UNIT- I

CURVE FITTING

Least square Method:

Suppose (x_i, y_i) for $i=1$ to n be the given data.
Let $Y = f(x)$ be the required curve which passes through the given data. Let $Y_i = f(x_i)$ be the estimated data for $i=1$ to n.

$\boxed{d_i = y_i - Y_i}$, called 'residuals' or 'errors'

For the best fitting curve, sum of the squares of the residuals should be minimum.

$\sum_{i=1}^{n} d_i^2$ should be minimum.

This curve is called the best fitting curve by least square method.

Fitting of a straight line by Least square Method:

Let (x_i, y_i) for $i=1$ to n be the given data.
Let $\boxed{Y = a+bx}$ be a straight line which passes through the given data.
Let Y_i be the estimated data, then

$$d_i = y_i - Y_i,$$

for best fitting straight line,

$$S = \sum_{i=1}^{n} d_i^2 \text{, must be minimum.}$$

$$S = \sum_{i=1}^{n} [y_i - Y_i]^2$$

$$S = \sum_{i=1}^{n} [y_i - a - bx_i]^2$$

classmate

Date
Page

S will be minimum when

$$\frac{\partial S}{\partial a} = 0 \quad \& \quad \frac{\partial S}{\partial b} = 0$$

case 1: $\frac{\partial S}{\partial a} = 0$

$$\sum_{i=1}^{n} 2[y_i - a - bx_i][-1] = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - \sum_{i=1}^{n} bx_i = 0$$

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i$$

$$\sum y = na + b\sum x \quad — ①$$

case 2: $\frac{\partial S}{\partial b} = 0$

$$\sum_{i=1}^{n} 2[y_i - a - bx_i][-x_i] = 0$$

$$\sum_{i=1}^{n} x_iy_i - \sum_{i=1}^{n} ax_i - \sum_{i=1}^{n} bx_i^2 = 0$$

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

solving ① & ②, we get a & b which are unknown constants
The equations ① & ② are called "normal equations"
to fit a straight line by least square method.

Fitting of parabola by Least square method:

Let (x_i, y_i) for $i = 1$ to n be the given data.
Let $\boxed{Y = ax^2 + bx + c}$ be the parabola which passes through the given point, to find the best parabola, we use least square method.

classmate

Date

Page

Let Y_i be the estimated points and

$$\boxed{d_i = y_i - Y_i} \rightarrow \text{residuals.}$$

for LSM, sum of the squares of the residuals must be minimum i.e.,

$$S = \sum_{i=1}^{n} d_i^2 \text{, is minimum}$$

$$S = \sum_{i=1}^{n} [y_i - Y_i]^2$$

$$S = \sum_{i=1}^{n} [y_i - ax_i^2 - bx_i - c]^2$$

S will be minimum when,

$$\frac{\partial S}{\partial a} = 0 \quad , \quad \frac{\partial S}{\partial b} = 0 \quad , \quad \frac{\partial S}{\partial c} = 0$$

case 1: $\frac{\partial S}{\partial a} = 0$

$$\sum_{i=1}^{n} 2[y_i - ax_i^2 - bx_i - c][-x_i^2] = 0$$

$$\sum_{i=1}^{n} x_i^2 y_i - \sum_{i=1}^{n} a x_i^4 - \sum_{i=1}^{n} b x_i^3 - \sum_{i=1}^{n} c x_i^2 = 0$$

$$\Sigma x^2 y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2 \quad \text{— ①}$$

case 2: $\frac{\partial S}{\partial b} = 0$

$$\sum_{i=1}^{n} 2[y_i - ax_i^2 - bx_i - c][-x_i] = 0$$

$$\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} a x_i^3 - \sum_{i=1}^{n} b x_i^2 - \sum_{i=1}^{n} c x_i = 0$$

$$\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x \quad \text{— ②}$$

case 3: $\frac{\partial S}{\partial c} = 0$

$$\sum_{i=1}^{n} [y_i - ax_i^2 - bx_i - c][-1] = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a x_i^2 - \sum_{i=1}^{n} b x_i - \sum_{i=1}^{n} c = 0$$

$$\Sigma y = a\Sigma x^2 + b\Sigma x + nc \quad - ③$$

Solving ①, ②, ③ we obtain the unknown constants a, b, c

Equations ①, ②, ③ are called "normal equations" to fit a parabola by LSM.

PROBLEMS:

* Fit a straight line through the data by LSM

a) $(2,2), (5,4), (6,6), (9,9), (11,10)$

Let $y = ax + b$,

normal eqns are:

$$\Sigma y = a\Sigma x + nb$$

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

$n = 5$ (no. of data)

x	y	x^2	xy
2	2	4	4
5	4	25	20
6	6	36	36
9	9	81	81
11	10	121	110
$\Sigma x = 33$	$\Sigma y = 31$	$\Sigma x^2 = 267$	$\Sigma xy = 251$

$$33a + 5b = 31$$

$$267a + 33b = 251$$

$a = 0.943 \qquad b = 0.024$

* Find a straight line of the form $y = a + bx$ for the data

x	1	2	3	4	5	6	7	8
y	3	3	4	5	5	6	6	7

let $y = ax + b$

normal eqns are : $\Sigma y = a\Sigma x + nb$

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

$n = 8$

x	y	x^2	xy
1	3	1	3
2	3	4	6
3	4	9	12
4	5	16	20
5	5	25	25
6	6	36	36
7	6	49	42
8	7	64	56
$\Sigma x = 36$	$\Sigma y = 39$	$\Sigma x^2 = 204$	$\Sigma xy = 200$

$36a + 8b = 39$

$204a + 36b = 200$

$a = 0.5833 \quad b = 2.25$

$y = 0.5833 + 2.25x$

* Fit a parabola of the form $y = ax^2 + bx + c$ by SLSM

x	0.78	1.56	2.34	3.12	3.81
y	2.5	1.2	1.12	2.25	4.28

normal equations are: $\Sigma y = a\Sigma x^2 + b\Sigma x + nc$

$\boxed{n = 5}$ $\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x$

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

x	y	x^2	x^4	xy	x^2y	x^3
0.78	2.5	0.6084	0.37	1.95	1.52	0.474
1.56	1.2	2.4336	5.922	1.872	2.9196	3.796
2.34	1.12	5.4756	29.982	2.608	6.132	12.812
3.12	2.25	9.7344	94.758	7.02	21.902	30.371
3.81	4.28	14.5161	210.717	16.3068	62.128	55.306
$\Sigma x = 11.61$	$\Sigma y = 11.35$	$\Sigma x^2 = 32.768$	$\Sigma x^4 = 341.749$	$\Sigma xy = 29.768$	$\Sigma x^2y = 94.601$	$\Sigma x^3 = 102.75$

$32.768a + 11.61b + 5c = 11.35$

$102.75a + 32.768b + 11.61c = 29.768$

$341.749a + 102.75b + 32.768c = 94.601$

$a = 0.982$

$b = -3.921$

$c = 4.936$

$\boxed{y = 0.982x^2 + (-3.921)x + 4.936}$

* Fit a parabola of the form $y = ax^2 + bx + c$ for the data

x	0	1	2	3	4
y	-2.1	-0.4	2.1	3.6	9.9

x	y	x^2	x^3	x^4	xy	x^2y
0	-2.1	0	0	0	0	0
1	-0.4	1	1	1	-0.4	-0.4
2	2.1	4	8	16	4.8	8.4
3	3.6	9	27	81	10.8	32.4
4	9.9	16	64	256	39.6	158.4
$\Sigma x = 10$	$\Sigma y = 13.1$	$\Sigma x^2 = 30$	$\Sigma x^3 = 100$	$\Sigma x^4 = 354$	$\Sigma xy = 54.2$	$\Sigma x^2y = 198.8$

normal eqn: $\Sigma y = a\Sigma x^2 + b\Sigma x + nc$

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

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

$\boxed{n = 5}$

$30a + 10b + 5c = 13.1$

$100a + 30b + 10c = 54.2$

$354a + 100b + 30c = 198.8$

$a = 0.505$

$b = 0.457$

$c = -1.808$

$y = (0.505)x^2 + 0.457x + (-1.808)$

classmate
Date
Page

__Fitting of Geometric curves:__

Take a curve of the form $y = ax^b$ for the data

x	50	450	780	1200	4400	4800	5300
y	28	30	32	36	51	58	69

by LSM fit a curve.

