NAME: SHANMUHAPRIYAA RAJU

REG NO: 21MAI1003

DIGITAL ASSIGNMENT - 1

ANSWER ALL THE QUESTIONS:

1. Obtain the mean, median, mode, standard deviation, Quartiles and Quartile coefficient of dispersion for the following data:

(a)

x:	1	2	3	4	5	6	7	8	9
f:	8	10	11	16	20	25	15	9	6

Assignment Digital:-

1 (a)

	x	f	c.f.	$x_i f_i$
	1	8	8	8
	2	10	18	20
	3	11	29	33
	4	16	45	64
	5	20	65	100
	6	25	90	150
	7	15	105	105
	8	9	114	72
	9	6	120	54
Total	45	120	594	606

mean $= \dfrac{\Sigma f x}{\Sigma f}$

$= \dfrac{606}{120}$

$\boxed{\text{mean} = 5.05}$

median $= l + \dfrac{\frac{N}{2} - m}{f} + c$

(for continous data)

median = value of $\left(\frac{N}{2}^{th}\right)$ observation

= value of $\left(\frac{120}{2}^{th}\right)$ observation

= value of 60^{th} observation

= $(x = 5)$ values is > 60

$\boxed{\text{median} = 5}$

mode = freq has max frequency

= 25

$\boxed{\text{mode} = 6}$

① ⓐ

x	f	fx	$x-\mu$ (5.05)	$(x-\mu)^2$	$f(x-\mu)^2$
1	8	8	-4.05	16.4025	131.22
2	10	20	-3.05	9.301	93.01
3	11	33	-2.05	4.202	46.222
4	16	64	-1.05	1.102	17.632
5	20	100	-0.05	0.0025	0.05
6	25	150	0.95	0.9025	22.563
7	15	105	1.95	3.803	57.045
8	9	72	2.95	8.703	78.327
9	6	54	3.95	15.603	93.618

$\Sigma f = 120$ $\Sigma = 539.687$

$$\sigma = \sqrt{\frac{539.687}{(120-1)}}$$

$$\sigma = 2.121$$

$$Q_1 = l_1 + \left(\frac{\frac{N}{4} - m_1}{f_1}\right) * C_1$$

$$Q_3 = l_3 + \left(\frac{\frac{3N}{4} - m_3}{f_3}\right) * C_3.$$

$$Q_1 = 4 + \left(\frac{\frac{120}{4} - 29}{16}\right) * 1$$

$$Q_1 = 4.0625$$

$$Q_3 = 6 + \left(\frac{\frac{3(120)}{4} - 65}{25}\right) * 1$$

$$Q_3 = 7$$

Quartile deviation = $\frac{Q_3 - Q_1}{2} = \frac{7 - 4.0625}{2} = 1.4688$

Coefficient of Q.D = $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{7 - 4.0625}{7 + 4.0625} = 0.26655 \simeq 0.2666$

(b) The mean of the five numbers 6, 9, 2, x, y is 5 and the standard deviation is √6. Find the values of x and y.

(b) mean of 5 no's is 6, 9, 2, x, y is 5. S.D=

$5 = \frac{\Sigma fx}{\Sigma f} = \frac{\text{sum of data}}{\text{no of terms}}$

$5 = \frac{6 + 9 + 2 + x + y}{5}$

$\Rightarrow x + y = 25 - 17$

$\boxed{x + y = 8} \rightarrow y = 8 - x$ —(1)

$\sqrt{S.D} =$ variance $\Rightarrow 6 = \frac{1}{n}\left(\sum_i (x_i - \mu)^2\right).$

$6 = \frac{1}{5}\left(1 + 4^2 + (-3)^2 + (x-5)^2 + (y-5)^2\right]$

$30 = 1 + 16 + 9 + (x^2 + 25 - 10x) + (y^2 + 25 - 10y)$

$x^2 + y^2 - 10x - 10y + 46 = 0$

Sub (ii) $\Rightarrow x^2 + (8-x)^2 - 10x - 10(8-x) + 46 = 0$

$x^2 + 64 + x^2 - 16x - 10x - 80 + 10x + 46 = 0$

$2x^2 - 16x + 30 = 0$

$\div$ by (2)

$(x^2 - 8x + 15) = 0$

$(x-5)(x-3) = 0$

$\Rightarrow x = 5$ or $x = 3$

When, $x = 5$, | $x = 3$
$y = 8 - 5$ | $y = 8 - 3$
$= 3$ | $= 5$

so the other two numbers are (3, & 5)

2. Calculate Karl Pearson's Coefficient of skewness from the following data :

Profits	No of Cos.	Profits	No of Cos.
70-80	12	110-120	50
80-90	18	120-130	45
90-100	35	130-140	30
100-110	42	140-150	8

