Assignment 2

TAN XIN TIAN A24CS0198

ANGELA NGU XIN YI A24CS0226

1.

Class Interval	MP, x_i	Frequency, f_i	Cf	$f_i x_i$
$150 \leq x < 160$	155	12	12	1860
$160 \leq x < 170$	165	20	32	3300
$170 \leq x < 180$	175	5	37	875
$180 \leq x < 190$	185	3	40	555
	Total :	40		6590

(a) Mean $= \frac{f_i x_i}{f_i} = \frac{6590}{40} = 164.75$

(b) Median

$N \div 2 = 20$

Median Class : $(160 \leq x < 170)$ cm

$$\text{Median} = L + \frac{\frac{N}{2} - cf_p}{f_{med}}(W)$$

$$= 160 + \frac{20 - 12}{20}(170 - 160)$$

$$= 164$$

(c) Mode

Mode Class = 159.5 – 169.5

$$\text{Mode} = L + h\left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right)$$

$$= 160 + (10)\left(\frac{20 - 12}{2(20) - 12 - 5}\right)$$

$$= 163.4783$$

$$= 163.48 \approx 163 \text{ cm}$$

(d) Modal Class : $(160 \leq x < 170)$ cm

2. (a) Mean $= \frac{85 + 90 + 75 + 88 + 92 + 80 + 85 + 82 + 90 + 95}{10}$

$= 85.2$

75, 80, 82, 85, 85, 85, 88, 90, 90, 92

Median $= \frac{85 + 85}{2}$

$= 85$

Mode = 85

(b) Mean : gives the average participation (85.2), showing overall engagement

Median : reflects the middle score is 85, so it's unaffected by outliers

Mode : 85 is the most frequent score (many students scored this)

Interpretation :

- These value very close to each other, indicating a consistent and fairly high level of participation among most members.
- There's no extreme score, group performance is balanced

Most appropriate statistic :

mean (85.2) better represents the overall performance because there are no outliers in data

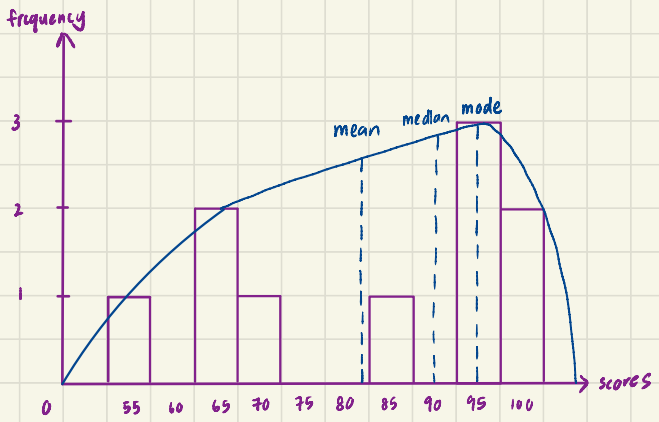
(c)(i) Mean $= \frac{55 + 65 + 65 + 70 + 85 + 95 + 95 + 95 + 100 + 100}{10}$

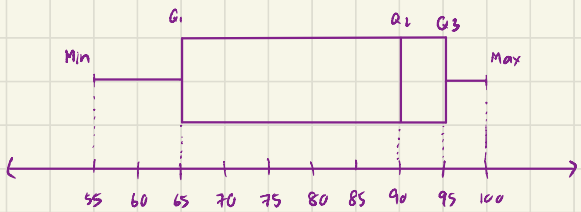
$= 82.5$

Median $= \frac{85 + 95}{2} = 90$

Mode = 95

(ii)





Q_1: $i = \frac{25}{100}(10) = 2.5 \approx 3$

$Y[3] = 65$

$Q_2 = 90$

Q_3: $i = \frac{75}{100}(10) = 7.5 \approx 8$

$Y[8] = 95$

$IQR = Q_3 - Q_1$

$= 30$

[lower Limit, Upper Limit]

$= [Q_1 - 1.5(IQR), Q_3 + 1.5(IQR)]$

$= [65 - 1.5(30), 95 + 1.5(30)]$

$= [20, 140]$

↳ Means no have outliers

(iii)

	Original	Corrected
Mean	85.2	82.5
Median	85	90
Mode	85	95

- The mean dropped from 85.2 to 82.5, which indicates more low-scoring participation in the corrected data.
- The median increased from 85 to 90, showing that at least half of the students participating at higher level.
- The mode changed from 85 to 95 suggesting more students in the corrected data are consistently scoring high.

