UNIT 11

Information Handling

Some useful Formulae

Correlation Coefficient Formulae

- $r = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\ \text{Var}(Y)}}$
- $r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$
- $r = \frac{n\sum XY - nXY}{\sqrt{[\sum X^2 - nX^2][\sum Y^2 - nY^2]}}$
- $r = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum (X-\bar{X})^2 \sum (Y-\bar{Y})^2}}$
- $r = \frac{\sum XY - nXY}{\sqrt{[\sum X^2 - nX^2][\sum Y^2 - nY^2]}}$
- $r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}}$

Types of Correlation

Positive or direct correlation: If r is positive, the relationship between the domain and range values has a **positive or direct correlation**. In this case, if the domain value increases, the range value also tends to increase and vice versa. In this case both variables move in the same direction. The value of correlation coefficient for positive correlation is between 0 and 1. i.e. $0 < r < 1$.

Negative or inverse correlation: If r is negative, the linear relationship between the domain and range values has a **negative or inverse correlation**. In this case, if the domain value increases, the range value tends to decrease. In this case both variables move in the opposite direction. The value of correlation coefficient for negative correlation is between -1 and 0. i.e. $-1 < r < 0$.

Zero or null correlation: The absence of any relation between the variables is called zero correlation. In this case variables are independent to each other. i.e. $r = 0$.

Estimated Regression Line

The estimated regression line of Y on X; $Y = a_{YX} + b_{YX}X$	The estimated regression line of X on Y; $X = a_{XY} + b_{XY}Y$
$b_{YX} = \dfrac{\sum(X-\overline{X})(Y-\overline{Y})}{\sum(X-\overline{X})^2}$	$b_{XY} = \dfrac{\sum(X-\overline{X})(Y-\overline{Y})}{\sum(Y-\overline{Y})^2}$
$b_{YX} = \dfrac{\sum(XY-n\overline{X}\,\overline{Y})}{\sum(X-\overline{X})^2}$	$b_{XY} = \dfrac{\sum(XY-n\overline{X}\,\overline{Y})}{\sum(Y-\overline{Y})^2}$
$b_{YX} = \dfrac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}}$	$b_{XY} = \dfrac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}}$
$b_{YX} = \dfrac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$	$b_{XY} = \dfrac{n\sum XY - (\sum X)(\sum Y)}{n\sum Y^2 - (\sum Y)^2}$
$b_{YX} = \dfrac{S_{XY}}{S_X^2} = r\dfrac{S_Y}{S_X}$	$b_{XY} = \dfrac{S_{XY}}{S_Y^2} = r\dfrac{S_X}{S_Y}$
$a_{YX} = \dfrac{\sum Y - b_{YX}\sum X}{n}$	$a_{XY} = \dfrac{\sum X - b_{XY}\sum Y}{n}$
$a_{YX} = \overline{Y} - b_{YX}\overline{X}$	$a_{XY} = \overline{X} - b_{XY}\overline{Y}$

EXERCISE 11.1

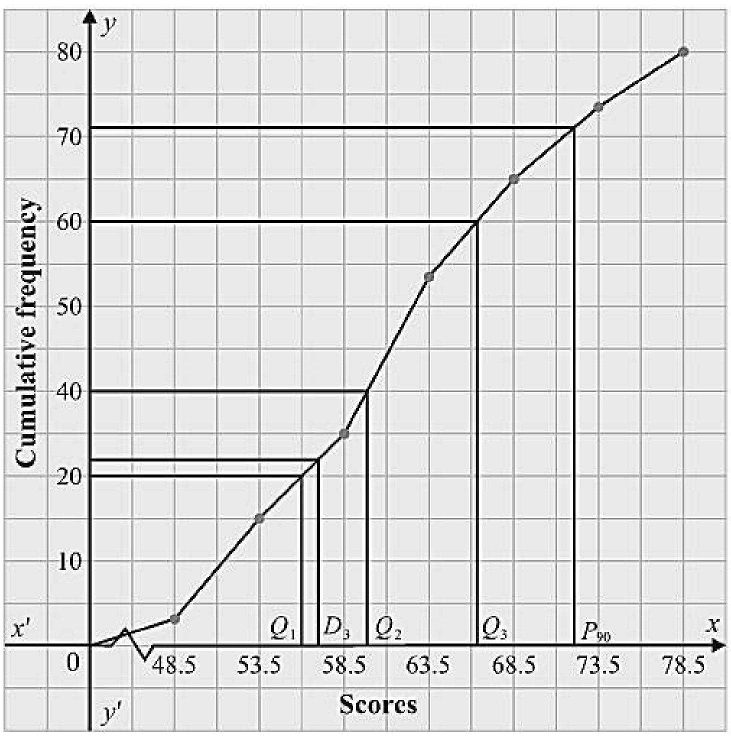
1. The following frequency distribution represents the achievement of a student's group upon a memory test:
   (i) Plot cumulative frequency graph of group scores.
   (ii) Determine median, lower and upper quartiles, D_3, P_{90} and interquartile range graphically.

Note: Write your answers in whole numbers.

Scores	44 – 48	49 – 53	54 – 58	59 – 63	64 – 68	69 – 73	74 – 78
No. of Students	3	12	15	23	12	8	7

Solution

(i)



(ii) Median = 61,
$Q_1 = 55$,
$Q_3 = 66$,
$D_3 = 57$,
$P_{90} = 73$,
$IQR = 11$

(ii) $N = 80$, Class width, $h = 5$

Formulae: $\text{Median} = L + \dfrac{\frac{N}{2} - c_f}{f} \times h$,

$$Q_1 = L + \frac{\frac{N}{4} - c_f}{f} \times h, \quad Q_3 = L + \frac{\frac{3N}{4} - c_f}{f} \times h$$

$$D_k = L + \frac{\frac{kN}{10} - c_f}{f} \times h, \quad P_k = L + \frac{\frac{kN}{100} - c_f}{f} \times h$$

Median: $\dfrac{N}{2} = \dfrac{80}{2} = 40$

Median class $= 59-63$

$L = 58.5$, $c_f = 30$, $f = 23$

$$\text{Median} = 58.5 + \frac{40 - 30}{23} \times 5 = 58.5 + \frac{10}{23} \times 5 = 58.5 + 2.1739 = 60.67 \approx 61$$

Q_1: $\dfrac{N}{4} = \dfrac{80}{4} = 20$

Class $= 54-58$

$L = 53.5$, $c_f = 15$, $f = 15$

$$Q_1 = 53.5 + \frac{20 - 15}{15} \times 5 = 53.5 + \frac{5}{15} \times 5 = 53.5 + 1.6667 = 55.17 \approx 55$$

