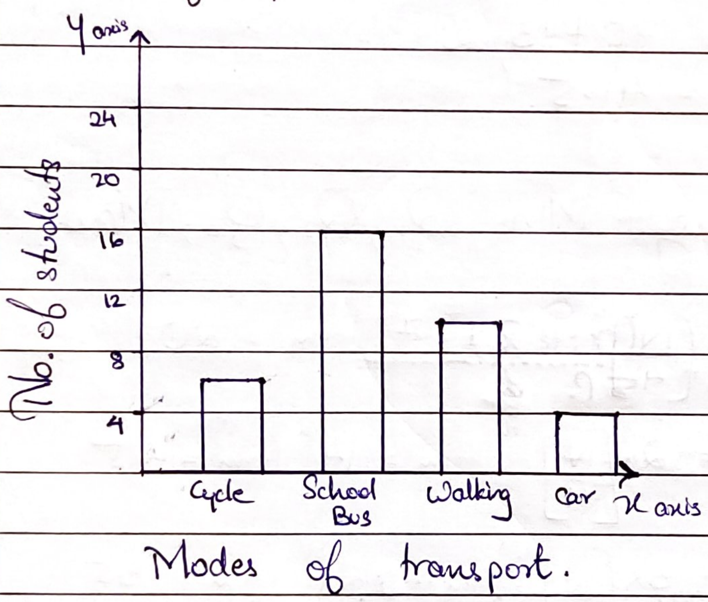
Statistics Assignment 1.

Q1. A survey of 36 students of a class was done to find out the mode of transport used by them while commuting to the school. The collected data is shown in the table given below. Represent the data in the form of a bar graph.

Mode of transport	Cycle	School Bus	Walking	Car
Number of students	6	16	10	4



Q2. Data analysis on the following attribute:
13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

a. Mean and Median : $n = 27$ $\therefore$ Mean $= \frac{\text{Sum}}{n} = \frac{809}{27} = 30.$

for Median, (n is odd) $\therefore$ $n+1/2^{th}$ element is median.
$\therefore$ 14^{th} element. Median = 25.

b. Mode and Data's modality:

The given dataset has two values that occur with the highest same frequency. Hence, the given data has a Dual modality.

Modes are 25 & 35, respectively.

c. Midrange : Average of largest & smallest values

$\therefore$ Midrange $= 70 + 13/2$
$= 41.5$.

d. Five number summary: Min, Q_1, Q_2, Q_3, Max.

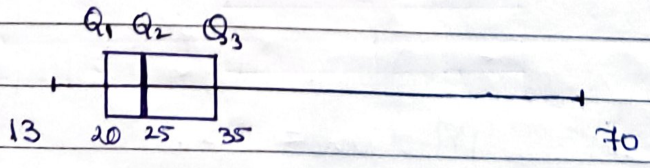
Min = 13, $Q_1 = \left[\frac{N}{4}\right]^{th}$ term = 7^{th} term = 20

$Q_2 = 2 \times \left[\frac{N}{4}\right]^{th}$ term = 14^{th} term = 25

Max = 70, $Q_3 = 3 \times \left[\frac{N}{4}\right]^{th}$ term = 21^{st} term = 35

$\therefore$ The five number summary of the given data is 13, 20, 25, 35, 70, respectively.

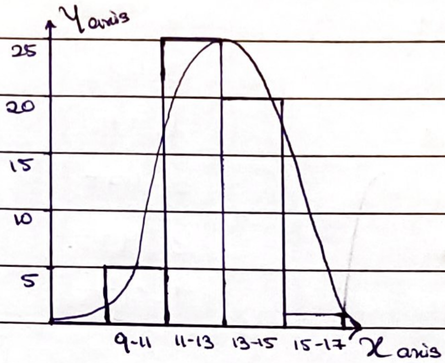
e. Box plot.



