 A die is rolled 100 times with the following distribution. At the 1% level of significance determine whether die is true or not.

$\chi^2_{tab} = 15.086$

Number	Observed Frequency	Expected Freq (E)	$(O-E)$	$(O-E)^2$	$(O-E)^2/E$
1	17	16.67	0.33	0.1089	0.006
2	14	16.67	-2.67	7.1289	0.427
3	20	16.67	3.33	11.0889	0.665
4	17	16.67	0.33	0.1089	0.006
5	17	16.67	0.33	0.1089	0.006
6	15	16.67	-1.67	2.7889	0.167
	100				1.277

∵ No of observations are 6 then No. of degree of freedom will be

$N-1 = 6-1 = 5$

Null Hypothesis (H_0) : The die is true

Expected frequency $= \frac{\text{Total freq}}{\text{No. of observations}} = \frac{100}{6} = 16.67$

$\chi^2_{cal} = 1.2796$ (for value)

$\chi^2_{tab} = 15.086$

∵ $\chi^2_{tab} > \chi^2_{cal}$ ∴ Null Hypothesis is accepted

∴ Die is uniform.

2×2 Contigency Table

a	b	a+b
c	d	c+d
a+c	b+d	N= a+b+c+d

→ No of observation.

Q-3 - In a certain sample of 2000 families 1400 families are consumers of Tea. Out of 1800 Hindu families 1236 families consumer Tea. Then using 'χ^2 test' state whether there is any significant difference b/w consumption of Tea among Hindu & Non Hindu families.

Solution →

Number of families	Hindu	Non-Hindu	Total
Families consuming Tea	1236	164	1400
Not consuming Tea	564	36	600
	1800	200	2000

Let Null Hypothesis is that ~~conjunction~~ of Tea there is no significant difference b/w conjunction of Tea around Hindu & non-Hindu family / H_0

$E(1236) = \frac{1400 \times 1800}{2000} = 1260$

$E(564) = \frac{600 \times 1800}{2000}$

$E(36) = \frac{600 \times 200}{2000} = 60$

$E(164) = \frac{1400 \times 200}{2000} = 140$

$\chi^2_{cal} = \sum\left(\frac{(O-E)^2}{E}\right)$

$= 576\left(\frac{1}{1260} + \frac{1}{540} + \frac{1}{140} + \frac{1}{60}\right)$

$= 15.23$

$\chi^2_{tab} = 3.841$

O	E	$(O-E)^2$	$(O-E)^2/E$
1236	1260	576	0.45
564	540	576	0.000000018
164	140	576	0.000001
36	60	576	0.000014

∵ $\chi^2_{tab} < \chi^2_{cal}$ ∴ Null hypothesis is rejected. Hence there is significant difference b/w consumption of Tea among hindu and non-hindu families.

t-test

Q-4 10 cartons are taken at random from automatic filling machine. The mean weight of the 10 cartons is 11.8 KG. And standard deviation is 0.15 kg. Does the sample mean differ significantly from the weight of 12 KG.

(Given that $V = 9$ (degree of freedom) $t_{0.05} = 2.26$ ← tabulated value at 5% level of significance.

Sol → $t = \dfrac{\bar{x} - \mu}{\sqrt{s^2/(n-1)}}$ $n = 10$ $\bar{x} = 11.8$

$v = 9$ $s = 0.15$

$t_{0.05} = 2.26$

Null Hypothesis ⇒ $H_0 : \mu = 12$ Alternative Hypo ⇒ $H_1 : \mu \neq 12$.

$t = \dfrac{11.8 - 12}{\sqrt{(0.15)^2/9}}$

$t_{cal} = -4$ ⇒ $t_{cal} > t_{tab}$ ⇒ Null Hypothesis is rejected at 5% level of significance.

$|t_{cal}| = 4$

~~Hence the sample mean differ from~~

Hence there is significant difference b/w sample mean & population

Q.2 → A machine is designed to produce insulating washers for electrical devices of average thickness of 0.025 cm. A random sample of 10 washers was found to have an average thickness of 0.024 cm. with standard deviation of 0.002 cm. Test the significance of the deviation. (value of t for 9° of freedom at 5% level is 2.262).

Sol → $\bar{x} = 0.024$ cm d.f ⇒ $t_{0.05} = 2.262$

$s = 0.002$ $t_{cal} = -1.5$

Null Hypothesis ⇒ $H_0 : \mu = 0.025$. $|t_{cal}| = 1.5$

There is no significant difference b/w deviations $t_{tab} > t_{cal}$

$t = \dfrac{0.024 - 0.025}{\sqrt{(0.002)^2/9}}$ ∴ Null Hypo is accepted

⇒ there is no significant difference b/w deviations.

— X — X —