Q_3: $\dfrac{3N}{4} = \dfrac{3(80)}{4} = 60$

Class $= 64-68$

$L = 63.5$, $c_f = 53$, $f = 12$

$$Q_1 = 63.5 + \frac{60 - 53}{12} \times 5 = 63.5 + \frac{7}{12} \times 5 = 63.5 + 2.9167 = 66.42 \approx 66$$

D_3: $\dfrac{3N}{10} = \dfrac{3(80)}{10} = 24$

Class $= 54-58$

$L = 53.5$, $c_f = 15$, $f = 15$

$$D_3 = 53.5 + \frac{24 - 15}{15} \times 5 = 53.5 + \frac{9}{15} \times 5 = 53.5 + 3 = 56.5 \approx 57$$

P_{90}: $\dfrac{90N}{100} = \dfrac{90(80)}{100} = 72$

Class $= 69-73$

$L = 68.5$, $c_f = 65$, $f = 8$

$$P_{90} = 68.5 + \frac{72 - 65}{8} \times 5 = 68.5 + \frac{7}{8} \times 5 = 68.5 + 4.375 = 72.875 \approx 73$$

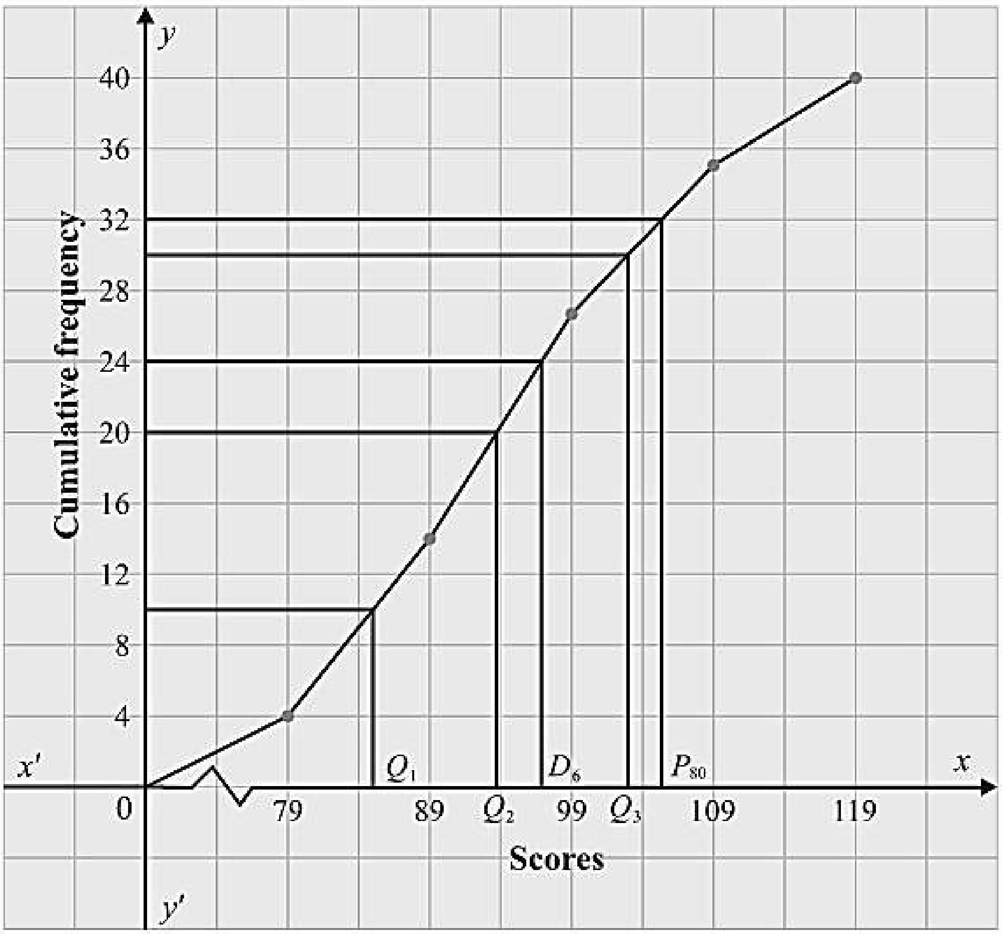
$$\text{IQR} = Q_3 - Q_1 = 66.42 - 55.17 = 11.25 \approx 11$$

2. Construct an ogive for the following distribution of scores:

Scores	69 – 79	79 – 89	89 – 99	99 – 109	109 – 119
f	4	10	13	8	5

Determine median, lower and upper quartiles, D_6, P_{80} and interquartile range graphically.

Solution



Median = 94,

$Q_1 = 85$,

$Q_3 = 103$,

$D_6 = 97$,

$P_{80} = 105$

$IQR = 18$

Scores	69–79	79–89	89–99	99–109	109–119
f	4	10	13	8	5
C.F.	4	14	27	35	40

$N = 40$, $h = 10$

Formulae: $\text{Median} = L + \dfrac{\frac{N}{2} - c_f}{f} \times h$, $Q_1 = L + \dfrac{\frac{N}{4} - c_f}{f} \times h$,

$Q_3 = L + \dfrac{\frac{3N}{4} - c_f}{f} \times h$, $D_6 = L + \dfrac{\frac{6N}{10} - c_f}{f} \times h$,

$P_{80} = L + \dfrac{\frac{80N}{100} - c_f}{f} \times h$

Median: $\dfrac{N}{2} = \dfrac{40}{2} = 20$
Class = 89–99
$L = 88.5$, $c_f = 14$, $f = 13$

$$\begin{aligned}\text{Median} &= 88.5 + \frac{20-14}{13} \times 10\\ &= 88.5 + \frac{6}{13} \times 10\\ &= 88.5 + 4.6154\\ &= 93.12 \approx 93\end{aligned}$$

Q_1: $\dfrac{N}{4} = \dfrac{40}{4} = 10$
Class = 79–89
$L = 78.5$, $c_f = 4$, $f = 10$

$$\begin{aligned}Q_1 &= 78.5 + \frac{10-4}{10} \times 10\\ &= 78.5 + \frac{6}{10} \times 10\\ &= 78.5 + 6\\ &= 84.5 \approx 85\end{aligned}$$

Q_3: $\dfrac{3N}{4} = \dfrac{3(40)}{4} = 30$
Class = 99–109
$L = 98.5$, $c_f = 27$, $f = 8$

$$\begin{aligned}Q_3 &= 98.5 + \frac{30-27}{8} \times 10\\ &= 98.5 + \frac{3}{8} \times 10\\ &= 98.5 + 3.75\\ &= 102.25 \approx 102\end{aligned}$$

