Assignment - 4

① find the mean, median, mode, harmonic mean, geometric, standard deviation from the following distribution of marks obtained by 90 students

Marks	20	30	40	50	60	70	80	90
No. of students	5	12	15	20	18	10	6	4

i) Mean $= \bar{x} = \frac{\Sigma x_i}{n}$ where ~~n=8~~

$\Sigma x_i = 20+30+40+50+60+70+80+90.$

$= 440$

$\bar{x} = \frac{\cancel{440}^{110}}{\cancel{8}_{2}} = 55.$

$N = \Sigma f_i = 5+12+15+20+18+10+6+4 = 90$

$\bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{100+360+600+1000+700+480+360+1080}{90}$

$= \cancel{45}52$

ii) Mode = 50 (20 times occurred.)

iii)

x	f	Cummulative freq.
20	5	5
30	12	17
40	15	32
50	20	52
60	18	70
70	10	80
80	6	86
90	4	90

Meadian = Avg of $\left[\frac{90}{2}\right]^{th}$ & $\left[\frac{90+1}{2}\right]^{th}$ observation

= 45th & 46th observation

$\boxed{\text{Meadian} = 50}$

iv) Harmonic mean $= H.M = \dfrac{N}{\Sigma\left(\frac{f}{x}\right)}$

x	f	f/x	$f\log x$
20	5	1/4	$5\log(20) = 6.505$
30	12	2/5	17.725
40	15	3/8	24.030
50	20	2/5	33.979
60	18	3/10	32.006
70	10	1/7	18.450
80	6	3/40	11.418
90	4	2/45	7.816
		$\frac{626}{315} = 1.9873015873$	151.929

9/80

18/60

$H.M = \dfrac{90}{1.9873} = 45.28$

v) Geometric Mean $= G_m = \text{Antilog}\left[\dfrac{\Sigma f \ln x}{N}\right]$

$= \text{Antilog}\left[\dfrac{151.929}{90}\right]$

$= 10^{1.6881} = 48.764$

② Find the mean, median, mode, harmonic mean, geometric mean standard deviation from the following distribution.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students.	5	8	11	21	35	30	22	18

i)

Marks	No. of students (f_i)	m_i	$f_i m_i$	cf
0-10	5	5	25	5
10-20	8	15	120	13
20-30	11	25	275	24
30-40	21 f_0	35	735	$\boxed{45}$
40-50	35 f_1	45	1575	80
50-60	30 f_2	55	1650	110
60-70	22	65	1430	132
70-80	18	75	1350	150
	$\Sigma f_i = 150$		$\Sigma f_i m_i = 7160$	

i) Mean $= \bar{x} = \dfrac{7160}{150} = 47.73.$

ii) Median $= l + \left(\dfrac{\frac{N}{2} - c.f}{f}\right) \times i = 40 + \dfrac{(75-45)}{35} \times 10$

$= 48.57.$

$N = \dfrac{150}{2} = 75$

iii) Mode $= L + \left(\dfrac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times i$

$= 40 + \dfrac{(35-21)}{(2.35-21-30)} \times 10 = 48.42$

$= 47.368$

iv) H.M $= \text{Antilog}\left(\dfrac{\Sigma f \log m}{N}\right) = 10^{\frac{244.471}{150}}$

$= 42.549$

x	f	m	$f\log m$
0-10	5	5	3.494
10-20	8	15	9.408
20-30	11	25	15.377
30-40	21	35	32.428
40-50	35	45	57.862
50-60	30	55	52.210
60-70	22	65	39.884
70-80	18	75	33.751
	$\Sigma N = 150$		$\Sigma f\log m = 244.471$

(3) find the coefficients of skewness & kurtosis for the following distribution of serum cholestrol levels of patients in a hospital. $\bar{x} = \frac{159}{2} = 79.5$

Serum cholestrol	50-60	60-70	70-80	80-90	90-100
patients	6	10	24	22	18

x_i	m_i	f_i	$(m_i-\bar{x})$	$(m_i-\bar{x})^2$	$(m_i-\bar{x})^3$	$(m_i-\bar{x})^4$
50-60	55	6	$(55-79.5) = -24.5$	600.25	-14706.125	360300.0625
60-70	65	10	$(65-79.5) = -14.5$	210.25	-3048.625	44205.0625
70-80	75	24	$(75-79.5) = 0$	0	0	0
80-90	85	22	$(85-79.5) = 5.5$	30.25	166.375	915.0625
90-100	95	18	$(95-79.5) = 15.5$	240.25	3723.875	57720.0625
		$\Sigma f_i = 80$				