②

C.I	f	m	$d = \left(\frac{m-A}{C}\right)$	fd	fd^2	cf
70-80	12	75	-4	-48	192	12
80-90	18	85	-3	-54	162	30
90-100	35	95	-2	-70	140	65
100-110	42	105	-1	-42	42	107
110-120	50	115	0	0	0	157
120-130	45	125	1	45	45	202
130-140	30	135	2	60	120	232
140-150	8	145	3	24	72	240
N = 240	A = 115	C = 10		-85	773	

$$\bar{X} = A + \left(\frac{\sum fd}{N} \times C\right) = 115 - \frac{85}{240} \times 10$$

$$\bar{X} = 111.458$$

Median class $= \frac{N}{2}^{th}$ item $= \frac{240}{2}^{th} = 120^{th}$ item $= (110-120)$

$$M = L + \left[\frac{\left(\frac{N}{2} - cf\right)}{f} \times i\right]$$

$$= 110 + \left(\frac{120 - 107}{50} \times 10\right)$$

$$= 110 + 2.6 \Rightarrow M = \underline{\underline{112.6}}$$

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times C$$

$$= \sqrt{\frac{773}{240} - \left(\frac{-85}{240}\right)^2} \times 10 = \sqrt{3.221 - 0.1254}$$

$$= \sqrt{3.0926} \times 10$$

$$\Rightarrow \sigma = 1.7593 \times 10 = \underline{\underline{17.593}}$$

$$\text{K.P.C. skewness} = 3\left(\frac{\bar{x} - M}{\sigma}\right)$$

$$= 3\left(\frac{111.458 - 112.6}{17.593}\right)$$

$$= -0.1947$$

Distribution is negatively skewed.

3. The grades of a class of 9 students on a midterm report (x) and on the final examination (y) are as follows:

x	77	50	71	72	81	94	96	99	67
y	82	66	78	34	47	85	99	99	68

Estimate the linear regression line of x on y.

③ linear regression of x on y : [$x = a + by$]

x	y	xy	x^2	y^2
77	82	6314	5929	6724
50	66	3300	2500	4356
71	78	5538	5041	6084
72	34	2448	5184	1156
81	47	3807	6561	2209
94	85	7990	8836	7225
96	99	9504	9216	9801
99	99	9801	9801	9801
67	68	4556	4489	4624
Σ of all: 707	658	53258	57557	51980

Regression line x on y:

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

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

$$= \frac{9 \times 53258 - 707 \times 658}{9 \times 51980 - (658)^2} = \frac{14116}{34856}$$

$$= 0.4049$$

Regression eq of x on y

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

$\bar{x} = \frac{\sum x}{n} = 707/9 = 78.555$

$\bar{y} = \frac{\sum y}{n} = 658/9 = 73.111$

$x - 78.555 = 0.4049(y - 73.111)$

$x = 0.4049y + 48.947 \Rightarrow x = 0.4049y + 48.947$

Vellore Institute of Technology

4. The table shows the weights X1 to the nearest pound (lb), the heights X2 to the nearest inch (in) , and the ages X3 to the nearest year of 12 boys.

Weight (X1)	64	71	53	67	55	58	77	57	56	51	76	68
Height (X2)	57	59	49	62	51	50	55	48	52	42	61	57
Age (X3)	8	10	6	11	8	7	10	9	10	6	12	9

(i) Find the regression equation of X1 on X2 and X3

(ii) Determine the estimated values of X1 for the given value X2=56 and X3=10.

(iii) Compute r(X1,X2) for the above data.

(4)

X_1	X_2	X_3	X_2^2	X_3^2	X_2X_3	X_1X_2	X_1X_3	X_1^2
64	57	8	3249	64	456	3648	512	4096
71	59	10	3481	100	590	4189	710	5041
53	49	6	2401	36	294	2597	318	2809
67	62	11	3844	121	682	4154	737	4489
55	51	8	2601	64	408	2805	440	3025
58	50	7	2500	49	350	2900	406	3364
77	55	10	3025	100	550	4235	770	5929
57	48	9	2304	81	432	2736	513	3249
56	52	10	2704	100	520	2912	560	3136
51	42	6	1764	36	252	2142	306	2601
76	61	12	3721	144	732	4636	912	5776
68	57	9	3249	81	513	3876	612	4624
Σ total. 753	643	106	34843	976	5779	40830	6796	4813

x_1 on $x_2, x_3 \Rightarrow x_1 = b_{12.3}\, x_2 + b_{13.2}\, x_3$

normal equations are,

$$\Sigma X_1 = nb_0 + b_1 \Sigma X_2 + b_2 \Sigma X_3 \quad \text{—} ①$$

$$\Sigma X_1 X_2 = b_0 \Sigma X_2 + b_1 \Sigma X_2^2 + b_2 \Sigma X_2 X_3 \quad \text{—} ②$$

$$\Sigma X_1 X_3 = b_0 \Sigma X_3 + b_2 \Sigma X_3^2 + b_1 \Sigma X_2 X_3 \quad \text{—} ③$$

on substituting values in ①, ②, ③.

$$753 = 12b_0 + 643b_1 + 106\, b_2 \quad \text{—} ①$$

$$40830 = 643b_0 + 34843\, b_1 + 5779\, b_2 \quad \text{—} ②$$

$$6796 = 106\, b_0 + 5779\, b_1 + 976\, b_2 \quad \text{—} ③$$

$1+4.0625 \simeq \underline{\underline{0.2666}}$

(4) (a)