$y = ax^b$

taking log on both sides.

$\log y = \log (ax^b)$

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

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

$Y = A + bX$

normal equations are.

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

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

x	y	$X = \log x$	$Y = \log y$	X^2	XY
50	28	3.912	3.332	15.303	13.034
450	30	6.109	3.401	37.319	20.776
780	32	6.659	3.465	44.342	23.073
1200	36	7.09	3.583	50.268	25.403
4400	51	8.389	3.931	70.375	32.977
4800	58	8.476	4.060	71.842	34.412
5300	69	8.575	4.234	73.53	36.306
		49.21	26.066	362.979	185.981

$\boxed{n = 7}$

$7A + 49.21b = 26.006$

$49.21A + 362.979b = 185.981$

$A = \log a$

$a = e^A = 11.14$

$A = 2.411 \qquad b = 0.185$

$Y = 2.411 + (0.185)X$

$y = ax^b \Rightarrow y = 11.14x^{0.185}$

* The pressure and volume of a gas related by the equation $PV^{\gamma} = K$, where γ & K are constants, fit this eqn to the following set of observations

P	0.5	1	1.5	2	2.5	3
V	1.62	1	0.75	0.62	0.52	0.46

$PV^{\gamma} = K$

$\log P + \gamma \log V = \log K$

$\log P = -\gamma \log V + \log K$

$Y = -\gamma X + B$

$Y = AX + B$

normal of eqn : $\Sigma Y = A \Sigma X + nb$

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

P	V	X	Y	X^2	XY
0.5	1.62	0.482	-0.693	0.232	-0.334
1	1	0	0	0	0
1.5	0.75	-0.287	0.405	0.082	-0.116
2	0.62	-0.478	0.693	0.228	-0.331
2.5	0.52	-0.653	0.916	0.427	-0.598
3	0.46	-0.776	1.098	0.602	-0.852
		-1.712	2.419	1.571	2.231

$-1.712A + 6B = 2.419$

$1.571A + (-1.712)B = -2.331$

$A = -1.515$

$A = -\gamma = 1.515$

$B = -0.029$

$B = \log K$; $K = e^{B} = 0.971$

$\boxed{PV^{1.515} = 0.971}$

classmate

Date ____

Page ____

* Fit an exponential curve $y = ae^{bx}$ for the data.

x	2	4	6	8
y	25	38	56	84

normal eqns are: $\sum Y = b\sum x + nA$

$\sum xY = b\sum x^2 + A\sum x$

$\boxed{n = 4}$

$y = ae^{bx}$

$\log y = bx + \log a$

$Y = bx + A$

x	y	$Y = \log y$	x^2	xY
2	25	3.218	4	6.436
4	38	3.637	16	14.548
6	56	4.025	36	24.15
8	84	4.430	64	35.44
$\sum x = 20$	$\sum y = 203$	$\sum Y = 15.31$	$\sum x^2 = 120$	$\sum xY = 80.574$

$20b + 4A = 15.31$

$120b + 20A = 80.574$

$b = 0.2012$

$A = 2.8215$

$a = e^A = 16.802$

$\boxed{y = 16.802e^{0.2012x}}$

* Find a curve of the form $y = ax^b$ for the data

x	2.3	3.1	4	4.92	5.91	7.2
y	33	39.1	50.3	67.2	85.6	125

$y = ax^b$

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

$Y = A + bX$

normal equations are:

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

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

x	y	$X = \log x$	$Y = \log y$	X^2	XY	x^2
2.3	33	0.832	3.496	0.692	2.908	5.29
3.1	39.1	1.131	3.666	1.279	4.146	9.61
4	50.3	1.386	3.918	1.92	5.430	16
4.92	67.2	1.593	4.207	2.537	6.701	24.206
5.91	85.6	1.776	4.449	3.154	7.901	34.928
7.2	125	1.974	4.828	3.896	9.530	51.84
$\Sigma x = 27.43$	$\Sigma y = 400.2$	$\Sigma X = 8.692$	$\Sigma Y = 24.564$	$\Sigma X^2 = 13.478$	$\Sigma XY = 36.616$	$\Sigma x^2 = 141.314$

$\boxed{n=6}$

$$6A + 8.692b = 24.564$$

$$8.692A + 13.478 = 36.616$$

$$A = 2.408, \quad b = 1.163$$

$$a = e^A = 11.111$$

$$\boxed{y = 11.111\,(x)^{1.163}}$$

* For the above same data, try for $y = ab^x$

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

$$Y = A + xB$$

normal equations are:

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

$$\Sigma xY = A\Sigma x + B\Sigma x^2$$

$$\Sigma xY = 116.828$$

$$6A + 27.43B = 24.564$$

$$27.43A + 141.314B = 116.828$$

$$A = 2.792 \qquad B = 0.284$$

$$a = e^A = 16.313 \qquad b = 1.309$$

$$y = (16.313)(1.309)^x$$

* $y = a + \frac{b}{x}$

x	2	4.5	6	7.5
y	1.52	1.8	1.98	2.5

$$\frac{1}{x} = X$$

$$y = a + \frac{b}{x}$$

$$y = a + bX$$

normal equations are:

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

$$\Sigma Xy = a\Sigma X + b\Sigma X^2$$

x	y	$X = \frac{1}{x}$	X^2	Xy
2	1.52	0.5	0.25	0.76
4.5	1.8	0.222	0.049	0.399
6	1.98	0.167	0.027	0.33
7.5	2.5	0.133	0.017	0.332
	$\Sigma y = 7.8$	$\Sigma X = 1.022$	$\Sigma X^2 = 0.343$	$\Sigma Xy = 1.821$

$$4a + 1.022b = 7.8$$

$$1.022a + 0.343b = 1.821$$

$a = 2.482$ $\quad$ $b = -2.087$

$$\boxed{y = 2.482 + \frac{(-2.087)}{x}}$$

Correlation and Regression

Let x and y be two independent variables, the co-variation between the 2 independent variables 'x' and 'y' is called 'correlation'.

* If both x and y together increases or together decreases, then we say that there is a "positive correlation" between variables x and y.
* If 'x' increases and 'y' decreases or vice versa, then we say that there is a "negative correlation" between variables 'x' and 'y'.
* The measure of correlation is usually calculated with the correlation coefficient 'r', defined as

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

where

n = number of data points
$\bar{x}, \bar{y}$ = Mean of x and mean of y
σ_x, σ_y = standard deviation in x, standard deviation in y

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

NOTE: $\boxed{-1 \leq r \leq 1}$

If $0 < r < 1$
then 'x' & 'y' are positively correlated.
If $-1 < r < 0$
then 'x' and 'y' are negatively correlated.
If $r = \pm 1$
then x and y are perfectly correlated
If $r = 0$
then x and y are not correlated, they are uncorrelated variables.