Central moment $= \mu_k = \frac{\Sigma f_i (m_i-\bar{x})^k}{N}$

$k=1 - \mu_1 = \frac{\Sigma\, 6(-24.5) + 10(-14.5) + 22(5.5) + 18(15.5)}{80}$

$= 1.35$

$k=2 = \mu_2 = \frac{6(-24.5)^2 + 10(-14.5)^2 + 22(5.5)^2 + 18(15.5)^2}{80}$

$= 133.675.$

$k=3,\ \mu_3 = \frac{6(-24.5)^3 + 10(-14.5)^3 + 22(5.5)^3 + 18(15.5)^3}{80}$

$= -600.41$

$k=4,\ \mu_4 = \frac{6(-24.5)^4 + 10(-14.5)^4 + 22(5.5)^4 + 18(15.5)^4}{80} = 45786.79$

$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-600.4)^2}{(133.675)^3} = 0.1509$, $\gamma_1 = \sqrt{\beta_1}$, $\gamma_1 = 0.388$

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{45786}{(133.675)^2} = 2.56 < 3$

$\gamma_2 = \beta_2 - 3 = -0.44$

(4) Calculate the 1st 4 moments about the mean, moments about the point 4 of the following distribution & hence calculate β_1, β_2.

Weekly earnings	0	1	2	3	4	5	6	7	8
No. of persons	1	8	28	56	70	56	28	8	1

Moments about Mean

x_i	f_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^3$	$(x_i-\bar{x})^4$
0	1	-4	16	-64	256
1	8	-3	9	-27	81
2	28	-2	4	-8	16
3	56	-1	1	-1	1
4	70	0	0	0	0
5	56	1	1	1	1
6	28	2	4	8	16
7	8	3	9	27	81
8	1	4	16	64	256
	$\Sigma f_i = 256$	0	60	0	708

$\bar{x} = \frac{\Sigma x_i f_i}{N} = \frac{8 + 56 + (168) + (280) + (280) + (168) + 56 + 8}{256}$

$= 4$

for discrete case central moments are given by $\mu_k = \frac{\Sigma f_i (x_i - \bar{x})^k}{N}$

$\mu_1 = \frac{\sum 256 \sum (x_i - \bar{x})}{256} = 0$

$\mu_2 = \frac{\cancel{\sum 256 \sum (x_i-\bar{x})^2}}{\cancel{256}} = \frac{\cancel{256 \times 60}}{\cancel{256}} = 60$

$\mu_2 = \frac{\sum 2(1\times 16 + 8\times 9 + 28\times 4 + 56\times 1)}{256}$

$= 2$

$\mu_3 = \frac{\sum \cancel{2.64} - 1\cancel{.64} + 8.\cancel{27} - 8\cancel{27} + 28.8 - 28.8 + 56 - 56}{256}$

$\mu_3 = 0$

$\mu_4 = \frac{\sum 2(1\times 256 + 8\times 81 + 28\times 16 + 56\times 1)}{256}$

$\mu_4 = 11$

$\therefore$ The first 4 central moments are $\mu_1, \mu_2, \mu_3, \mu_4$ = 0, 2, 0, 11.

& The moments about point 4 also $\mu_1 = 0$, $\mu_2 = 2$, $\mu_3 = 0$, $\mu_4 = 11$ ($\because$ mean & point coincide)

$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{0}{8} = 0$

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{11}{4} = 2.75$.

(5) Find Karl Pearsons Coeff of correlation b/w x & y from the following data.

x	11	12	13	14	15	16	17	18	19	20
y	30	29	29	25	24	24	24	21	18	15

$N = 10$

$\bar{x} = \frac{\sum x}{N} = \frac{155}{10} = 15.5$

$\bar{y} = \frac{\sum y}{N} = \frac{239}{10} = 23.9$.

x	y	xy	x^2	y^2
11	30	330	121	900
12	29	348	144	841
13	29	377	169	841
14	25	340	196	625
15	24	360	225	576
16	24	384	256	576
17	24	408	289	576
18	21	378	324	441
19	18	342	361	324
20	15	300	400	225
$\Sigma x = 155$	$\Sigma y = 239$	$\Sigma xy = 3567$	2485	5929

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

$r = \frac{(10\times 3567) - (155\times 239)}{\sqrt{10\times 2485 - (155)^2} \times \sqrt{10\times 5929 - (239)^2}} = -1.0374$

$\therefore r < 0$, negatively correlated.