Q3. A pedestrian has obtained the following table which represents the number of children who begin to walk for the first time at different ages

Months	9	10	11	12	13	14	15
Children	1	4	9	16	11	8	1

a) Frequency : Plot



b) Mode = 12 months

$n = 50$

Median = 12 months

Mean = 610/50

= 12.2 months

$$\text{Variance} = \frac{\sum_{i=0}^{n} (x_i - \bar{x}) f}{n} = 1.68.$$

Months (x_i)	Children (f)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f$
9	1	-3.2	10.24	10.24
10	4	-2.2	4.84	19.36
11	9	-1.2	1.44	12.92
12	16	-0.2	0.04	0.64
13	11	0.8	0.64	7.04
14	8	1.8	3.24	25.92
15	1	2.8	7.84	7.84

Q4 Result of throwing two dice 120 times is given:

Sums	2	3	4	5	6	7	8	9	10	11	12
Frequency	3	8	9	11	20	19	16	13	11	6	4

i) Calculate mean, std. deviation. ii) draw box plot.

Sums (x_i)	Frequency (f_i)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
2	3	−5.025	25.2506	75.7518
3	8	−4.025	16.2006	129.6048
4	9	−3.025	9.1506	82.3554
5	11	−2.025	4.1006	45.1066
6	20	−1.025	1.0506	21.0120
7	19	−0.025	0.0006	0.0114
8	16	0.975	0.9506	15.2096
9	13	1.975	3.9006	50.7078
10	11	2.975	8.8506	97.3566
11	6	3.975	15.8006	94.8036
12	4	4.975	24.7506	99.0024

$\Sigma (x_i - \bar{x})^2 f_i = 710.9322$

$$\text{Mean} = \frac{\Sigma x_i f_i}{\Sigma f_i} = 7.025$$

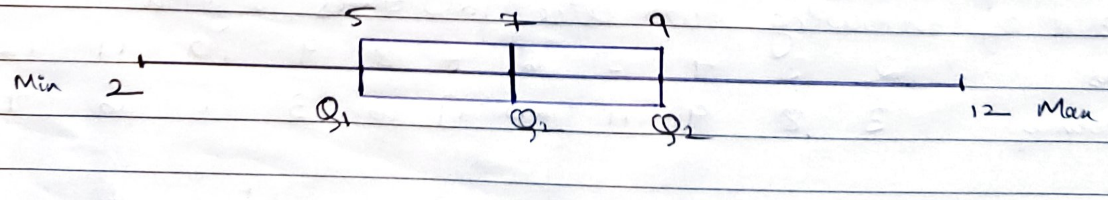
$$\text{Variance} = \frac{\sum (x_i - \bar{x})^2 f_i}{\Sigma f_i} = \frac{710.9322}{120}$$

$$\text{Variance} = 5.924435$$

$\therefore$ Standard Deviation $= 2.4340$.

Now, for Boxplot Min $= 2$, $Q_1 = \left(\frac{120}{4}\right)^{th}$ term $= 30^{th}$ term $= 5$

$Q_2 = (2 \times 30)^{th}$ term $= 60^{th}$ term $= 7$, $Q_3 = (3 \times 30)^{th}$ term $= 9$,
Max $= 12$.



Q5. Construct a frequency distribution table for the following weights (in gm) of 30 oranges using the equal class intervals, one of them is 40-45 (45 not included). The weights are :

31, 41, 46, 33, 44, 51, 56, 63, 71, 71, 62, 63, 54, 53, 51, 43, 36, 38, 54, 56, 66, 71, 74, 75, 46, 47, 59, 60, 61, 63.

Intervals	Frequency
30-35	2
35-40	2
40-45	3
45-50	3
50-55	5
55-60	3
60-65	6
65-70	1
70-75	5
Total =	30

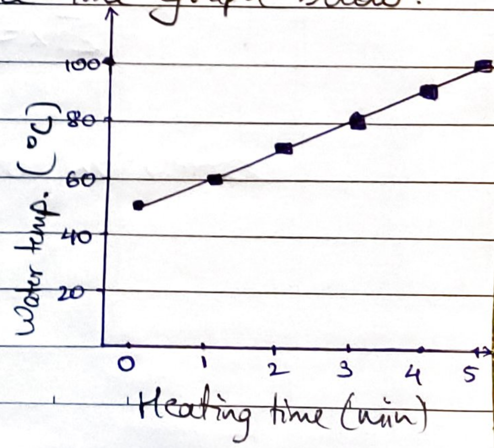
a) Class mark of class interval 50-55

$= \frac{50+55}{2} = 52.5$

b) For the given ~~range~~ weights, there are
c) 9 intervals ranging from 30 to 75.

d) Class interval of 65-70 has the lowest frequency of 1 unit.