The coefficient of correlation of two variables x & y is

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{n \sigma_x \sigma_y} \quad \text{— ①}$$

take $x-\bar{x} = X$ $\quad y-\bar{y} = Y$

$$\sigma_x = \sqrt{\frac{\sum (x-\bar{x})^2}{n}} = \sqrt{\frac{\sum X^2}{n}}$$

$$\sigma_y = \sqrt{\frac{\sum (y-\bar{y})^2}{n}} = \sqrt{\frac{\sum Y^2}{n}}$$

Now, equation ① becomes,

$$r = \frac{\sum XY}{n\sqrt{\frac{\sum X^2}{n}}\sqrt{\frac{\sum Y^2}{n}}}$$

$$\boxed{r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}}$$

Regression: The method of finding a mathematical relation between two variables x and y is called regression.
If the goal to find a relation of the form $y = f(x)$ is called regression of 'y' on 'x'.
If to find a relation of the form $x = f(y)$ is called regression of 'x' on 'y'.

Formula 1: The equation of the regression line of 'y' on 'x' is

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

where $b_{yx} = \frac{r \cdot \sigma_y}{\sigma_x}$ is called the coefficient of regression of y on x

classmate

Date
Page

FORMULA 2 The equation of the regression of 'x' on 'y' is

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

where b_{xy} is called the coefficient of regression of x on y. $b_{xy} = r.\frac{\sigma_x}{\sigma_y}$

THEOREM:

Show that the coefficient of correlation 'r' is a geometric mean of the coefficients of regression.

Proof: We know that the coefficient of correlation $= r$ and, if 'a' & 'b' are any 2 numbers, then the GM between a & b is $\sqrt{ab}$

We have the coefficients of regression,

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

Now, consider

$$b_{xy} \times b_{yx} = \frac{r\,\cancel{\sigma_x}}{\cancel{\sigma_y}} \times r\,\frac{\cancel{\sigma_y}}{\cancel{\sigma_x}}$$

$$b_{xy} \times b_{yx} = r^2$$

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

NOTE: ① If both b_{xy} & b_{yx} are positive, r is positive

② If both b_{xy} & b_{yx} are negative, r is negative.

$r \to$ Coefficient of correlation

* Obtain the coefficient of correlation & hence find the equations of regression lines for the following data.

x	1	2	3	4	5	6	7	8
y	10	12	16	28	25	36	41	49

x	y	$X = x-\bar{x}$	$Y = y-\bar{y}$	X^2	Y^2	XY
1	10	-3.5	-17.125	12.25	293.265	59.937
2	12	-2.5	-15.125	6.25	228.765	37.812
3	16	-1.5	-11.125	2.25	123.765	16.687
4	28	-0.5	0.875	0.25	0.765	-0.437
5	25	0.5	-2.125	0.25	4.515	-1.062
6	36	1.5	8.875	2.25	78.765	13.312
7	41	2.5	13.875	6.25	192.515	34.687
8	49	3.5	21.875	12.25	478.515	76.562
$\Sigma x = 36$	$\Sigma y = 217$			$\Sigma X^2 = 42$	$\Sigma Y^2 = 1400.9$	$\Sigma XY = 237.498$

$\bar{x} = \frac{\Sigma x}{8} = 4.5$

$\bar{y} = \frac{\Sigma y}{8} = 27.125$

The coefficent of correlation is

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}} = \frac{237.498}{\sqrt{42}\sqrt{1400.9}} = 0.979$$

$$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} \qquad b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$$

$$\sigma_x = \sqrt{\frac{\Sigma X^2}{n}} = 2.291 \qquad \sigma_y = \sqrt{\frac{\Sigma Y^2}{n}} = 13.233$$

The equation of regression line of y on x

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

$$b_{yx} = \frac{0.979 \times 13.233}{2.291} = 5.654$$

$$y - 27.125 = 5.654(x - 4.5)$$

$$\boxed{5.654x - y + 1.682 = 0}$$

The equation of regression line of x on y

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

$$b_{xy} = \frac{0.979 \times 2.291}{13.233} = 0.169$$

$$x - 4.5 = 0.169(y - 27.125)$$

$$\boxed{x - 0.169y + 0.0841 = 0}$$

Verify:

$$r = \sqrt{b_{xy} \times b_{yx}}$$

$$= \sqrt{5.654 \times 0.169}$$

$$\boxed{r = 0.977}$$

Verified.

classmate

Date
Page

NOTE ① The point $(\bar{x}, \bar{y})$ passes through the lines of regression

$$b_{xy} = x - \bar{x} = r \frac{\sigma x}{\sigma y} (y - \bar{y}) \Rightarrow x \text{ on } y$$

$$b_{yx} = y - \bar{y} = r \frac{\sigma y}{\sigma x} (x - \bar{x}) \Rightarrow y \text{ on } x$$

②

$$r \frac{\sigma x}{\sigma y} \times r \frac{\sigma y}{\sigma x} = r^2 = \text{product of regression coefficients}$$

$$r = \pm\sqrt{\text{product of regression coefficients}}$$

If we know regression coefficients, we can find the correlation coefficient. (r)

* If $x = 0.44y + 0.4$
$y = 1.763x + 2.68$, Find the mean of x, mean of y and the correlation coefficent.

Since the point $(\bar{x}, \bar{y})$ passes through the lines of regression, we have

$$\bar{x} = 0.44\bar{y} + 0.4$$
$$\bar{y} = 1.73\bar{x} + 2.68$$

$$\bar{x} = 0.44(1.763\bar{x}) + 0.44(2.68) + 0.4$$
$$0.2\cancel{38}\bar{x} = 1.5792$$
$$\boxed{\bar{x} = 7.043}$$

$$\boxed{\bar{y} = 14.8658}$$

$$r = \pm\sqrt{0.44 \times 1.763}$$
$$= + 0.8807$$

* $y = 4.686x + 4.927$
$x = 0.197y - 0.548$, find the mean of x & y & the correlation coefficient

$\bar{x} = 0.197\bar{y} - 0.548$
$\bar{y} = 4.686\bar{x} + 4.927$

$\bar{x} = 5.4986$
$\bar{y} = 30.6939$

$r = \sqrt{4.686 \times 0.197}$
$= 0.9608$

Multiple Regression

The multiple regression examines how a number of multiple independent variables are related to one dependent variable. If x_1, x_2 are two independent variables & 'y' be a dependent variable, then we can form a regression plane through the given 'n' data points in the form

$y = ax_1 + bx_2 + c$ where, a, b & c are unknown constants to be determined.

To determine constants a, b, c, we use the following equations.

$$\Sigma y = a\Sigma x_1 + b\Sigma x_2 + nc$$
$$\Sigma x_1 y = a\Sigma x_1^2 + b\Sigma x_1 x_2 + c\Sigma x_1$$
$$\Sigma x_2 y = a\Sigma x_1 x_2 + b\Sigma x_2^2 + c\Sigma x_2$$

* Find 'y' when $x_1 = 10$, $x_2 = 6$ from the least square regression equation of 'y' on x_1, x_2 for the following data.

y	90	72	54	42	30	12
x_1	3	5	6	8	12	14
x_2	16	10	7	4	3	2

$\boxed{n = 6}$

$$y = ax_1 + bx_2 + c$$

$$\Sigma y = a\Sigma x_1 + b\Sigma x_2 + nc$$

$$\Sigma x_1 y = a\Sigma x_1^2 + b\Sigma x_1 x_2 + c\Sigma x_1$$

$$\Sigma x_2 y = a\Sigma x_1 x_2 + b\Sigma x_2^2 + c\Sigma x_2$$

y	x_1	x_2	x_1^2	x_2^2	x_1x_2	x_1y	x_2y
90	3	16	9	256	48	270	1440
72	5	10	25	100	50	360	720
54	6	7	36	49	42	324	378
42	8	4	64	16	32	336	168
30	12	3	144	9	36	360	90
12	14	2	196	4	28	168	24
$\Sigma y = 300$	$\Sigma x_1 = 48$	$\Sigma x_2 = 42$	$\Sigma x_1^2 = 474$	$\Sigma x_2^2 = 434$	$\Sigma x_1x_2 = 236$	$\Sigma x_1y = 1818$	$\Sigma x_2y = 2820$

$$300 = 48a + 42b + 6c$$

$$1818 = 474a + 236b + 48c$$

$$2820 = 236a + 434b + 42c$$

$$a = -5.6461$$

$$b = 2.5385$$

$$c = 84.400$$

$$y = -5.6461x_1 + 2.5385b + 84.400$$

classmate

Date
Page

Fit/

* Make a regression plane to estimate $\beta_0, \beta_1, \beta_2$ to the following data of a transport company on the weights of six shipments with the distances they have moved & the damage of goods that was incured.
Estimate the damage when a shipment of 3700 kg is moved to 260kms.