(or)

x	y	$x = x_i - \bar{x}$	$y = y_i - \bar{y}$	xy	x^2	y^2
11	30	-4.5	6.1	-27.45	20.25	37.21
12	29	-3.5	5.1	-17.85	12.25	26.01
13	29	-2.5	5.1	-12.75	6.25	26.01
14	25	-1.5	1.1	-1.65	2.25	1.21
15	24	-0.5	0.1	-0.05	0.25	0.01
16	24	0.5	0.1	0.05	0.25	0.01
17	24	1.5	0.1	0.15	2.25	8.41
18	21	2.5	-2.9	-7.25	6.25	34.81
19	18	3.5	-5.9	-20.65	12.25	79.21
20	15	4.5	-8.9	-40.05	20.25	
$\Sigma x = 155$	$\Sigma y = 239$	$\Sigma x = 0$	$\Sigma y = 0$	-127.5	82.5	212.9

we have $\Sigma xy = -127.5$, $\Sigma x^2 = 82.5$, $\Sigma y^2 = 212.9$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{82.5}{10}} = 2.872$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{N}} = \sqrt{\frac{212.9}{10}} = 4.609$$

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y} = \frac{-127.5}{10 \times 2.872 \times 4.609}$$

$$r = -0.9629$$

$\therefore r < 0$, it is $-ve$ co-relation.

(8) Calculate the rank correlation coefficient for the following data.

x	81	78	73	69	68	62	58	73
y	10	12	18	18	22	20	24	18

w.k.t rank co-relation coeff $= R(\sigma)\rho = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} d_i^2$

$n = 7$

x	y	$d_i = (x-y)$	d_i^2
81	10	71	5041
78	12	66	4356
73	18	55	3025
73	18	55	3025
69	18	51	2601
68	22	46	2116
62	20	42	1764
58	24	34	1156

$\Sigma d_i^2 = 23084$

$$R = 1 - \frac{6}{8(64-1)}(23084)$$

$R = 1 - 358.19$

x	R_1	y	R_2	$d_i = (R_1 - R_2)$	d_i^2
81	1	10	8	-7	49
78	2	12	7	-5	25
73	3.5	18	5	-1.5	2.25
73	3.5	18	5	-1.5	2.25
69	5	18	5	0	0
68	6	22	2	4	16
62	7	20	3	4	16
58	8	24	1	7	49

$\Sigma d_i^2 = 159.5$

$$R = 1 - \frac{6}{8(63)} \cancel{\times 141.5} = \cancel{-0.6845}$$

$$R = 1 - \frac{6}{8 \times 63}\left[159.5 + \frac{1}{12}(2^2 - 2) + \frac{1}{12}(3^2 - 3)\right]$$

$r = -0.928$

(9) Obtain the rank correlation coefficient for the following data:

x	68	64	75	50	64	80	75	40	55	64
y	62	58	68	45	81	60	68	48	50	70

x	R_1	y	R_2	$d_i = (R_1 - R_2)$	d_i^2
68	4	62	5	-1	1
64	6	58	7	-1	1
75	2.5	68	3.5	-1	1
50	9	45	10	-1	1
64	6	81	1	5	25
80	1	60	6	-5	25
75	2.5	68	3.5	-1	1
40	10	48	9	1	1
55	8	50	8	0	0
64	6	70	2	4	16

$\Sigma d_i^2 = 72$

$$R = 1 - \frac{6}{n(n^2-1)}\left(\Sigma d_i^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \ldots\right)$$

$$= 1 - \frac{6}{10\times 99}\left(72 + \frac{1}{12}(2^3-2) + \frac{1}{12}(3^3-3) + \frac{1}{12}(2^3-2)\right)$$

$$R = 0.545$$

(10) The following are the measurements of air velocity & evaporation coefficient of burning fuel droplets in an impulse engine.

Air velocity (cm/sec) x	20	60	100	140	180	220
Evaporation coeff (mm^2/sec) y	0.18	0.37	0.35	0.78	0.56	0.75
	300	340	380	260		
	1.36	1.17	1.65	1.18		

Fit a straight line to these data by the method of least squares & use it to estimate evaporation coeff of a droplet when air velocity 190 cm/sec.