D_6: $\dfrac{6N}{10} = \dfrac{6(40)}{10} = 24$
Class = 89–99
$L = 88.5$, $c_f = 14$, $f = 13$

$$\begin{aligned}D_6 &= 88.5 + \frac{24-14}{13} \times 10\\ &= 88.5 + \frac{10}{13} \times 10\\ &= 88.5 + 7.6923\\ &= 96.19 \approx 96\end{aligned}$$

P_{80}: $\dfrac{80N}{100} = \dfrac{80(40)}{100} = 32$
Class = 99–109
$L = 98.5$, $c_f = 27$, $f = 8$

$$\begin{aligned}P_{80} &= 98.5 + \frac{32-27}{8} \times 10\\ &= 98.5 + \frac{5}{8} \times 10\\ &= 98.5 + 4.25\\ &= 104.75 \approx 105\end{aligned}$$

$$\begin{aligned}\text{IQR} &= Q_3 - Q_1\\ &= 102.25 - 84.5\\ &= 17.75 \approx 17\end{aligned}$$

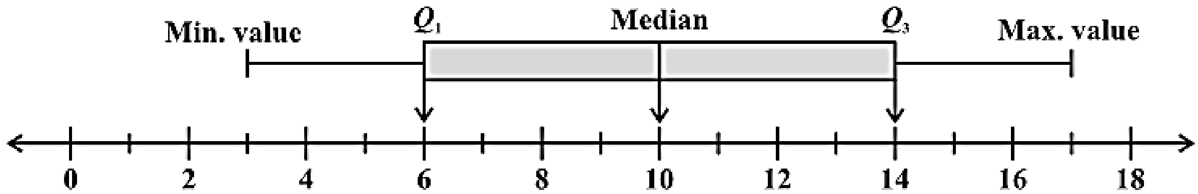
Answers are in approximation so both of above true either from book or my given solution

3. Find Q_1, Q_3, median, range and IQR for the dataset given below:

3, 6, 8, 4, 7, 5, 10, 11, 13, 9, 14, 12, 15, 16, 17

Also draw a box-and-whisker plot.

Solution



Ordered data:

3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 $\qquad n = 15$

Median $= \dfrac{n+1}{2} = \dfrac{15+1}{2} = 8^{\text{th}}$ item

Median = 10

Q_1 position $= \dfrac{n+1}{4} = \dfrac{16}{4} = 4^{\text{th}}$ item

$Q_1 = 6$

Q_3 position $= \dfrac{3(n+1)}{4} = \dfrac{3(16)}{4} = 12^{\text{th}}$ item

$Q_3 = 14$

Range = Max - Min = 17 − 3 = 14

IQR = $Q_3 - Q_1$ = 14 − 6 = 8

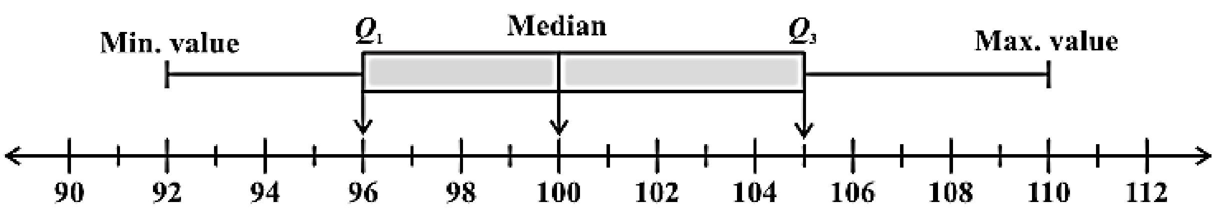
Five-number summary: (3, 6, 10, 14, 17)

4. Find Q_1, Q_3, median, range, IQR and extreme values plot for the following data:

102, 98, 95, 100, 93, 110, 108, 104, 97, 96, 92, 101, 99, 105, 107

Also draw a box-and-whisker plot.

Solution



Ordered data:

92, 93, 95, 96, 97, 98, 99, 100, 101, 102, 104, 105, 107, 108, 110 $\qquad n = 15$

Median $= \dfrac{n+1}{2} = \dfrac{15+1}{2} = 8^{\text{th}}$ item

Median = 100

Q_1 position $= \dfrac{n+1}{4} = \dfrac{16}{4} = 4^{\text{th}}$ item

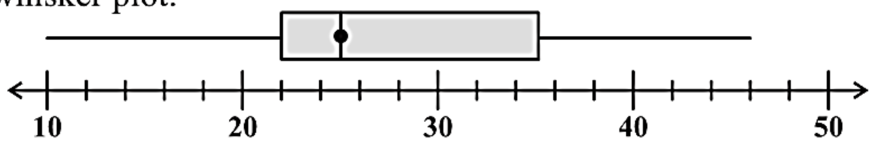
$Q_1 = 96$

Q_3 position $= \dfrac{3(n+1)}{4} = \dfrac{3(16)}{4} = 12^{\text{th}}$ item

$Q_3 = 105$

Five-number summary: (92, 96, 100, 105, 110)