Weight x_1 (1000kg)	4.0	3.0	1.6	1.2	3.4	4.8
Distance x_2 (100km)	1.5	2.2	1.0	2.0	0.8	1.6
Damage y (RS)	160	112	69	90	123	186

$$y = \beta_0 x_1 + \beta_1 x_1 + \beta_2 \qquad \boxed{n=6}$$

$$\Sigma y = \beta_0 \Sigma x_1 + \beta_1 \Sigma x_2 + n\beta_2$$
$$\Sigma x_1 y = \beta_0 \Sigma x_1^2 + \beta_1 \Sigma x_1 x_2 + \beta_2 \Sigma x_1$$
$$\Sigma x_2 y = \beta_0 \Sigma x_1 x_2 + \beta_1 \Sigma x_2^2 + \beta_2 \Sigma x_2$$

x_1	x_2	y	x_1^2	x_2^2	x_1x_2	x_1y	x_2y
4	1.5	160	16	2.25	6	640	240
3	2.2	112	9	4.84	6.6	336	246.4
1.6	1.0	69	2.56	1	1.6	110.4	69
1.2	2	90	1.44	4	2.4	108	180
3.4	0.8	123	11.56	0.64	2.72	418.2	98.4
4.8	1.6	186	23.04	2.56	7.68	892.8	297.6
$\Sigma x_1 = 18$	$\Sigma x_2 = 9.1$	$\Sigma y = 740$	$\Sigma x_1^2 = 63.6$	$\Sigma x_2^2 = 15.29$	$\Sigma x_1x_2 = 27$	$\Sigma x_1 y = 2505.4$	$\Sigma x_2 y = 1131.4$

$$740 = 18\beta_0 + 9.1\beta_1 + 6\beta_2$$
$$2505.4 = 63.6\beta_0 + 27\beta_1 + 18\beta_2$$
$$1131.4 = 27\beta_0 + 15.29\beta_1 + 9.1\beta_2$$

$$\beta_0 = 30.1091$$
$$\beta_1 = 12.1608$$
$$\beta_2 = 14.5617$$

$$y = 30.1091 x_1 + 12.1608 x_2 + 14.45617$$

at $x_1 = 3.7$, $x_2 = 2.6$ $\Rightarrow y = 157.583$ ₹

Random variables

Random variables:
In a random experiment, if a real variable is associated with every outcome then it is called a random variable.
In other words, it is a function that assigns a real number to every sample point in the sample space of a random experiment.

They are usually denoted by X, Y, Z etc

Discrete Random Variables:

If X takes some finite or countably infite no. of values then we say that, `X` is a discrete random variable

continuous Random variables:

If `X` takes infinite no. of values of a random experiment, then it is called continuous Random variable.

NOTE: whenever there is a counting process, involved, we talk of discrete Random variable,
whereas when a measuring process is involved, we talk of continuous random variable.

Discrete Probability Distribution:

If for each value x_i of X of a discrete random variable, we assign a real value $P(x_i)$ such that,

i) $P(x_i) \geq 0$
ii) $\sum_i P(x_i) = 1$

Here P(X) is called the "probability - Mass function" or "probability - Density Function".

cummulative Distributive Function:

$f(x)$ is a cummulative distributive function, if

$$f(x) = P(X=x_i) = \sum_{i=1}^{n} P(x_i)$$

Mean, variance of discrete probability Distributions:

Let 'X' be a discrete random variable, then the

i) Mean of X is usually denoted by E(X) which is called "Expected value of X" & is defined as

$$\boxed{E(X) = \sum x . P(x)}$$

ii) The variance of X is denoted by 'Var(X)' and is defined as

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

where $\boxed{E(X^2) = \sum x^2 P(x)}$

* The probability distribution of a random variable 'x' is given by the following table

X	0	1	2	3	4	5
P(X=x)	K	5K	10K	10K	5K	K

i) Find k

ii) Find mean & variance

iii) Find $P(X \geq 3)$, $P(X < 2)$

i) $\sum_{i=1}^{6} P(x_i) = 1$ & $P(x_i) \geq 0$

$K + 5K + 10K + 10K + 5K + K = 1$

$32K = 1$

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

ii) Mean = $E(X)$

$E(X) = \sum x\, P(x)$

$= (0 \times K) + 1(5K) + 2(10K) + 3(10K) + 4(5K) + 5(K)$

$= 80K$

$= 80 \times \frac{1}{32} = \frac{10}{4} = \frac{5}{2} = 2.5$

$\boxed{E(X) = 2.5}$

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

$E(X^2) = \sum x^2\, P(x)$

$0 + 1^2(5K) + 4(10K) + 9(10K) + 16(5K) + 25(K)$

$= 240K$

$= \cancel{240}^{30} \times \frac{1}{\cancel{32}_4} = \frac{30}{4} = \frac{15}{2} = 7.5$

$Var(X) = 7.5 - (2.5)^2$

$= 7.5 - 6.25$

$= 1.25$

$\boxed{Var(X) = 1.25}$

iii) $P(X \geq 3)$

$P(X=3) + P(X=4) + P(X=5)$

$= 10K + 5K + K$

$= 16K = \frac{16}{32} = \frac{1}{2} = 0.5$

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

$= K + 5K$

$= 6K$

$= \frac{6}{32} = \frac{3}{16}$

* The probability distribution of a finite random variable 'x' is given by the following table. Find the value of k, mean & variance.

x	-2	-1	0	1	2	3
$P(x_i)$	0.1	K	0.2	2K	0.3	k.

$\sum_{i=1}^{6} P(x_i) = 1$

$0.1 + K + 0.2 + 2K + 0.3 + K = 1$

$0.6 + 4K = 1$

$4K = 0.4$

$\boxed{K = 0.1}$

Mean $E(X)$

$E(X) = \sum x\, P(x)$

$= -0.2 - k + 0 + 2K + 0.6 + 3K$

$= 0.4 + 4K$

$= 0.4 + 0.4$

$= 0.8$

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

$E(X^2) = \sum x^2\, P(x)$

$= \cancel{0 + 2^2(0K) + 4(10K) + 9(10K)}$

$= 4(0.1) + K + 0 + 2K + 4(0.3) + 9(k)$

$= 0.4 + K + 2K + 1.2 + 9k$

$= 12K + 1.6$

$= 1.2 + 1.6$

$= 2.8$

$var(X) = 2.8 - (0.8)^2 = 2.8 - 0.64 = 2.16$

* A random variable 'x' has the following probability function

x	1	2	3	4	5	6	7
$P(x_i)$	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

i) Find k, iv) $P(1 \le X < 5)$
ii) $P(x \ge 6)$. v) $E(X)$
iii) $P(X<6)$ vi) $var(x)$.