Let eq^n of straight line $y = a + bx$ — (1)

Normal eq^{n}s are $\Sigma y = na + b\Sigma x$ — (2)

$\Sigma xy = a\Sigma x + b\Sigma x^2$ — (3)

x	y	xy	x^2
20	0.18	3.6	400
60	0.37	22.2	3600
100	0.35	35	10000
1400	0.78	109.2	19600
180	0.56	100.8	32400
220	0.75	165	48400
260	1.18	306.8	67600
300	1.36	408	90000
340	1.17	397.8	115600
380	1.65	~~639.4~~	144400
2000 ~~1940~~	~~7.686~~ 8.35	1989.4 ~~1424.8~~	~~464400~~ 532000

substituting in 2 & 3 $\Rightarrow$ ~~7.86~~ 8.35 $= 10a + b2000$

~~21~~1989.4 $= 2000a + b532000$

~~$a = 1.0447$~~ ~~$a = 0.1038$~~

~~$b = 0.0031880$~~ ~~1.813~~

$a = 0.069$, $b = 0.0038$

$\Rightarrow$ $y = ax + b \Rightarrow y = 0.069x + 0.0038$

when $x = 190 \Rightarrow y = 0.069(190) + 0.0038$

$\boxed{y = 0.79}$

(11) The following data on the percentage of high performance radial tires made by a certain manufacturer that are still usable after having been driven for the given no. of times.

Miles driven (thousand) x	1	2	5	10	20	30	40	50
Percentage usable y	98.2	91.7	81.3	64.0	36.4	32.6	17.1	11.3

fit an exponential curve $y = ab^x$ by applying the method of least squares to the data points. Use the result to estimate what percentage of the manufacturer's high performance radial tires will last at least 25,000 miles.

$y = ab^x$

apply log on b.s $\log y = \log a + x\log b$.

this is same as $Y = A + Bx$.

where $A = \log a$, $B = \log b$, $Y = \log y$.

Normal eq^n are $\Sigma Y = nB + B\Sigma x$ — (1)

$\Sigma xY = B\Sigma x + B\Sigma x^2$ — (2)

x	$Y = \log y$	x^2	xy
1	1.992	1	1.992
2	1.962	4	3.924
5	1.9100	25	9.55
10	1.806	100	18.06
20	1.561	400	31.22
30	1.513	900	45.39
40	1.232	1600	49.28
50	1.053	2500	52.65
$\Sigma x = 158$	$\Sigma Y = 13.029$	$\Sigma x^2 = 4930$	$\Sigma xY = 212.066$

$\Rightarrow$ from (1) & (2) $13.029 = 8A + B\,158$ — (3)

$212.066 = 158A + B\,4930$ — (4)

solving (3) & (4) we have $A = \cancel{1.782} \approx 2.122$

$B = \cancel{-0.014} = -0.015$

$\Rightarrow a = 10^{A} = 10^{2} = 100$ $\cancel{A = \ln a \Rightarrow a = e^{A} = e^{2.782}} = 5.941$

$b = 10^{B} = 10^{-0.015}$ $\cancel{B = \ln b \Rightarrow b = e^{b} = e^{-0.014}} = 8.34 = 0.986 = 0.975$

$= 0.94$

$y = ab^{x} = 100(0.94)^{x}$

$y = 100(0.94)^{25}$

$y = 21.29$

(12) The following are data on the drying time of a certain paint & the amount of an additive that is intended to reduce the drying time

Amount of paint additive (grms) x	0	1	2	3	4	5	6	7	8
Drying time (hrs) y	12	10.5	10	8	7	8	7.5	8.5	9

Fit a second degree polynomial (parabola) by the method of least squares to the data points. Use the result to estimate predict the drying time of the paint when 6.5 grams of the additive is being used.

let parabola eq^n $y = a + bx + cx^2$ — (1)

normal eq^n's are

$\Sigma y = na + b\Sigma x + c\Sigma x^2$ — (2)

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

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

x	y	x^2	x^3	x^4	xy	x^2y
0	12	0	0	0	0	0
1	10.5	1	1	1	10.5	10.5
2	10	4	8	16	20	40
3	8	9	27	81	24	72
4	7	16	64	256	28	112
5	8	25	125	625	40	200
6	7.5	36	216	1296	45	270
7	8.5	49	343	2401	59.5	416.5
8	9	64	512	4096	72	576
36	80.5	204	1296	8772	299	1697

$\frac{8(9)}{2}$

$\Rightarrow$ $80.5 = \cancel{a}\,9a + 36b + 204c$ — (5)

$299 = 36a + 204b + 1296c$ — (6)

$1697 = 204a + 1296b + 8772c$ — (7)

solving (5), (6), (7) we get $a = 12.184$

$b = -1.846$

$c = 0.1829$.