5. Find Q_1, Q_3, median, minimum and maximum values for the following box-and-whisker plot:

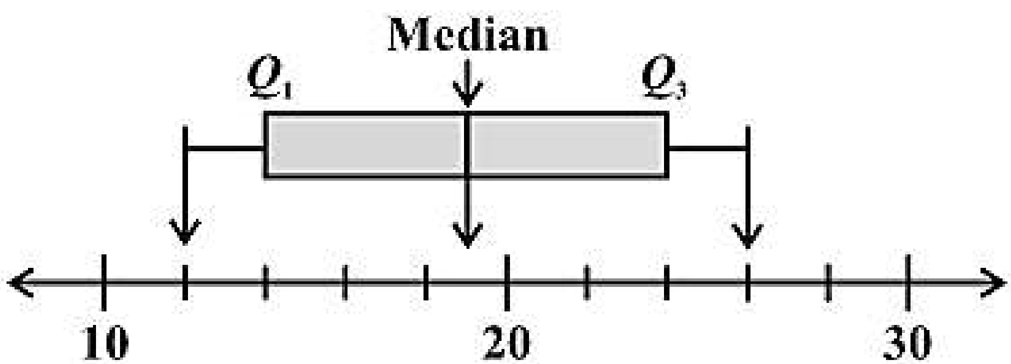


Solution

Minimum	=	10
Q_1	=	22
Median	=	25
Q_3	=	35
Maximum	=	46

6. Draw the box-and-whisker plot, if min. value = 12, max. value = 26, median = 19, Q_1 = 14 and Q_3 = 24.

Solution

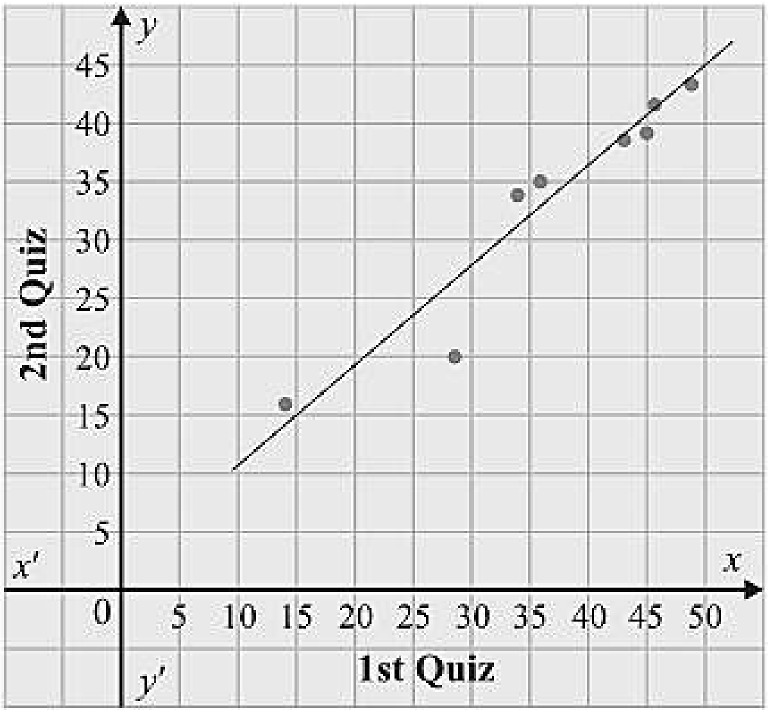


7. Construct a scatter diagram and draw a line of best fit for the following quiz scores for 8 students in a class:

1^{st} quiz	14	28	34	36	43	45	46	49
2^{nd} quiz	16	20	33	35	38	39	42	43

Describe correlation between 1^{st} and 2^{nd} quiz scores also.

Solution



Strong positive correlation between 1st and 2nd quiz.

1^{st} quiz	14	28	34	36	43	45	46	49
2^{nd} quiz	16	20	33	35	38	39	42	43

To describe correlation:
As the 1^{st} quiz scores increase, the 2^{nd} quiz scores also increase.
The points would lie close to an upward sloping line.
Therefore, there is a **strong positive correlation.**

8. The following table shows the test scores (one measuring speed and the other measuring strength) of football players out of 200:

1^{st} Test	20	38	70	40	100	110	50	136	150	80
2^{nd} Test	160	142	130	120	70	50	140	44	36	100

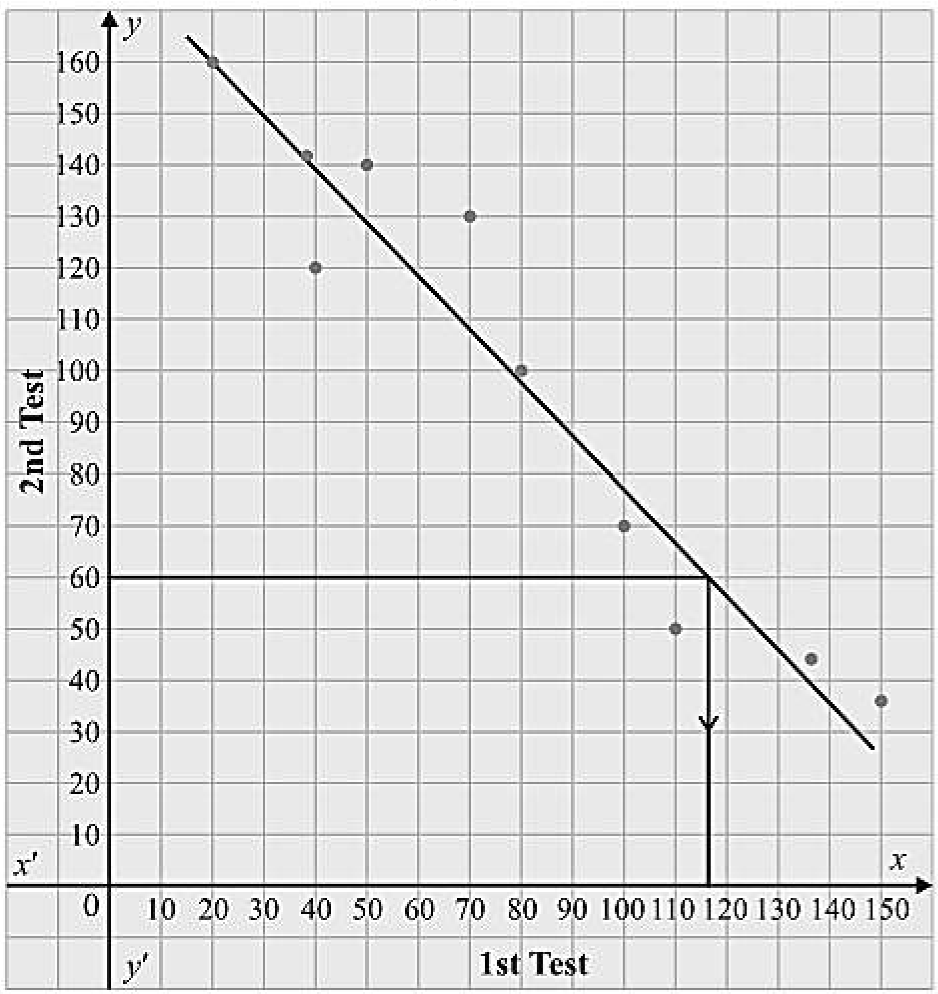
(i) Construct a scatter diagram and draw a line of best fit.

(ii) Describe correlation between the 1^{st} and 2^{nd} tests.

(iii) Abdullah scores 60 in the 2^{nd} test. Estimate his score in the 1^{st} test.