i) $k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$

$10k^2 + 9k = 1$

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

$10k^2 + 10k - k - 1 = 0$

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

$k = -1, \frac{1}{10}$

but $k \ne -1$ because $P(x_i) \ge 0$

so, $\boxed{k = 0.1}$

ii) $P(X \ge 6) = P(6) + P(7)$

$= 2k^2 + 7k^2 + k$

$= 9k^2 + k$

$\frac{9}{100} + \frac{1}{10} = \frac{19}{100} = 0.19.$

iii) $P(X<6) = P(1) + P(2) + P(3) + P(4) + P(5)$

$= k + 2k + 2k + 3k + k^2$

$= k^2 + 8k$

$\frac{1}{100} + \frac{8}{10} = \frac{81}{100} = 0.81$

iv) $P(1 \le X < 5) = P(1) + P(2) + P(3) + P(4).$

$= k + 2k + 2k + 3k$

$= 8k = 0.8$

v) $E(X) = \sum x P(x)$

$= k + 4k + 6k + 12k + 5k^2 + 12k^2 + 49k^2 + 7k$

$= 66k^2 + 30k$

$= \frac{66}{100} + \frac{30}{10} = \frac{366}{100} = 3.66.$

vi) $var(x) = E(x^2) - [E(x)]^2$

$E(x^2) = \sum x^2 p(x)$

$= k + 8k + 18k + 48k + 25k^2 + 72k^2 + 343k^2 + 49k.$

$= 437k^2 + 124k$

$= \frac{437}{100} + \frac{1240}{100} = \frac{1677}{100} = 16.77.$

$var(x) = 16.77 - (3.66)^2 = 3.3744$

* Find k, $E(x)$, $E(x^2)$, σ^2 for the probability function $P(x)$ defined by the following table.

x	1	2	3	-------	n
$P(x)$	k	$2k$	$3k$	-------	nk.

i) $k + 2k + 3k + \cdots + nk = 1$

$\frac{n(n+1)k}{2} = 1$

$k = \frac{2}{n^2+n}$

ii) $E(x) = \sum x P(x)$

$k + 4k + 9k + \cdots + n^2 k$

$= k \frac{(n)(2n+1)(n+1)}{6}$

$= \frac{2}{\cancel{n}\cancel{(n+1)}} \cdot \frac{\cancel{(n)}(2n+1)\cancel{(n+1)}}{6} = \frac{2n+1}{3}$

iii) $E(x^2) = \sum x^2 P(x)$

$= k + 8k + 27k + \cdots + n^3 k$

$= k \frac{(n)^2(n+1)^2}{4} = \frac{2}{\cancel{n}\cancel{(n+1)}} \times \frac{n^2(n+1)^2}{4} = \frac{n(n+1)}{2}$

classmate

Date

Page

iv) σ^2 = variance $= E(X^2) - (E(X))^2$

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

$$= \frac{n(n+1)}{2} - \frac{(4n^2+1+4n)}{9}$$

$$= \frac{9n^2+9n-8n^2-2-8n}{18}$$

$$= \frac{n^2+n-2}{18}$$

continuous probability Distributions

Let 'x' is a continuous random variable, defined in the interval $(-\infty, \infty)$, then we assign a real value $f(x)$ for each 'x' of X in the same interval defined as follows:

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

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

here, $f(x)$ is called "probability density function" i.e., "pdf" of the random variable 'x'.

* probability that $X \in [a, b]$ is defined as

$$P(a \leq x \leq b) = P(a < x \leq b) = P(a \leq x < b)$$
$$= P(a < x < b) = \int_a^b f(x)\,dx.$$

* Area under the curve $f(x)$ with x-axis = 1 i.e., $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

cummulative distribution function (cdf):

The cdf of a continuous random variable 'x' is defined as

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

Mean, variance of continuous probability function

Mean:

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

variance:

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

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx.$$

* If a continuous random variable 'X' has a pdf,

$$f(x) = \begin{cases} \frac{1}{4} & , -2 < x < 2 \\ 0 & , \text{else where} \end{cases}$$

, obtain

i) $P(X < 1)$

ii) $P(|x| > 1)$

iii) $P(2x + 3 > 5)$

i) $P(X < 1) = P(-2) + P(-1) + P(0)$

$$= \int_{-2}^{1} \frac{1}{4}\, dx = \left(\frac{x}{4}\right)_{-2}^{1} = \frac{1}{4} + \frac{2}{4} = \frac{3}{4}$$

ii) $P(|x| > 1) = 1 - P(|x| \leq 1)$

$$= 1 - P(-1 \leq x \leq 1)$$

$$= 1 - \int_{-1}^{1} \frac{1}{4}\, dx = 1 - \frac{2}{4} = \frac{1}{2}$$

iii) $P(2x + 3 > 5) = P(x > 1) = 1 - P(X < 1)$

$$= \int_{1}^{2} \frac{1}{4}\, dx = \left(\frac{x}{4}\right)_{1}^{2} = \frac{1}{4}$$

* A continuous random variable 'x' that can assume any value between $2 \le x \le 5$ has a density function.
$f(x) = K(1+x)$ find the value of k,
find $P(x<4)$

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

$$\int_{2}^{5} K(1+x)\,dx = 1$$

$$\left[Kx + \frac{Kx^2}{2}\right]_2^5 = 1$$

$$5K + \frac{25K}{2} - 2K - \frac{4K}{2} = 1$$

$$3K + \frac{21K}{2} = 1$$

$$27K = 2$$

$$\boxed{K = \frac{2}{27}}$$

$$P(x<4) = \int_{2}^{4} K(1+x)\,dx = \frac{2}{27}\left(x + \frac{x^2}{2}\right)_2^4 = \frac{2}{27}\left(4 + \frac{16}{2} - 2 - \frac{4}{2}\right)$$

$$= \frac{2}{27}(8) = \frac{16}{27}$$

* A continuous random variable has a pdf $f(x) = 3x^2$ for $0 \le x \le 1$, find a & b such that
i) $P(x \le a) = P(x > a)$
ii) $P(x > b) = 0.05$

i) $$\int_{0}^{a} 3x^2\,dx = \int_{a}^{1} 3x^2\,dx$$

$$\Rightarrow \left(\frac{3x^3}{3}\right)_0^a = \left(\frac{3x^3}{3}\right)_a^1$$

$$\Rightarrow a^3 = 1 - a^3$$

$$2a^3 = 1$$

$$a^3 = \frac{1}{2} \Rightarrow a = \frac{1}{\sqrt[3]{2}} = 0.7937$$

classmate

Date
Page

ii) $$\int_b^1 3x^2 dx = 0.05$$

$$(x^3)_b^1 = 0.05$$

$$1 - b^3 = 0.05$$

$$b^3 = 0.95$$

$$b = 0.9830$$

* A continuous random variable 'x' has pdf, $f(x) = kx^2e^{-x}$ for $x \geq 0$, find k, mean & variance.

$$\int_{-\infty}^{\infty} kx^2 e^{-x} dx = 1 \Rightarrow \text{condition}$$

$$\Rightarrow \int_0^{\infty} kx^2 e^{-x} dx = 1$$

Bernoulli's part of integration $\Rightarrow k\left[x^2(-e^{-x}) - 2x(e^{-x}) + 2(-e^{-x})\right]_0^{\infty} = 1$

$$\Rightarrow k\left[-x^2e^{-x} - 2xe^{-x} - 2e^{-x}\right]_0^{\infty} = 1$$

$$\Rightarrow k\left[-0-0-0-(-0-0-2)\right] = 1$$

$$k[2] = 1$$

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

Mean $\Rightarrow \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_0^{\infty} x \cdot kx^2 e^{-x} dx$

$$= \int_0^{\infty} kx^3 e^{-x} dx$$

$$= k\left[x^3(-e^{-x}) - 3x^2(e^{-x}) + 6x(-e^{-x}) - 6(e^{-x})\right]_0^{\infty}$$

$$= k\left[0-0+0-0-(0-0+0-6)\right]$$

$$= \frac{6}{2} = 3$$

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

$$E(X^2) = \int_0^{\infty} kx^4 e^{-x} dx = k(24) = 12$$

$$\text{var}(x) \Rightarrow 12 - (3)^2 = 12 - 9 = 3$$

classmate

Date
Page

* If the pdf of a continuous random variable 'x' is given by

$$f(x) = \begin{cases} ax & ; 0 \le x \le 1 \\ a & ; 1 \le x \le 2 \\ 3a - ax & ; 2 \le x \le 3 \\ 0 & ; \text{elsewhere} \end{cases}$$