$b_0 = 3.651 \qquad b_1 = 0.8546 \qquad b_2 = 1.5063.$

(i) The Regression equation is

$$X_1 = b_0 + b_1 X_2 + b_2 X_3$$

$$X_1 = 3.651 + 0.8546 X_2 + 1.5063 X_3.$$

(ii) when $X_2 = 56$ & $X_3 = 10.$

$$X_1 = +3.651 + 0.8546(56) - 10(1.5063).$$

$$X_1 = 66.5718.$$

(iii) $$r(X_1, X_2) = \frac{n\sum x_1 x_2 - \sum x_1 \sum x_2}{\sqrt{n\sum x_1^2 - (\sum x_1)^2}\sqrt{n\sum x_2^2 - (\sum x_2)^2}}$$

$$= \frac{12(40830) - (753)(643)}{\sqrt{12(48139) - 753^2}\sqrt{12(34843) - (643)^2}}$$

$$= \frac{5781}{\sqrt{10659}\sqrt{4667}} = \frac{5781}{103.24 * 68.3}$$

$$r(X_1, X_2) = \underline{\underline{0.8196}}$$

5. Given that $r_{12}=-0.89$, $r_{13}=-0.97$, $r_{23}=0.96$ then find $r_{12.3}$ and $R_{3.12}$.

(5) $r_{12} = -0.89$ $r_{13} = -0.97$ $r_{23} = 0.96$

$r_{12.3} = ?$ (partial) $R_{3.12} = ?$ (multiple)

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

$$= \frac{-0.89 - (-0.97)(0.96)}{\sqrt{(1 - (-0.97)^2)(1 - 0.96^2)}}$$

$$r_{12.3} = \frac{0.0412}{\sqrt{(0.0591)(0.0784)}}$$

$$r_{12.3} = \frac{0.0412}{0.06806} = 0.60526.$$

$$r_{12.3} \approx 0.6053$$

$$R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2 r_{31} r_{32} r_{12}}{1 - r_{12}^2}}$$

$$= \sqrt{\frac{(-0.97)^2 + (0.96)^2 - 2(-0.89)(-0.97)(0.96)}{1 - (-0.89)^2}}$$

$$= \sqrt{\frac{0.9409 + 0.9216 - 1.657}{1 - 0.7921}}$$

$$R_{3.12} = 0.9929.$$

6. A statistical investigator obtains the following regression equations in a survey: $X - Y - 6 = 0$ and $0.64X + 4.08 = Y$. Here X = age of husband and Y = age of wife. Find (i) means of X and Y (ii) correlation coefficient between X and Y and (iii) Standard deviation of X and Y.

6) $X - Y - 6 = 0$ and $Y = 0.64X + 4.08$.

X = husband age Y = wife age.

(i) $X - Y - 6 = 0$ — ①

$0.64X - Y + 4.08 = 0$ — ②

As both lines pass through regression lines, mean values, point $(\bar{x}, \bar{y})$ satisfies both equations.

$\bar{x} - \bar{y} = 6$ — ①

$0.64\bar{x} - \bar{y} = -4.08$

(−) (+) (+)

$0.36\bar{x} = 10.08$

$\Rightarrow \bar{x} = 28$

& $\bar{y} = 28 - 6 = 22$

So, the mean value for X & Y is 28 and 22

(ii) for correlation b/w X & Y, calculate b_{yx} & b_{xy}

$X = Y + 6$ $\quad\quad Y = 0.64x + 4.08$

$b_{xy} = \frac{1}{1} = 1$ $\quad\quad b_{yx} = \frac{0.64}{1} \neq 1$

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

$= 1 \times 0.64 = 0.64$

$r = \sqrt{0.64} \Rightarrow r = \pm 0.8$

Both regression coefficient values are +ve, and

hence, $r = 0.8$.

(iii) Since variance is not given,

and the ratio of

$\frac{y}{x} = 0.8$

so, $\frac{x}{y} = \frac{1}{0.8} = 1.25$

Hence the ratio would be, for

x/y its 1.25. and for

y/x its 0.8.

7. (a) A fund has a sample R-squared value close to 0.5 and it is doubtlessly offering higher risk adjusted returns with the sample size of 50 for 5 predictors. Find Adjusted R square value.

(b) Give the interpretation of Regression plots using scale location.

7) (a) sample size = n = 50

$R^2 = 0.5$ $\quad k = 5$

$$\bar{R}^2 = 1 - (1 - R^2)\left[\frac{n-1}{n-(k+1)}\right]$$

$$= 1 - (1 - 0.5)\left[\frac{50-1}{50-(5+1)}\right]$$

$$R^2_{adj} = 1 - (0.5)\left(\frac{49}{44}\right)$$

$$\bar{R}^2_{adj} = 0.4431$$

(b) Give interpretation using scale location for regression plots.

→ It usually helps to show if the residuals are spread equally along the range of input variables.

→ It displays the fitted values of model along x axis & square root of standardized residuals in y axis.

→ We check for homoscedasticity to check the assumption for equal variance.

→ It should be randomly scattered along the regression line with roughly the values being fitted with equal variability.

Using these, scale plots can be interpreted.