Solution

8. (i)



(ii) Strong negative correlation. **(iii)** 115.5

(iii)

x	y	xy	y^2
20	160	3200	25600
38	142	5396	20164
70	130	9100	16900
40	120	4800	14400
100	70	7000	4900
110	50	5500	2500
50	140	7000	19600
136	44	5984	1936
150	36	5400	1296
80	100	8000	10000
794	**992**	**62280**	**114500**

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum y^2 - (\sum y)^2} = \frac{622800 - 787648}{1145000 - 984064}$$

$$b = \frac{-164848}{160936} = -1.0243$$

$$a = \bar{x} - b\bar{y} = 79.4 + 1.0243(99.2) = 181.01$$

$$x = 181.01 - 1.0243y$$

$$x = 181.01 - 1.0243(60) = 115.5$$

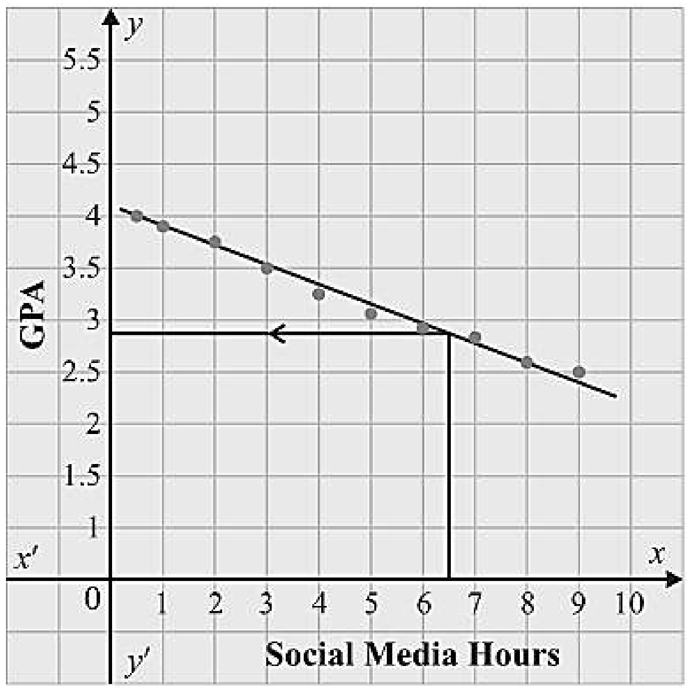
9. The following table shows the number of daily social media hours and GPA of university students.

Social Media Hours	0.5	1	2	3	4	5	6	7	8	9
GPA	4.0	3.9	3.7	3.5	3.3	3.1	2.9	2.8	2.6	2.5

(i) Construct a scatter diagram and draw a line of best fit.
(ii) Identify the nature of correlation between social media use and GPA.
(iii) Estimate the GPA for a student who spends 6.5 hours on social media.

Solution

(i)



(ii) Strong negative correlation. **(iii)** 2.9

(iii) Estimate GPA for $x = 6.5$ hours

n	$\sum x$	$\sum y$	$\sum xy$	$\sum x^2$
10	45.5	32.3	135.55	295.25

Least squares regression of y on x:

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}, \quad a = \frac{\sum y - b\sum x}{n}$$

$$b = \frac{10(135.55) - (45.5)(32.3)}{10(295.25) - (45.5)^2} = \frac{1,355.5 - 1,470.65}{2,952.5 - 2,070.25} = \frac{-115.15}{882.25} = -0.1305$$

$$a = \frac{32.3 - (-0.1305)(45.5)}{10} = \frac{32.3 + 5.939}{10} = \frac{38.239}{10} = 3.8239$$

Regression equation: $y = 3.8239 - 0.1305x$

For $x = 6.5$, $y = 3.8239 - 0.1305(6.5) = 3.8239 - 0.84825 = 2.97565 \approx 2.98$

Estimated GPA for 6.5 hours is about 2.98.

EXERCISE 11.2

1. Find the range of the following data sets:

(i) 63, 89, 98, 125, 79, 108, 117, 60 (ii) 43.5, 13.6, 18.9, 38.4, 61.4, 29.4

Solution

(i) Data: 63, 89, 98, 125, 79, 108, 117, 60

Highest value = 125, Lowest value = 60

Range = Highest – Lowest = 125 – 60 = 65

Therefore, $\boxed{\text{Range} = 65}$

(ii) Data: 43.5, 13.6, 18.9, 38.4, 61.4, 29.4

Highest value = 61.4, Lowest value = 13.6

Range = Highest – Lowest = 61.4 – 13.6 = 47.8

Therefore, $\boxed{\text{Range} = 47.8}$

---

2. If the range and the lowest value of a set of data are 46.7 and 13.4 respectively, then find the highest value.

Solution

Given, Range = 46.7 and Lowest value = 13.4

Range = Highest – Lowest

46.7 = Highest – 13.4

Highest = $\boxed{46.7 + 13.4 = 60.1}$

Therefore, $\boxed{\text{Highest value} = 60.1}$

---

3. Calculate the range of the following data:

Income (in Rs.)	4000 – 4500	4500 – 5000	5000 – 5500	5500 – 6000	6000 – 6500
No. of workers	8	12	30	21	6

Solution

Highest income = 6500, Lowest income = 4000

Range = Highest – Lowest = 6500 – 4000 = 2500

Therefore, **Range = Rs. 2500**

4. A group of 7 workers reported the number of items they assembled in a day as:

52, 55, 50, 53, 54, 56, 52

Find the standard deviation and variance of the items assembled.

Solution

Data: 52, 55, 50, 53, 54, 56, 52 $(n = 7)$

Mean, $\bar{x} = \frac{\sum x}{n} = \frac{52+55+50+53+54+56+52}{7} = \frac{372}{7} = 53.14$

Variance, $\sigma^2 = \frac{\sum (x-\bar{x})^2}{n}$

$$= \frac{(52-53.14)^2+(55-53.14)^2+(50-53.14)^2+(53-53.14)^2+(54-53.14)^2+(56-53.14)^2+(52-53.14)^2}{7}$$

$$= \frac{24.86}{7} = 3.55$$

Standard deviation, $\sigma = \sqrt{\sigma^2} = \sqrt{3.55} = 1.88$

Variance = 3.55, Standard Deviation = 1.88

5. A librarian recorded the number of visitors during 5 days of a week.

120, 135, 130, 125, 140

Calculate the variance and standard deviation of visitors.

Solution

Data: 120, 135, 130, 125, 140 $(n = 5)$

Mean, $\bar{x} = \dfrac{120+135+130+125+140}{5} = \dfrac{650}{5} = 130$

Variance, $\sigma^2 = \dfrac{\sum(x-\bar{x})^2}{n}$

$= \dfrac{(120-130)^2+(135-130)^2+(130-130)^2+(125-130)^2+(140-100)^2}{7}$

$= \dfrac{100+25+0+25+100}{5} = \dfrac{250}{5} = 50$

Standard deviation, $\sigma = \sqrt{50} = 7.07$

Variance = 50, Standard Deviation = 7.07

6. Find the range, variance and standard deviation of first 23 odd numbers.

Solution

First 23 odd numbers are 1,3,5,…,45

Range $= 45 - 1 = 44$

$\overline{\mathrm{X}} = \frac{\text{First+Last}}{2} = \frac{1+45}{2} = 23$

Variance $= \sigma^2 = \frac{n^2-1}{n} = \frac{23^2-1}{3} = \frac{529-1}{3} = \frac{528}{3} = 176$

Standard Deviation $= \sigma = \sqrt{\frac{n^2-1}{n}} = \sqrt{176} = 13.27$

7. The rainfall recorded in various places of five districts in a week is given below. Find its variance and standard deviation.