Find i) value of 'a'

ii) The cummulative distributive function (cdf)

i)

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

$$\int_0^1 ax\,dx + \int_1^2 a\,dx + \int_2^3 (3a - ax)\,dx + 0 = 1$$

$$\left(\frac{ax^2}{2}\right)_0^1 + (ax)_1^2 + \left(3ax - \frac{ax^2}{2}\right)_2^3 = 1$$

$$\frac{a}{2} + 2a - a + 9a - \frac{9a}{2} - 6a + \frac{4a}{2} = 1$$

$$\frac{3a}{2} - \frac{9a}{2} + 3a + 2a = 1$$

$$-\frac{6a}{2} + 5a = 1$$

$$2a = 1$$

$$\boxed{a = \frac{1}{2}}$$

ii) CDF $\Rightarrow \int_{-\infty}^{x} f(x)\, dx$

at $a = \frac{1}{2} \Rightarrow f(x) = \begin{cases} x/2 & , 0 \le x \le 1 \\ 1/2 & ; 1 \le x \le 2 \\ 3/2 - \frac{x}{2} & ; 2 \le x \le 3 \\ 0 & ; \text{elsewhere} \end{cases}$

For $\Rightarrow \int_{-\infty}^{0} 0\, dx = 0 \quad — ①$

classmate

Date
Page

For $\Rightarrow \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{0} \cancel{f(x)}^{0}\,dx + \int_{0}^{x} f(x)\,dx$

$$CDF = 0 + \int_{0}^{x} \frac{x}{2}\,dx = \left(\frac{x^2}{4}\right)_0^x = \frac{x^2}{4}$$

For $\Rightarrow \int_{-\infty}^{2} f(x)\,dx = \int_{-\infty}^{0} \cancel{f(x)}^{0}\,dx + \int_{0}^{1} f(x).dx + \int_{1}^{x} f(x)\,dx.$

$$= 0 + \left(\frac{x^2}{4}\right)_0^1 + \left(\frac{x}{2}\right)_1^x$$

$$= \frac{1}{4} + \frac{x}{2} - \frac{1}{2} = \frac{x}{2} - \frac{1}{4}$$

$$= \frac{2x-1}{4}$$

For $\Rightarrow \int_{-\infty}^{3} f(x)\,dx = \int_{-\infty}^{0} \cancel{f(x)}^{0}\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{x} f(x)\,dx$

$$= 0 + \left(\frac{x^2}{4}\right)_0^1 + \left(\frac{x}{2}\right)_1^2 + \left(\frac{3x}{2} - \frac{x^2}{4}\right)_2^x$$

$$= \frac{1}{4} + 1 - \frac{1}{2} + \frac{3x}{2} - \frac{x^2}{4} - 3 + 1$$

$$= \frac{3x}{2} - \frac{x^2}{4} - \frac{1}{4} - 1$$

$$= \frac{6x - x^2 - 5}{4}$$

For $\Rightarrow \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} \cancel{f(x)}^{0}\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{3} f(x)\,dx$

$$+ \int_{3}^{\infty} \cancel{f(x)}^{0}\,dx$$

$$= 0 + \frac{1}{4} + 1 - \frac{1}{2} + \frac{9}{2} - \frac{9}{4} - 3 + 1 + 0$$

$$= \frac{1}{4} + \frac{1}{2} + \frac{9}{2} - \frac{9}{4} - 3 + 1$$

$$= 1$$

classmate

Date
Page

$$cdf = \begin{cases} 0 & ; -\infty \le x \le 0 \\ x^2/4 & ; 0 \le x \le 1 \\ x/2 - \frac{1}{4} & ; 1 \le x \le 2 \\ \frac{3x}{2} - \frac{x^2}{4} - \frac{5}{4} & ; 2 \le x \le 3 \\ 1 & ; x > 3 \end{cases}$$

* The Pdf of a random variable X is given by

$$f(x) = \begin{cases} x & , 0 \le x \le 1 \\ 2-x & , 1 < x \le 2 \\ 0 & , \text{otherwise} \end{cases}$$

i) Find cdf of $f(x)$

ii) $P(X > 1.5)$

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

$\Rightarrow$

* For $\int_{-\infty}^{0} f(x) \cancel{dx}^{0} \Rightarrow 0$

* For $\int_{-\infty}^{1} f(x)\,dx \Rightarrow$ $\int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{0} \cancel{f(x)}^{0} dx + \int_{0}^{x} f(x)\,dx$.

$$= \int_{0}^{x} f(x)\,dx$$

$$= \int_{0}^{x} x\,dx = \frac{x^2}{2}$$

* For $\int_{-\infty}^{2} f(x)\,dx \Rightarrow \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{0} \cancel{f(x)}^{0}\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{x} f(x)\,dx$

$$\Rightarrow \int_{0}^{1} x\,dx + \int_{1}^{x} (2-x)\,dx$$

$$= \left(\frac{x^2}{2}\right)_0^1 + \left(2x - \frac{x^2}{2}\right)_1^x = \frac{1}{2} + 2x - \frac{x^2}{2} - 2 + \frac{1}{2}$$

$$= 2x - \frac{x^2}{2} - 1$$

* For $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{\infty} f(x)\,dx$

$$= \left(\frac{x^2}{2}\right)_0^1 + \left(\frac{2x - x^2}{2}\right)_1^2 + 0$$

$$= \frac{1}{2} + 4 - \frac{4}{2} - 2 + \frac{1}{2}$$

$$= 5 - 2 - 2 = 1.$$

$$\text{cdf} = \begin{cases} 0 \; ; & -\infty \le x \le 0 \\ \frac{x^2}{2} \; ; & 0 \le x \le 1 \\ 2x - \frac{x^2}{2} - 1 \; ; & 1 \le x \le 2 \\ 1 \; ; & x \ge 2 \end{cases}$$

ii) $P(X > 1.5)$

$$\int_{1.5}^{2} f(x)\,dx = \int_{1.5}^{2} (2-x)\,dx$$

$$= \left(2x - \frac{x^2}{2}\right)_{1.5}^{2}$$

$$= 4 - \frac{4}{2} - 3 + \frac{2.25}{2}$$

$$= 2 - 3 + \frac{2.25}{2}$$

$$= 1.125 - 1$$

$$= 0.125$$

Binomial Distribution

This distribution is a discrete probability distribution. we apply this distribution in the following cases:

i) There should be only two outcomes of a trial
i.e., (True, False), (Success, Failure), (Yes, No)

ii) The no. of trials is finite

iii) All trials are independent

Then the pdf of the binomial distribution is defined as

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

here, n = no. of trials
p = probability of success
q = probability of Failure

i.e., $\boxed{q = 1-p}$

* The pdf $P(X=r)$ represents the probability of r^{th} success out of n trials

Mean, variance of Binomial distribution

Mean:

we have

$$E(X) = \sum r\, P(r)$$

$$E(X) = \sum_{r=0}^{n} r \, {}^nC_r \, p^r q^{n-r}$$

$$= \sum_{r=0}^{n} r \cdot \frac{n!}{r!(n-r)!} p^r q^{n-r}$$

$$= \sum_{r=1}^{n} \cancel{r} . \frac{n(n-1)!}{\cancel{r}(r-1)!\,(n-r+1-1)!} p^{r-1} . p . q^{n-r-1+1}$$

$$= \sum_{r=1}^{n} \frac{n\,(n-1)!}{(r-1)!\,((n-1)-(r-1))!} p^{(r-1)} . p . q^{(n-1)-(r-1)}$$

