1]

Stem	leaf
1	3 8 9
2	1 4 7 8 9
3	0 1 2 3 4 5 8 8 9
4	0 2 4 5 6 7 7
5	7

key: 5|7 means 57

a] Calculate mean, median, and mode

b] Calculate 25th percentile value
75th percentile value

c] Draw box and whisker plot

d] Detect outliers.

Solutions:

Mean: $\bar{X} = \dfrac{13 + 18 + 19 + \cdots + 57}{25} = \dfrac{857}{25} = 34.28$

Median: Here, $n = 25$, an odd number.

$\therefore$ Median $= \left(\dfrac{n+1}{2}\right)^{th}$ observation

$= \left(\dfrac{25+1}{2}\right)^{th}$ obs.

$= 13$ th obs.

$= 34$

Mode: The most frequent obs.   38 and 37.

$\boxed{Q_1 \text{ or } P_{25}}$

25th percentile:

$$\text{Position of } Q_1 = \frac{1 \times 25}{4}$$

$$= 6.25$$

$\left[P_{25} = \dfrac{25 \times 25}{100} = 6.25 \quad \text{same} \right]$

Since, the position of Q_1 is not integer, we go for next integer value.

$$\therefore \quad Q_1 = 7\text{th} \quad \text{obs.}$$

$$= 28$$

Ans:

$\boxed{Q_3 \text{ or } P_{75}}$

75th percentile:

$$\text{Position of } Q_3 = \frac{3 \times 25}{4}$$

$$= 18.75$$

Since the position of Q_3 is not integer,

$$\therefore \quad Q_3 = 19\text{th} \quad \text{obs.}$$

$$= 42$$

$*$ Interpretation must $*$

Box and whisker plot:

For box-and-whisker plot, we need,

Minimum = 13

Q_1 = 28

Q_2/Median/50th percentile = 34

Q_3 = 42

Max = 57
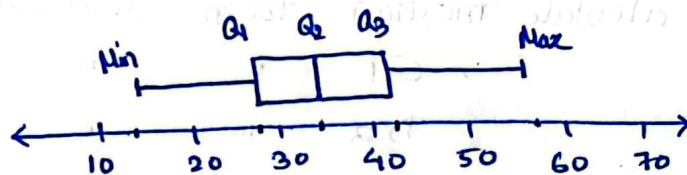
Outliers

$$IQR = Q_3 - Q_1$$
$$= 14$$

$$Lower\ fences = Q_1 - 1.5\ IQR$$
$$= 28 - (1.5 \times 14)$$
$$= 7$$

$$Upper\ fences = Q_3 + 1.5\ IQR$$
$$= 42 + (1.5 \times 14)$$
$$= 63$$

Since, there is no value outside of these inner fences range.

So, there are no outlier in this data set.

21 Suppose, you are making a table for 15 observations with four classes: (5-9); (9-13); (13-17); and (17-21). The percentage frequencies for first class is 20% and last class is 20%.
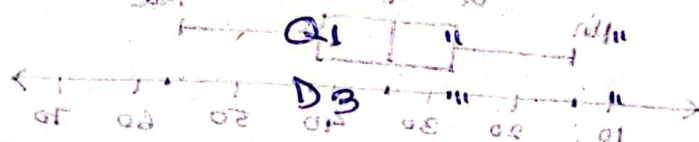
a] Calculate the frequency of middle two classes, when average is 12·6

Median from Grouped data → **b]** Calculate median/Q_2/50th percentile value.

Median from ogive curve ← **c]** Calculate median from cumulative freq. curve

$$\longmapsto \boxed{Q_1} \quad " \quad \overset{'''}{} \quad \longmapsto$$

$$\longleftarrow \quad D3 \quad ''' \quad \longrightarrow$$

Solution:

Here, N = 15

Relative freq × N = freq.

Percent. freq × N / 100 = freq.

class	Percent	freq(f_i)	(x_i) → Mid value	$f_i x_i$
5-9	20%	3	7	21
9-13		f_1	11	$11f_1$
13-17		f_2	15	$15f_2$
17-21		3	19	57

Now,

$$\overline{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

$$\Rightarrow 12\cdot6 = \frac{21 + 11f_1 + 15f_2 + 57}{15} \qquad [\Sigma f_i = N = 15]$$

$$\Rightarrow 189 = 78 + 11f_1 + 15f_2$$

$$\Rightarrow 11f_1 + 15f_2 = 111 \quad \underline{\hspace{2cm}} (I)$$

Again

$$\Sigma f_i = 15$$

$$\Rightarrow 6 + f_1 + f_2 = 15$$

$$\Rightarrow f_1 + f_2 = 9 \quad \underline{\hspace{2cm}} (II)$$

$$\therefore f_1 = ??? \qquad \therefore f_2 = ???$$

CamScanner

3] Data set A: 15, 12, 18, 10, 16

Data set B: 11, 17, 14, 20, 13

a) Which data set shows more consistency?

b) If the last value in data set B is corrected to 10, compute the corrected standard deviation.

