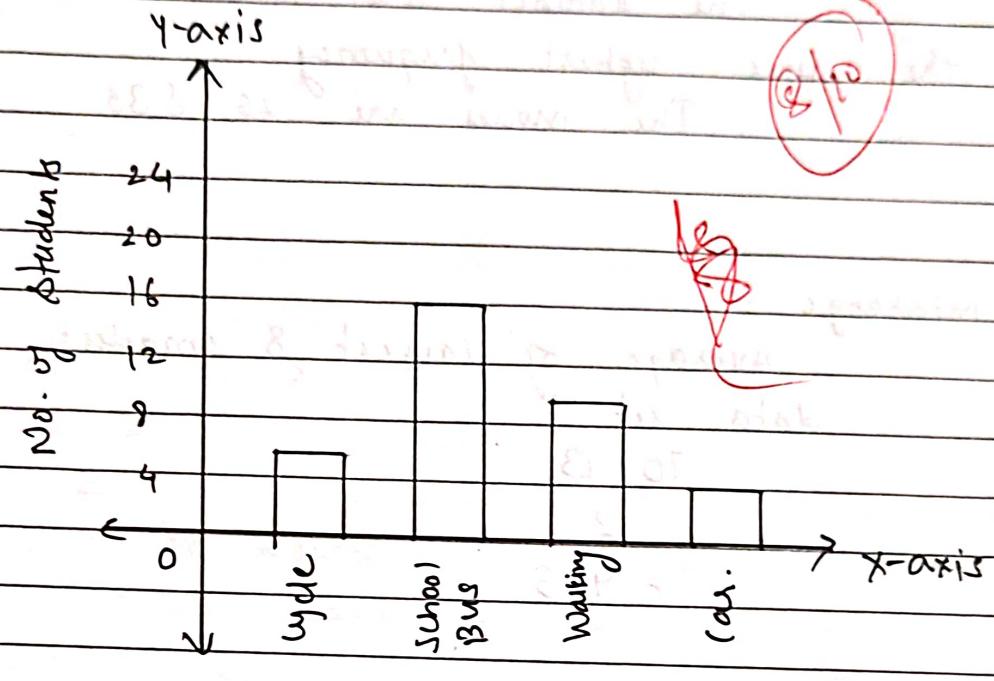
Page No.
Date

# Statistics in AI-DS

Manav Pahilwani
37

## Assignment 1.

Q1.




Q2.

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

1] Mean & median

$n = 27$

$\therefore$ Mean = $\frac{\text{sum}}{n}$

$= \frac{809}{27}$

$= 29.22 \approx 30$

$\therefore$ Mean = 30

Median = 25

2. Mode :-

The dataset has two values that occur with the same highest frequency

∴ The modes are 25 & 35.

3. Midsarge :-

Average of largest & smallest values in data set

$$\frac{70 + 13}{2}$$

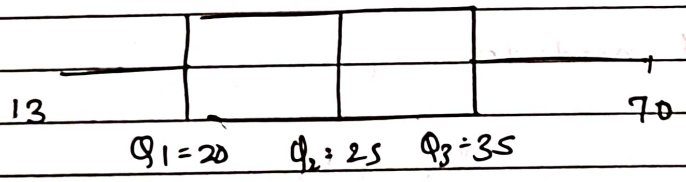
$$= 41.5$$

4. Five numbes summary

the $Q_1 = 20$

$Q_3 = 35$

∴

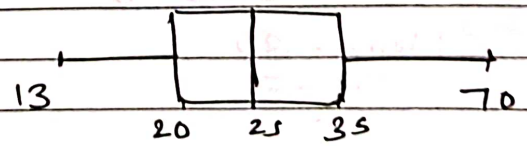



∴ The five numbes summary is

13, 20, 25, 35, 70.

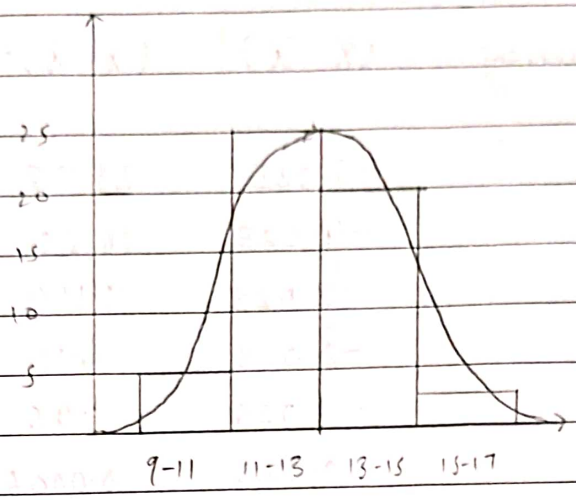
5.