$$= np . (p+q)^{n-1}$$

but $p+q=1$

so, $\boxed{np = E(X)}$

<u>Variance:</u>

$$var(X) = \sum r^2 p(r) - \left(\sum r p(r)\right)^2$$

$$E(X^2) = \sum_{r=0}^{n} r^2 p(r)$$

$$= \sum_{r=0}^{n} r^2 . {}^nC_r\, p^r q^{n-r}$$

$$= \sum_{r=0}^{n} r(r+1-1) \frac{n!}{r!\,(n-r)!} p^r q^{n-r}$$

$$= \sum_{r=0}^{n} r(r-1) \frac{n!}{r!\,(n-r)!} p^r q^{n-r} + \sum_{r=0}^{n} r \frac{n!}{r!\,(n-r)!} p^r q^{n-r}$$

$$= \sum_{r=0}^{n} r(r-1) \frac{n!}{r!\,(n-r)!} p^r q^{n-r} + np$$

$$= \sum_{r=2}^{n} \cancel{r(r-1)} \frac{n(n-1)(n-2)!}{\cancel{r(r-1)}(r-2)!\,(n-2-r+2)!} p^{r-2} . p^2 . q^{n-2-r+2} + np$$

$$= n(n-1)p^2 \sum_{r=2} \frac{(n-2)!}{(r-2)!\,((n-2)-(r-2))!} p^{r-2} q^{(n-2)-(r-2)} + np$$

$$= n(n-1)p^2 \; (p+q)^{n-2} + np$$

$= n(n-1)p^2 (1)^{n-2} + np$

$= n(n-1)p^2 + np$

$= np(np - p + 1)$

$= np(np + q)$

$\Rightarrow Var(x) = np(np+q) - (np)^2$

$= \cancel{(np)^2} + npq - \cancel{(np)^2}$

$$\boxed{Var(x) = npq}$$

* The mean & variance of binomial distribution are 16 & 8. Find the distribution of x.

$np = 16$

$npq = 8$

$16 \cdot q = 8$

$q = \frac{8}{16} = \frac{1}{2}$

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

$1 - p = q$

$1 - q = p$

$$\boxed{p = \frac{1}{2}}$$

$n \times p = 16$

$\frac{n}{2} = 16$

$$\boxed{n = 32}$$

$P(X = r) = {}^{n}C_{r} \left(\frac{1}{2}\right)^{r} \left(\frac{1}{2}\right)^{n-r} = {}^{32}C_{r} \left(\frac{1}{2}\right)^{r} \left(\frac{1}{2}\right)^{32-r}$

$P(X) = {}^{32}C_{r} \cdot \left(\frac{1}{2}\right)^{32}$

* The no. of telephone lines busy at an instant of time is a binomial variate with probability 0.2. If at an instant 10 lines are chosen at random, exactly, what is the probability that i) 5 lines are busy exactly ii) atmost 2 lines are busy

$p = 0.2 \qquad n = 10$
$q = 0.8$

i) $P(X=r) = {}^{10}C_r (0.2)^r (0.8)^{10-r}$

put $r = 5$

$= {}^{10}C_5 (0.2)^5 (0.8)^5$

$= \frac{\cancel{10}^2 \times \cancel{9}^3 \times \cancel{8} \times 7 \times 6}{\cancel{8} \times 4 \times 3 \times \cancel{2} \times 1} (0.2)^5 (0.8)^5$

$= 252 \times 4^5 (0.2)^{10}$

$= 0.0264$

ii) $P(X \leq 2) \Rightarrow {}^{10}C_0 (0.2)^0 (0.8)^{10} + {}^{10}C_1 (0.2)^1 (0.8)^9 + {}^{10}C_2 (0.2)^2 (0.8)^8$

$= 0.1073 + 0.2684 + 0.3019$

$= 0.6776$

* The probability that a man aged 60 will live up to 70 is 0.65. what is the probability that out of 10 men, now aged 60, at least 7 will live up to 70.

$p = 0.65$
$q = 1-p = 0.35$
$n = 10$

$P(X=r) = {}^nC_r p^r q^{n-r}$

$= {}^{10}C_r (0.65)^r (0.35)^{10-r}$

P(atleast upto 70 years) $= P(X \geq 7)$

$P(7) + P(8) + P(9) + P(10)$

$= {}^{10}C_7 (0.65)^7 (0.35)^3 + {}^{10}C_8 (0.65)^8 (0.35)^2 + {}^{10}C_9 (0.65)^9 (0.35)$

$+ {}^{10}C_{10} (0.65)^{10}$

$= 0.2522 + 0.1756 + 0.0724 + 0.0134$

$= 0.5136$

* Determine the probability of getting 9 exactly twice in 3 throws with a pair of fair dice

Sample space $\Rightarrow \{(1,1),(1,2),(1,3),\ldots\ldots,\ldots(6,5),(6,6)\} = 36$

$P(\text{getting } 9) \Rightarrow \{(4,5),(5,4),(3,6),(6,3)\} \Rightarrow \frac{4}{36} = \frac{1}{9}$

$p = \frac{1}{9}$

$q = 1 - p = \frac{8}{9}$

$n = 3$

$r = 2$

$P(X=2) = {}^3C_2 \left(\frac{1}{9}\right)^2 \left(\frac{8}{9}\right)^1$

$= \frac{\cancel{3}\times\cancel{2}}{\cancel{2}\times 1} \times \frac{1}{81} \times \frac{8}{\cancel{9}3}$

$P(X=2) = \frac{8}{243} = 0.0329$

* In 100 sets of 10 tosses of an unbiased coin, in how many cases should we expect (i) 7 heads & 3 tails
(ii) atleast 7 heads

$p = \frac{1}{2}$ (getting a head)

$q = \frac{1}{2}$ (getting a tail)

$n = 10$ (tosses in one set)

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

$$P(X=r) = {}^{10}C_r \left(\frac{1}{2}\right)^r \left(\frac{1}{2}\right)^{10-r} = {}^{10}C_r \left(\frac{1}{2}\right)^{10}$$

i) P(getting 7 heads & 3 tails)

$$P(X=7) = {}^{10}C_7 \left(\frac{1}{2}\right)^{10}$$

$P(X=7) = 0.1171 \Rightarrow$ for one set

for 100 sets $\Rightarrow P(X=7) = 11.71 \; (100 \times 0.1171)$

ii) P(atleast 7 heads)

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

$$\left({}^{10}C_7 + {}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10}\right) \times \frac{1}{2^{10}}$$

$= 0.1718$ for 1 set

for 100 sets $\Rightarrow P(X \geq 7) = 17.18$

Poisson Distribution:

This is also a discrete probability distribution. we apply this distribution in the following cases:

i) when probality of success is very small
ii) when no. of trials is very large.

* The pdf of poisson distribution is defined as

$$P(X=r) = \frac{e^{-m} m^r}{r!}$$

$P(X=r)$ represents the probability of r^{th} success.