Q6. A fifth grader conducted a 5 minute experiment that involved heating time and water temperature. The results of the experiment are represented in the line graph below:



Prediction from information gathered:

D) The water temperature increases as the heating time continues.

Q7) The weights of 10 people were recorded in kg as 35, 41, 42, 56, 58, 62, 70, 71, 90, 77. Find the percentile for the weight 58 kg.

→

Ascending order of data: 35, 41, 42, 56, 58, 62, 70, 71, 77, 90.

No of people with weight below 58 = 4.

∴ Percentile $= \frac{4}{10} \times 100 = 40\%$ile.

Q8) The scores for some candidates in a test are 40, 45, 49, 53, 61, 65, 71, 79, 85, 91. What will be the percentile for the score 71?

→

No of candidates who scored less than 71 = 6, Total = 10.

∴ Percentile $= \frac{6}{10} \times 100 = 60\%$ile.

Q9) Find the value of the correlation coefficient from given data.

→

Subject	Age (x)	Glucose level (y)	Σxy	x^2	y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
	$\Sigma x = 249$	$\Sigma y = 486$	$\Sigma xy = 20485$	$\Sigma x^2 = 11409$	$\Sigma y^2 = 40022$

Correlation coefficient $r = \dfrac{n(\Sigma xy) - (\Sigma x \Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$

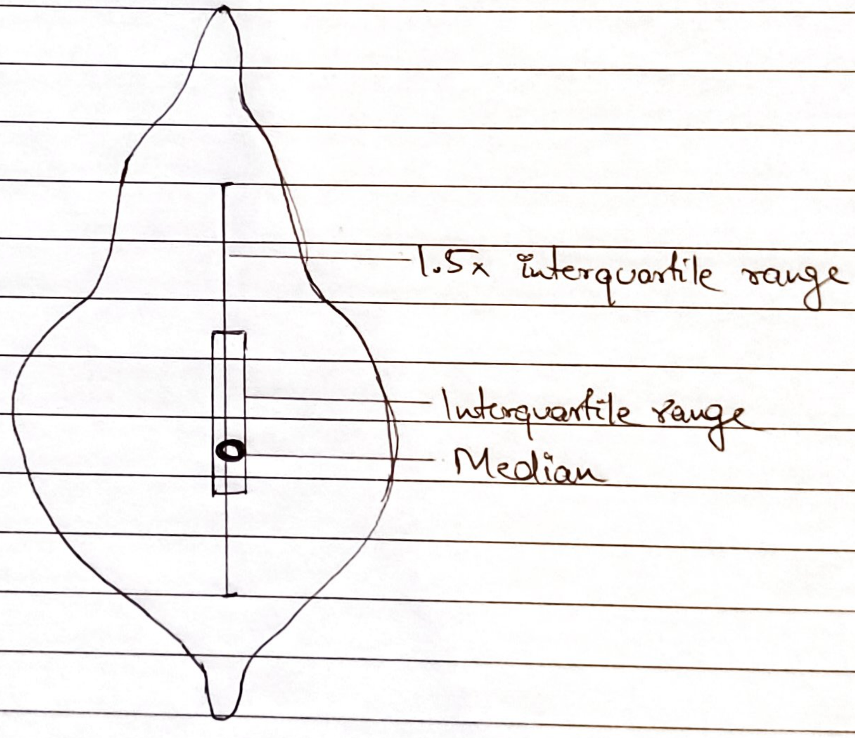
$\boxed{r = 0.5298}$

Sundaram®

FOR EDUCATIONAL USE

Q10) Violin plot
→ A violin plot is a hybrid of box plot and kernel density plot [a non-parametric method to estimate the probability density function of a random variable based on kernels as weights], which shows peaks in the data.

It is used to visualize the distribution of numeric data. Violin plot depicts the summary statistics and the density of each variable.



The thicker part depicts high probability and as on both ends it gets thinner, it displays lower probability (outliers)