Rainfall (in mm)	42	51	54	61	63	71
Number of places	5	13	4	9	5	4

Solution

x	f	x - $\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
42	5	-13.925	194.006	970.03
51	13	-4.925	24.256	315.33
54	4	-1.925	3.706	14.82
61	9	5.075	25.756	231.80
63	5	7.075	50.056	250.28
71	4	15.075	227.256	909.02
Total	**40**			**2691.28**

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{2237}{40} = 55.925$$

$$\sigma^2 = \frac{\sum f(x-\bar{x})^2}{N} = \frac{2691.28}{40} = 67.282$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{67.282} \approx 8.20$$

8. Machine A Output (units): 98, 100, 102, 101, 99
Machine B Output (units): 95, 100, 105, 90, 110
(i) Which machine has better performance?
(ii) Which machine is more consistent?

Solution

$$\bar{x}_A = \frac{98+100+102+101+99}{5} = 100$$

$$\sigma_A^2 = \frac{(98-100)^2+(100-100)^2+(102-100)^2+(101-100)^2+(99-100)^2}{5} = \frac{10}{5} = 2$$

$$\sigma_A = \sqrt{2} = 1.41$$

Machine B: 95, 100, 105, 90, 110

$$\bar{x}_B = \frac{95+100+105+90+110}{5} = 100$$

$$\sigma_B^2 = \frac{(95-100)^2+(100-100)^2+(105-100)^2+(90-100)^2+(110-100)^2}{5} = \frac{250}{5} = 50$$

$$\sigma_B = \sqrt{50} = 7.07$$

(i) Both machines have equal performance.

(ii) Machine A is more consistant.

9. The monthly sales (rupees in lacs) for two salespersons over 6 months are:

Person A: 5.5, 5.7, 5.4, 5.6, 5.8, 5.6

Person B: 5.5, 6.5, 4.5, 6.0, 5.0, 6.0

Compare their performance and consistency.

Solution

Person A: 5.5, 5.7, 5.4, 5.6, 5.8, 5.6

$\bar{x}_A = 5.60$, $\sigma_A = 0.14$

Person B: 5.5, 6.5, 4.5, 6.0, 5.0, 6.0

$\bar{x}_B = 5.58$, $\sigma_B = 0.72$

Comparison:

(i) Mean: A (5.60) > B (5.58) $\Rightarrow$ Person A has slightly better average sales.

(ii) Standard deviation: A (0.14) < B (0.72) $\Rightarrow$ Person A is more consistent.

10. The table given below shows the daily wages of workers in a textile mill, grouped into six income brackets:

Daily Wage (Rs)	800 – 1000	1000 – 1200	1200 – 1400	1400 – 1600	1600 – 1800	1800 – 2000
Frequency	2	4	6	8	2	1

Calculate the mean, variance and standard deviation of the wages.

Solution

Wage (Rs)	f_i	Midpoint x_i	$f_i x_i$	$x_i - \bar{x}$	$f_i(x_i - \bar{x})^2$
800-1000	2	900	1800	-460.87	425,025.8
1000-1200	4	1100	4400	-260.87	272,300.5
1200-1400	6	1300	7800	-60.87	22,227.6
1400-1600	8	1500	12,000	139.13	154,938.3
1600-1800	2	1700	3400	339.13	230,045.2
1800-2000	1	1900	1900	539.13	290,663.5
Total	**23**		**31,300**		**1,395,200.9**

Mean

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{31300}{23} = 1360.87$$

Variance

$$\sigma^2 = \frac{\sum f_i (x_i - \bar{x})^2}{N} = \frac{1,395,200.9}{23} = 60642.72$$

Standard Deviation

$$\sigma = \sqrt{\sigma^2} = \sqrt{60642.72} = 246.26$$

11. A company forecasts monthly sales (rupees in millions): 15, 18, 14, 20, 13. Find variability in sales predictions.

Solution

Mean $\mu = \frac{15+18+14+20+13}{5} = \frac{80}{5} = 16$

x	$x-\mu$	$(x-\mu)^2$
15	-1	1
18	2	4
14	-2	4
20	4	16
13	-3	9
Total		**34**

Variance $= \sigma^2 = \frac{\sum(x-\mu)^2}{N} = 6.8 \text{ million}^2$

Standard Deviation $= \sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}} = \sqrt{6.8} = 2.61$ million

12. Unemployment rates (%) in five provinces are 5.2, 6.0, 4.8, 5.5, 6.2. Calculate standard deviation and describe is there balanced unemployment rate?

Solution

Mean $\mu = \frac{5.2+6.0+4.8+5.5+6.2}{5} = 5.54$

x	$x-\mu$	$(x-\mu)^2$
5.2	-0.34	0.1156
6.0	0.46	0.2116
4.8	-0.74	0.5476
5.5	-0.04	0.0016
6.2	0.66	0.4356
Total		**1.312**

Variance $= \sigma^2 = \frac{\sum(x-\mu)^2}{N} = 0.2624$

Standard Deviation $= \sigma = \sqrt{0.2624} = 0.51$

Unemployment rates are fairly balanced

13. Find variance and standard deviation:

(i) $\Sigma x = 45, \Sigma x^2 = 421, n = 5$ (ii) $\Sigma fx = 210, \Sigma fx^2 = 7560, \Sigma f = 6$

(iii) $\bar{x} = 18, \Sigma fx^2 = 1670, \Sigma f = 5$

Solution

(i) $\sum x = 45, \sum x^2 = 421, n = 5$

$$\sigma^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \frac{421}{5} - \left(\frac{45}{5}\right)^2 = 84.2 - 81 = 3.2$$

$$\sigma = \sqrt{3.2} = 1.79$$

(ii) $\sum fx = 210, \sum fx^2 = 7560, \sum f = 6$

$$\sigma^2 = \frac{7560}{6} - \left(\frac{210}{6}\right)^2 = 1260 - 1225 = 35$$

$$\sigma = \sqrt{35} = 5.92$$

(iii) $\bar{x} = 18, \sum fx^2 = 1670, \sum f = 5$

$$\sigma^2 = \frac{1670}{5} - 18^2 = 334 - 324 = 10$$

$$\sigma = \sqrt{10} = 3.16$$

REVIEW EXERCISE 11

1. Four possible answers are given for the following questions. Choose the correct answer.

(i) ________ is used to get the cumulative frequencies.