$\Rightarrow$ from (1) $y = 12.18 - 1.85x + 0.182x^2$.

when $x = 6.5$ $\boxed{y = 7.8}$

(13) Predict y at $x = 3.75$ by fitting a power curve $y = ax^b$ to the given data.

x	1	2	3	4	5	6
y	2.98	4.26	5.21	6.10	6.80	7.50

$$y = ax^b$$

taking log on b.s $\log y = \log a + b\log x$.

which is same as $Y = A + bX$.

where $Y = \log y$, $A = \log a$, $X = \log x$

Now Normal Distrib^n eq^n are

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

$X = \log a$	$Y = \log y$	X^2	XY
0	0.474	0	0
0.301	0.629	0.0906	0.189
0.477	0.716	0.227	0.341
0.602	0.785	0.362	0.472
0.698	0.832	0.487	0.580
0.778	0.875	0.605	0.680
$\Sigma X = 2.856$	$\Sigma Y = 4.3111$	$\Sigma X^2 = 1.771$	$\Sigma XY = 2.262$

$\Rightarrow$ $4.311 = 6A + 2.856b$. — (1)

$2.262 = 2.856A + 1.771b$. — (2)

Solving from (1) & (2) we get $A = 0.4\cancel{7}5$

$B = 0.\cancel{51}0$

$\Rightarrow$ $a = 10^{0.475} = \cancel{3.019} = 2.985$

~~$b = 10^{0.49} = 3.090$~~

~~$y = 3x^3$~~.

$\Rightarrow$ $y = 2.98\,x^{0.51}$

$y = 2.98(3.75)^{0.51} = 5.847$

(14) The following results were worked out from scores in Mathematics (x) & statistics (y) of students in an examinations.

	x	y
Mean	39.5	47.5
Standard deviation	10.8	17.8

Karl pearsons Correlation coefficient $= +0.42$ find both the regression lines. Use these regression & estimate the value of y for $x = 50$ & also estimate the value for $y = 30$

x on y is given by.

$$x - \bar{x} = \underbrace{r\frac{\sigma x}{\sigma y}}_{b_{xy}}(y - \bar{y})$$

$$(x - 39.5) = 0.42\,\frac{10.8}{17.8}(y - 47.5)$$

$$x = 0.42(0.606)y - \cancel{0.42}((47.5)) + 39.5$$

$$x = 0.254y - 12.06 + 39.5$$

$$\boxed{x = 0.245y + 27.44}$$

y on x is given by b_{yx}.

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

$$(y - 47.5) = 0.42\,\frac{17.8}{10.8}(x - 39.5)$$

$$y = 0.42(1.648)(x - 39.5) + 47.5$$

$$y = 0.692x - 27.334 + 47.5$$

$$\boxed{y = 0.692x + 20.166.}$$

$$y = 0.69(50) + 20.166,$$

$$\boxed{y = 54.666}$$

$$x = 6.245(30) + 27.44$$

$$\boxed{x = 34.79}$$

(15) from the following data, obtain two regression eq^{ns}. Also calculate co-relation coefficient.

x	6	2	10	4	8
y	9	11	5	8	7

we have y on x $(y-\bar{y}) = \left(\frac{r\sigma y}{\sigma x}\right)(x-\bar{x})$

or $(y-\bar{y}) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}(x-\bar{x})$

x on y & $(x-\bar{x}) = \frac{r\sigma x}{\sigma y}(y-\bar{y})$

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

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
6	9	0	1	0	1	0
2	11	-4	3	16	9.	-12
10	5	4	-3	16	9	-12
4	8	-2	0	4	0	0
8	7	2	-1	4	1	-2
		0	0	40	20	-26

$\Sigma x = 30$ $\Sigma y = 40$

$\bar{x} = \frac{30}{5} = 6.$ $\bar{y} = 40/5 = 8.$

$\Rightarrow (y-8) = \frac{-\cancel{26}^{13}}{\cancel{40}_{20}}(x-6)$

$y = \frac{-13}{20}x + \frac{13\times 6}{20} + 8.$

$\boxed{y = -0.65x + 11.9}$

$\Rightarrow (x-6) = \frac{-26}{20}(y-8)$

$x = \frac{-26}{20}y + \frac{26\times 8}{20} + 6$

$\boxed{x = -1.3y + 16.4}$

w.k.t $b_{xy} = -1.3$, $b_{yx} = -0.65$

Co-relation coefficient $r = \pm\sqrt{b_{yx} . b_{xy}}$.

$r = \pm\sqrt{(1.3)(0.65)}$

$\boxed{r = -0.919}$ (because b_{yx} & b_{xy} is -ve)

(16) show that θ, an acute angle b/w the two lines of regression, is given by $\tan\theta = \left(\frac{1-r^2}{r}\right)\frac{\sigma x \sigma y}{\sigma_x^2 + \sigma_y^2}$, interpret the case when $r = 0, \pm 1$.