Mean, variance of Poisson distribution:

Mean:

$$E(X) = \sum_{r=0}^{\infty} r \, P(r)$$

$$= \sum_{r=0}^{\infty} r \cdot \frac{e^{-m} \cdot m^r}{r!}$$

classmate

Date ______
Page ______

$$= e^{-m}\left[m + \frac{2m^2}{2!} + \frac{3m^3}{3!} + \frac{4m^4}{4!} + \cdots\right]$$

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

$$= m \cdot e^{-m} \cdot e^{m}$$

$$= m$$

$$\boxed{E(x) = m}$$

Variance:

$$Var(x) = E(x^2) - (E(x))^2$$

$$E(x^2) = \sum_{r=0}^{\infty} r^2 \frac{e^{-m} m^r}{r!}$$

$$= e^{-m} \sum_{r=0}^{\infty} \frac{r^2 m^r}{r!}$$

$$= e^{-m} \sum_{r=0}^{\infty} r(r-1+1) \frac{m^r}{r!}$$

$$= e^{-m} \sum_{r=0}^{\infty} r(r-1)\frac{m^r}{r!} + e^{-m} \sum_{r=0}^{\infty} \frac{r m^r}{r!}$$

$$= e^{-m} \sum_{r=0}^{\infty} r(r-1)\frac{m^r}{r!} + m$$

$$= e^{-m} \sum_{r=2}^{\infty} \frac{m^r}{(r-2)!} + m$$

$$= e^{-m}\left[m^2 + \frac{m^3}{1!} + \frac{m^4}{2!} + \cdots\right] + m$$

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

$$= m^2 e^{-m} \cdot e^{m} + m$$

$$= m^2 + m$$

$$E(x^2) = m(m+1)$$

$$Var(x) = \cancel{m^2} + m - \cancel{m^2}$$

$$Var(x) = m$$

classmate

Date
Page

* A hospital switch board recieves an average of 4 emergency calls in a 10 minute interval. what is a probability that
i) there are atmost 2 emergency calls.
ii) there are exactly 3 emergency calls in a 10 minute interval

By data $m = 4$

$$P(X = r) = \frac{e^{-m} m^r}{r!}$$

at $m = 4 \Rightarrow P(X = r) = \frac{e^{-4} \cdot 4^r}{r!}$

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

$$= e^{-4}\left(\frac{4^0}{0!} + \frac{4}{1!} + \frac{4^2}{2!}\right)$$

$$= e^{-4}(1 + 4 + 8)$$

$$= 13 \cdot e^{-4}$$

$P(X\leq 2) = 0.2381$

ii) $P(X = 3) = \frac{e^{-4}(4)^3}{3!}$

$= 0.1953$

* If the probability of a bad reaction from a certain injection is 0.001. Determine the chance that out of 2000 individuals, more than 2 will get a bad reaction

$p = 0.001$

$m = np = 2000 \times 0.001 = 2$

$P(X \geq 2) = 1 - P(X \leq 2)$

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

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

$1 - e^{-2}[1 + 2 + 2]$

$1 - e^{-2}[5]$

$P(X>2) = 0.3233$

* A communication channel recieves independent pulses at the rate of 12 pulses / μ sec, the probability of transmission error is 0.001 for each μsec. compute the probability of

i) no error during during a μ sec

ii) one error per μ sec

iii) at least one error / μ sec

iv) 2 errors

v) at most 2 errors

$p = 0.001$

$m = np = 12 \times 0.001 = 0.012$

i) $P(X=r) = \frac{e^{-0.012} (0.012)^r}{r!}$

$P(X=0) = e^{-0.012} (1)$

$= 0.9880$

ii) $P(X=1) = e^{-0.012} \times 0.012$

$= 0.0118$

iii) $P(X \geq 1) = 1 - P(X<1)$

$= 1 - (P(X=0))$

$= 1 - 0.9880$

$= 0.012$

iv) $P(X=2) = \frac{e^{-0.012} (0.012)^2}{2}$

$= 0.000071$

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

$= 0.9998$

classmate

Date ______
Page ______

* A car hire firm has 2 cars, which it hires out day by day. The no. of demands for a car on a day is distributed as a Poisson distribution with mean 1.5 : calculate the proportion of day on which neither car is used & the proportion of days on which some demand is refused.

$$m = 1.5$$

$$P(X=x) = \frac{e^{-m} \cdot m^x}{x!}$$

i) neither car is hired

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

ii) demand is refused when demand is more than no. of available cars.

$$P(X>2) = 1 - P(X \leq 2)$$
$$1 - [P(0) + P(1) + P(2)]$$
$$1 - [0.2231 + 0.3346 + 0.2509]$$
$$1 - 0.8086$$
$$= 0.1914$$

* Fit a poisson distribution for the following data:

x	0	1	2	3	4
f	122	60	15	2	1

$N = \sum f$

$$\text{Mean} = \frac{\sum x \cdot f}{\sum f} = \frac{0+60+30+6+4}{200}$$

$$\text{Mean} = \frac{1}{2} = 0.5$$

$$m = 0.5$$

$$P(X=x) = \frac{e^{-0.5} (0.5)^x}{x!}$$

$P(X=0) = 0.6065$ $\quad$ $P(X=3) = 0.0126$

$P(X=1) = 0.3032$ $\quad$ $P(X=4) = 0.0015$

$P(X=2) = 0.0758$

Expected value = $N \times P(r)$
$= \sum f \times P(r)$

$E(0) = 121.3$
$E(1) = 60.64$
$E(2) = 15.16$
$E(3) = 2.52$
$E(4) = 0.3$

x	0	1	2	3	4
$E(X)$	121.3	60.64	15.16	2.52	0.3

For binomial distribution,
$$m = np$$
$$0.5 = 5 \times p$$
$$p = \frac{0.5}{5} = 0.1$$

* Stating the assumptions, derive poisson distribution as a limiting case of the binomial distribution.

Assumptions: p is very small
n is very large

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

Mean $(m) = np$

$p = \frac{m}{n}$ $q = 1 - \frac{m}{n}$

$$P(X=r) = {}^{n}C_{r} \left(\frac{m}{n}\right)^{r} \left(1-\frac{m}{n}\right)^{n-r}$$

$$= \frac{n!}{r!(n-r)!} \frac{m^{r}}{n^{r}} \frac{(n-m)^{n-r}}{n^{n-r}}$$

$$= \frac{n!}{r!(n-r)!} \frac{m^{r}(n-m)^{n-r}}{n^{n}}$$

$$= \frac{n(n-1)(n-2)\cdots(n-(r-1))(n-r)!}{r!\,(n-r)!} \frac{m^r (n-m)^{n-r}}{n^n}$$

$$= \frac{m^r}{r!} \frac{\cancel{n^r}(1)(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{r-1}{n}) \cdot \cancel{n^{n-r}}(1-\frac{m}{n})^{n-r}}{\cancel{n^n}}$$

taking limits $n \to \infty$

$$\Rightarrow \lim_{n\to\infty} \frac{m^r}{r!} (1-\tfrac{1}{n})(1-\tfrac{2}{n})(1-\tfrac{3}{n})\cdots(1-\tfrac{r-1}{n}) \; (1-\tfrac{m}{n})^{n-r}$$

$$\Rightarrow \lim_{n\to\infty} \frac{m^r}{r!} (1-\frac{m}{n})^{n-r}$$

1^∞ form.

$$\Rightarrow \frac{m^r}{r!} e^{\lim_{n\to\infty} (-\frac{m}{n})(n-r)} = \frac{m^r}{r!} e^{\lim_{n\to\infty} (-m + \frac{mr}{n})}$$

$$P(X=r) = \frac{m^r}{r!} e^{-m}$$

* A bag contains 1 red and 7 white marbles. A marble is drawn from the bag & it's colour is observed. then the marble is put back into the bag & contents are thoroughly mixed using
  i) Binomial distribution
  ii) Poisson approximation to Binomial distribution.

Find the probability that in 8 such drawings, a red ball is selected exactly 3 times.

p = probability of drawing a red ball $= \frac{1}{8}$

$q = 7/8$

$n = 8$

i) $P(X=r) = {}^{n}C_{r} (p)^r (q)^{n-r} = {}^{8}C_{r} (p)^r (q)^{8-r}$

$$P(X=3) = \frac{8\times7\times\cancel{6}}{3\times\cancel{2}\times1} \times \left(\frac{1}{8}\right)^3 \left(\frac{7}{8}\right)^5 = \frac{56 \times 7^5}{(8)^8}$$

$$P(X=3) = 0.0560$$

ii) $m = n \times p$

$m = 8 \times \frac{1}{8} = 1.$

$$P(X=r) = \frac{e^{-m} m^r}{r!} = \frac{e^{-1}(1)^r}{r!}$$

$$P(X=3) = \frac{e^{-1}}{3!} = 0.0613$$