$Q_2$ = median = 25




Teacher's Sign.: _____

Q 3.




mode → 12 months

median → 12 months

mean = $\frac{610}{50}$

= 12.2 months

| Months (x) | Children (f) | $(x-\bar{x})$ | $(x-\bar{x})^2$ | $(x-\bar{x})^2 f$ |
|---|---|---|---|---|
| 9 | 1 | -3.2 | 10.24 | 10.24 |
| 10 | 4 | -2.2 | 4.84 | 19.36 |
| 11 | 9 | -1.2 | 1.44 | 12.96 |
| 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 |

Variance = $\frac{\Sigma (x-\bar{x})^2 f}{n}$

= 1.68

Q 4.

| Sums (x) | frequency | $(x-\bar{x})$ | $(x-\bar{x})^2$ | $(x-\bar{x})^2 f$ |
|---|---|---|---|---|
| 2 | 3 | -5.025 | 25.25 | 75.75 |
| 3 | 8 | -4.025 | 16.20 | 129.6 |
| 4 | 9 | -3.025 | 9.150 | 82.35 |
| 5 | 11 | -2.025 | 4.10 | 45.1 |
| 6 | 20 | -1.025 | 1.05 | 21 |
| 7 | 19 | -0.025 | 0.0006 | 0.0114 |
| 8 | 16 | 0.975 | 0.95 | 15.2 |
| 9 | 13 | 1.975 | 3.9 | 50.7 |
| 10 | 11 | 2.975 | 8.85 | 97.35 |
| 11 | 6 | 3.975 | 15.80 | 94.8 |
| 12 | 4 | 4.975 | 24.75 | 99 |

mean = $(x) = 7.025$ variance : $\frac{\Sigma (x-\bar{x}) f}{n} = \frac{710.8614}{120}$

$Q_1 = \left(\frac{N+1}{4}\right)^{th}$ term $\therefore 30^{th}$ term $= 5.92$

$Q_1 = 5+6/2 = 5.5 \approx 5$

$Q_2 = (N+1/2)^{th}$ term : 7

$Q_3 = 3/4\,(N+1)^{th}$ terms

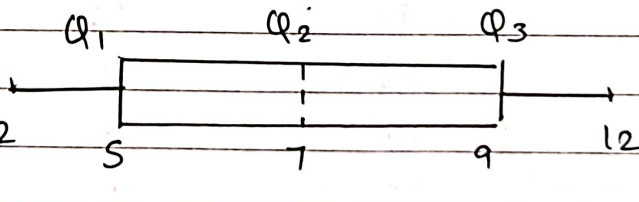
$= 3/4\,(121)^{th}$ term

$= 90.75^{th}$ term

$= 90 + 91/2$ term

$Q_3 = 9.$

IQR = 4.



Page No.
Date

Q 5.

| Interval | 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 | 4 |
| 75-80 | 1 |
| | Total → 30 |

a. Class mark of 50-55
is $\frac{55+50}{2}$
$= \underline{\underline{52.5}}$

b. Range $= 80-35 = 45$

c. There are 10 intervals

d. Class 65-70, has the lowest frequency.

75-80

Teacher's Sign.: _____

Page No.
Date

Q6. Option D, The water temperature increases as the heating time continues.

Q7. 35, 41, 42, 56, 58, 62, 70, 71, 90, 77

percentile for 58 kg

No. 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

∴ Percentile for weight 58 kg is 40.

Q8. 40, 45, 49, 53, 61, 65, 71, 79, 85, 91.

No. of candidates with marks below 71 = 6

∴ Percentile = $\frac{6}{10} \times 100$

= 60

∴ 71 score percentile is 60.

Teacher's Sign.: _____

Page No.
Date

Q9.

| Subject | Age (x) | Glucose lvl (y) | $\Sigma xy$ | $\Sigma x^2$ | $\Sigma y^2$ |
|---|---|---|---|---|---|
| 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 |
| | $\Sigma = 247$ | $\Sigma = 486$ | $\Sigma$ [20485 | 11409 | 40022] |

Correlation coefficient formula.

$$\frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

$$\frac{6(20485) - (247 \times 486)}{\sqrt{[6(11409) - (247)^2][6(40022) - (486)^2]}}$$

$$= \frac{2868}{5413.286}$$

$$= 0.5298$$

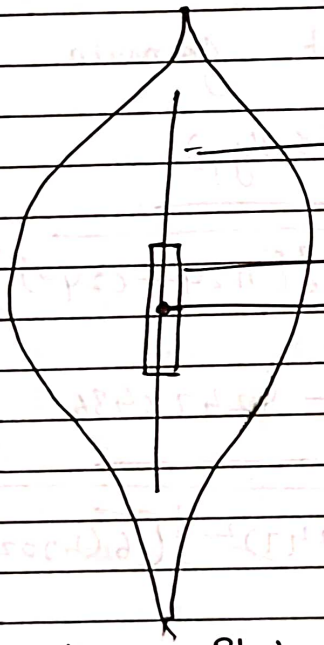
The correlation co-efficient is 0.5298.

Teacher's Sign.: _____

Page No.
Date

Q 10. Violin Plot.

A violin plot is a hybrid of box plot & 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 summary statistics & the density of each variable




Violin Plot.

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

Teacher's Sign.: _____