↳ Hints: Assignment Question.

Best measures of dispersion = Standard deviation

Over the years geometric growth / decrease = Geometric mean

Distance vs speed = Harmonic mean

For comparison = Coefficient of variance

Position of Q_i, D_i, or P_i is not integer = Go to next integer value

$$\begin{bmatrix} 6.75 & = 7\text{th obs.} \\ 8.25 & = 9\text{th obs.} \end{bmatrix}$$

Position of Q_i, D_i or P_i is integer = take average.

$$\begin{bmatrix} Q_i = 6 & . \text{ Then, } \dfrac{6\text{th obs.} + 7\text{th obs.}}{2} \\ Q_i = 2 & ; \text{ Then, } \dfrac{2\text{nd obs} + 3\text{rd obs.}}{2} \end{bmatrix}$$

Lowest 25% = Q_1 or P_{25}

Highest 25% = Q_3 or P_{75}

Lowest 30% = P_3 or P_{30}

Highest 30% = D_7 or P_{70}

4) Sales Revenue: 7, 8, 9, 11, 12, 14, 15

Advertising cost: 32, 35, 39, 44, 50, 52, 55

Here sales revenue influenced by the ad. cost.

     a] Determine the Direction and strength of relationship.

     b] Determine the fitted regression model

     c] Estimate sales revenue when ad. cost is 60.

     d] Estimate ad. cost when sales rev. is 10.

     e] How precise your prediction.

     f] If the last ad. value is corrected to 50, Calculate the corrected regression parameter.

Solution :

     Here, sales rev. influence by ad. cost.

     So, Sales revenue is Y

     Advertising cost is X.

Table 1:

Sales Revenue (y_i)	Ad. cost (x_i)	x_i^2	y_i^2	$x_i y_i$
7	32	1024	49	224
8	35	1225	64	280
9	39	1521	81	351
11	44	1936	121	484
12	50	2500	144	600
14	52	2704	196	728
15	55	3025	225	825
$\sum y_i = 76$	$\sum x_i = 307$	$\sum x_i^2 = 13935$	$\sum y_i^2 = 880$	$\sum x_i y_i = 3492$

a) From table 1,

$$\Sigma x_i = 307, \qquad \Sigma y_i^2 = 880$$
$$\Sigma y_i = 76, \qquad \Sigma x_i y_i = 3492$$
$$\Sigma x_i^2 = 13935,$$

$$\therefore \; r = \frac{\Sigma x_i y_i - \dfrac{\Sigma x_i \, \Sigma y_i}{n}}{\sqrt{\Sigma x_i^2 - \dfrac{(\Sigma x_i)^2}{n}} \; \sqrt{\Sigma y_i^2 - \dfrac{(\Sigma y_i)^2}{n}}} = ???$$

b) from table 1,

$$\vdots$$

$$\therefore \; \hat{\beta} = \frac{n\Sigma x_i y_i - \Sigma x_i \, \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{(7 \times 3492) - (307 \times 76)}{(7 \times 13935) - (307)^2} = 0.34$$

$$\therefore \; \hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x} = -4.05$$

$$\therefore \; \hat{y}_i = -4.05 + 0.34\, x_i$$

c) $x = 60 \quad ; \quad \hat{y}_i = ??$

d) $y_i = 10 \quad ; \quad x = ???$

e) $R^2 = 1 - \dfrac{SSE}{SST} \quad ; \quad [\text{ref: lecture slide}]$

f)

⑪

Incorrected part :

$$\Sigma x_i = 307$$

$$\Sigma y_i = 76$$

$$\Sigma x_i^2 = 13935$$

$$\Sigma y_i^2 = 880$$

$$\Sigma x_i y_i = 3492$$

Corrected part :

$$\Sigma x_i = 307 - 55 + 50 = \quad ???$$

$$\Sigma y_i = 76$$

$$\Sigma x_i^2 = 13935 - 55^2 + 50^2 = \quad ???$$

$$\Sigma y_i^2 = 880$$

$$\Sigma x_i y_i = 3492 - (15 \times 55) + (15 \times 50) = ???$$

$$\hat{\beta}_{corrected} =$$

$$\hat{\alpha}_{corrected} =$$

\# Strength of association = Karl pearson.

\# Direction of association = Scatter plot

\# Direction of and Strength = Karl pearson

\# y on x = Before "on" — Dependent variable

After "on" — Independent variable