✓(a) Addition (b) square root (c) multiplication (d) division

(ii) Second quartile represents:

(a) mean (b) mode ✓(c) median (d) variance

(iii) First quartile divides the data into ________ equal parts.

(a) Two (b) Three ✓(c) four (d) ten

(iv) Difference between the highest and the lowest values is called:

(a) mean (b) variance ✓(c) range (d) standard deviation

(v) Scatter diagram represents a relationship between ________ variables.

(a) five ✓(b) two (c) three (d) four

(vi) ________ is measure of dispersion:

(a) mean (b) median (c) mode ✓(d) variance

(vii) Positive square root of variance is called:

(a) mean (b) median
✓(c) standard deviation (d) range

(viii) ________ is not measure of dispersion.

(a) range ✓(b) arithmetic mean
(c) variance (d) standard deviation

(ix) Variance of the data 8, 8, 8, 8, 8, 8 is:

✓(a) 0 (b) 16 (c) 8 (d) 48

(x) Range of first 20 natural numbers is:

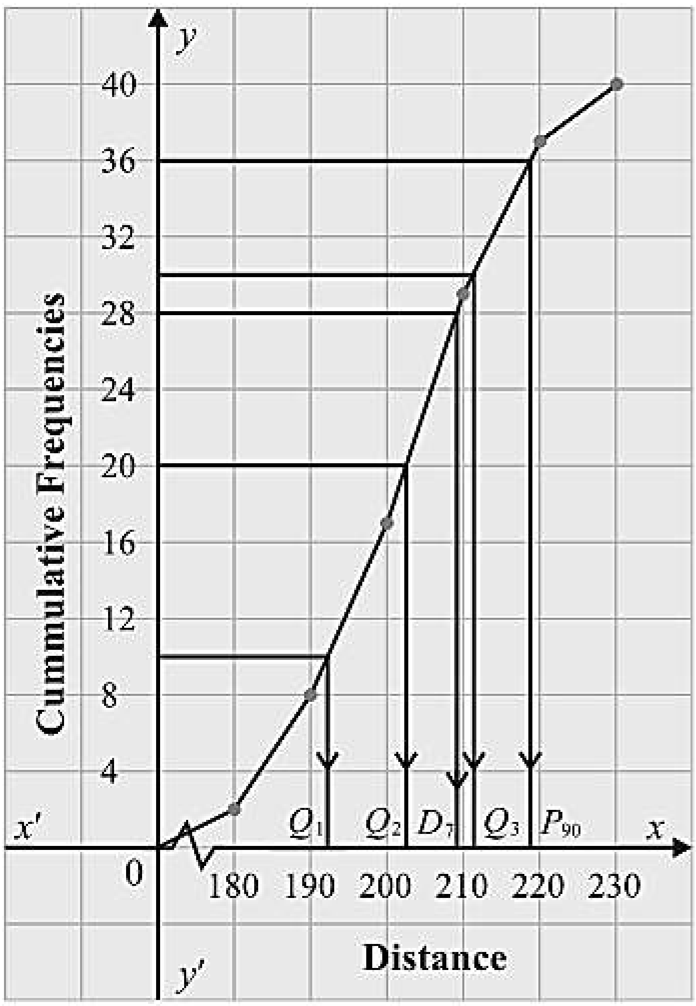
(a) 20 (b) 10 ✓(c) 19 (d) 30

2. The following results for the long jump were recorded:

Distance (in cm)	170 – 180	180 – 190	190 – 200	200 – 210	210 – 220	220 – 230
f	2	6	9	12	8	3

Construct the cumulative frequency polygon and locate median, Q_1, Q_3, D_7, P_{90} and interquartile range on it.

Solution



Median $= 202.5$, $Q_1 = 192.2$, $Q_3 = 211.3$,

$D_7 = 209.2$, $P_{90} = 218.8$, $IQR = 19.1$

(ii)

Distance (in cm)	f	cf
170 – 180	2	2
180 – 190	6	8
190 – 200	9	17
200 – 210	12	29
210 – 220	8	37
220 – 230	3	40

Total $N = 40$

Cumulative frequency (less-than type): (upper class boundary, c.f.)

$(170, 0), (180, 2), (190, 8), (200, 17), (210, 29), (220, 37), (230, 40)$

$$\text{Median} = l + \left(\frac{N/2 - c.f._{prev}}{f_m}\right)h = 200 + \left(\frac{20-17}{12}\right)10 = 202.50\ cm$$

$$Q_1 = l + \left(\frac{N/4 - c.f._{prev}}{f}\right)h = 190 + \left(\frac{10-8}{9}\right)10 = 192.22\ cm$$

$$Q_3 = l + \left(\frac{3N/4 - c.f._{prev}}{f}\right)h = 210 + \left(\frac{30-29}{8}\right)10 = 211.25\ cm$$

$$D_7 = l + \left(\frac{7N/10 - c.f._{prev}}{f}\right)h = 200 + \left(\frac{28-17}{12}\right)10 = 209.17\text{cm}$$

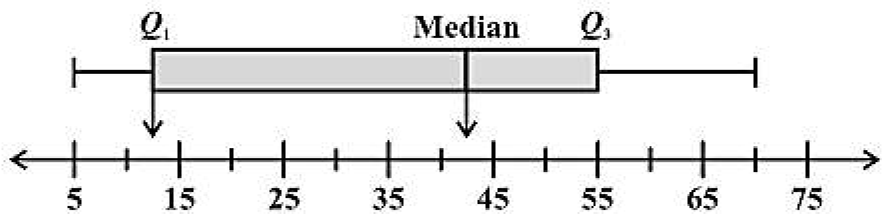
$$P_{90} = l + \left(\frac{90N/100 - c.f._{prev}}{f}\right)h = 210 + \left(\frac{36-29}{8}\right)10 = 218.8\ cm$$

$$\text{Interquartile range} = Q_3 - Q_1 = 211.25 - 192.22 = 19.03\ cm$$

3. The summary statistics for a data set is given below. Show it with a box-and-whisker plot.

Min. Value	Max. Value	Q_1	Median	Q_3
5	70	12.6	43	55.6

Solution



4. The following table shows the weekly hours spent exercising and resting heart rate (beats per minute) for ten individuals:

Exercise Hours	0	1	2	3	4	5	6	7	8	9
Heart Rate	90	85	83	80	76	74	72	70	68	65

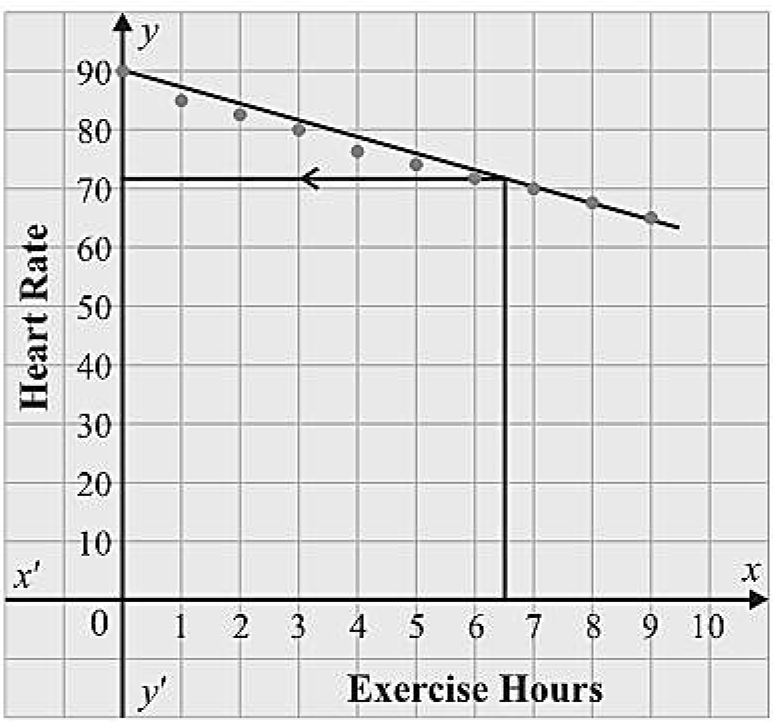
(i) Plot the data on a scatter diagram and draw a line of best fit.

(ii) State the type of correlation observed.

(iii) Predict the heart rate of Sakeena who exercises for 6.5 hours per week.

Solution

(i)



(ii) Strong negative correlation.

(iii) 71

x	y	xy	x^2	y^2
0	90	0	0	8100
1	85	85	1	7225
2	83	166	4	6889
3	80	240	9	6400
4	76	304	16	5776
5	74	370	25	5476
6	72	432	36	5184
7	70	490	49	4900
8	68	544	64	4624
9	65	585	81	4225
45	**763**	**3216**	**285**	**58799**

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} = -2.64 \text{ and } a = \frac{y - b\sum x}{n} = 88.16$$

The best fitted line is $\hat{y} = a + bx = 88.16 - 2.64x$

(ii)

Type of correlation: $r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} = -0.993$

Strong negative correlation

(iii)

Prediction for 6.5 hours

$$\hat{y} = a + bx = 88.16 - 2.64(6.5) = 71.03$$

5. Find the range for the given data, 25 , 30 , 35 , 40 , 50 , 60 , 65 , 75

Solution

Range = 75 – 25 = 50

6. Calculate range, variance and standard deviation for the following data set:

Class Interval	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
f	4	12	20	24	16	4

Solution

Class Interval	f	Midpoint x	fx	$x-\bar{x}$	$f(x-\bar{x})^2$
0 – 5	4	2.5	10	-14.75	870.25
5 – 10	12	7.5	90	-9.75	1140.75
10 – 15	20	12.5	250	-4.75	451.25
15 – 20	24	17.5	420	0.25	1.50
20 – 25	16	22.5	360	5.25	441.00
25 – 30	4	27.5	110	10.25	420.25
Total	**80**		**1240**		**3325.00**

Range

$$\text{Range} = 30 - 0 = 30$$

Mean

$$\bar{x} = \frac{\sum fx}{N} = \frac{1240}{80} = 15.5$$

Variance

$$\sigma^2 = \frac{\sum f(x-\bar{x})^2}{N} = \frac{3080}{80} = 38.5$$

Standard deviation

$$\sigma = \sqrt{38.5} \approx 6.2$$

7. The sum of 5 numbers is 45 and the sum of their squares is 421. Find the mean and standard deviation of the data.

Solution

Sum of numbers, $\sum x = 45$ (given)

Sum of squares, $\sum x^2 = 421$ (given)

Mean $\bar{x} = \frac{\sum x}{n} = \frac{45}{5} = 9$

Variance $\sigma^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \frac{421}{5} - 9^2 = 3.2$

Standard deviation $\sigma = \sqrt{3.2} = 1.789$

8. The monthly household expenses (rupees in thousands) of two families for 6 months were:

Family A: 45, 47, 46, 48, 46, 47

Family B: 38, 52, 40, 50, 42, 49

Calculate the mean and standard deviation of the monthly expenses. Which family spends more on average? Which family has more stable expenses?

Solution

Family A: 45, 47, 46, 48, 46, 47

$$\bar{x}_A = \frac{279}{6} = 46.50$$

$$\sum x_A^2 = 13031 \Rightarrow s_A^2 = \frac{13031}{6} - (46.5)^2 = 0.9167$$

$$s_A = 0.9574$$

Family B: 38, 52, 40, 50, 42, 49

$$\bar{x}_B = \frac{271}{6} = 45.17$$

$$\sum x_B^2 = 12521 \Rightarrow s_B^2 = \frac{12521}{6} - (45.17)^2 = 32.8056$$

$$s_B = 5.7285$$

9. The daily wages of 40 workers in a factory are grouped as follows:

Daily Wages (Rs.)	1000-1200	1200-1400	1400-1600	1600-1800	1800-2000
No. of Workers	6	10	12	8	4

Find the mean, variance and standard deviation of the daily wages.

Solution

Daily Wages (Rs.)	f	Midpoint x	fx	$x-\bar{x}$	$f(x-\bar{x})^2$
1000–1200	6	1100	6600	-370	822600
1200–1400	10	1300	13000	-170	289000
1400–1600	12	1500	18000	30	10800
1600–1800	8	1700	13600	230	423200
1800–2000	4	1900	7600	430	739600
Total	**40**		**58800**		**2285200**

Midpoint $x = \frac{\text{lower} + \text{upper}}{2}$

2. Mean

$$\bar{x} = \frac{\sum fx}{N} = \frac{58800}{40} = 1470$$

3. Variance

$$\sigma^2 = \frac{\sum f(x-\bar{x})^2}{N} = \frac{2285200}{40} = 57130$$

Using the given value: $\sigma^2 = 57100$

4. Standard deviation

$$\sigma = \sqrt{57100} \approx 238.96$$