3. (a)(i) $R = 40 - 25 = 15$

$$\text{Mean} = \frac{25 + 30 + 28 + 35 + 32 + 27 + 40 + 38 + 33 + 36 + 31 + 29}{12}$$

$$= 32 \ (\text{RM } 32\text{k})$$

(ii) Variance, $\sigma^2 = \frac{(25-32)^2 + (30-32)^2 + (28-32)^2 + (35-32)^2 + (32-32)^2 + (27-32)^2 + (40-32)^2 + (38-32)^2 + (38-32)^2 + (33-32)^2 + (36-32)^2 + (31-32)^2 + (29-32)^2}{12}$

$$= 19.1667 \quad (\text{RM } 19.17\text{k})$$

(iii) Standard deviation, $\sigma = \sqrt{19.1667}$

$$= 4.3780 \quad (\text{RM } 4.378\text{ k})$$

(b) The range of RM15k shows the spread of data which is difference between highest sales (RM40k) and lowest sales (RM25k). This shows that sales figures had some notable fluctations. The variance of RM19.1667k quantifies the dispersion of sales data points from mean in square units.

Standard deviation, which is RM 4.3780k reflets the average amount of which the monthly sales figures deviate from the mean. This moderate standard deviation shows that sales figures are not extremely inconsistent

monthly sales is
The variability in moderate, showing that the business has a stable monthly income. Although some months perform slightly better or worse than others, there is no alarming swing. This shows a predictable cash flow trends.

(c) It helps financial planning as the business can anticipate and plan for months with lower sales. Moreover, this understanding helps in inventory and staffing task also. For example, months with higher sales need more stock and staff while cost-cutting and some marketing strategies would be needed for lower months. Additionally, the standard deviation can help to create realistic monthly budgets and detect seasonal trends, which in case improving operational efficiency.

Question 4

a) $\bar{x} = 50$ $\quad\quad$ $\sigma = 10$

50% of employees showed an increase in productivity after the training program. For this reason, the mean is 50 and the scores are normally distributed so the increase can be seen as scores above the mean. So 50% of the values lie above the mean.

$P(X \geq 50) = P(Z \geq \frac{50-50}{10})$

$= P(Z \geq 0)$

$= 1 - P(Z \leq 0)$

$= 1 - 0.5$

$= 0.5$

b) $P(37 < X < 65)$

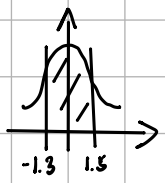
$= P(\frac{37-50}{10} < Z < \frac{65-50}{10})$

$= P(-1.3 < Z < 1.5)$

$= P(Z < 1.5) - P(Z < -1.3)$

$= 0.93319 - 0.0968$

$= 0.83639$

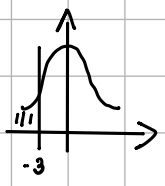


c) $P(X < 20)$

$= P(Z < \frac{20-50}{10})$

$= P(Z < -3)$

$= 0.00135$



$0.00135 \times 1000 = 1.35$

≈ 1 employee

Cost $= 1 \times 200$

$=$ RM200

d) $P(Z > x) = 0.05$

$P(Z > \frac{X-50}{10}) = 0.05$

$1 - P(Z \leq \frac{X-50}{10}) = 0.05$

$P(Z \leq \frac{X-50}{10}) = 0.95$

$\frac{X-50}{10} = 1.64$

$X = 66.4$

Question 5

a) $X \sim B(6, 0.25)$ X = number of correct answers in 6 MCQ

b)

X	Probabilities
0	$\binom{6}{0}(0.25)^0(0.75)^6 = 0.1780$
1	$\binom{6}{1}(0.25)^1(0.75)^5 = 0.3560$
2	$\binom{6}{2}(0.25)^2(0.75)^4 = 0.2966$
3	$\binom{6}{3}(0.25)^3(0.75)^3 = 0.1318$
4	$\binom{6}{4}(0.25)^4(0.75)^2 = 0.0330$
5	$\binom{6}{5}(0.25)^5(0.75)^1 = 4.3945 \times 10^{-3}$
6	$\binom{6}{6}(0.25)^6(0.75)^0 = 2.4414 \times 10^{-4}$

c) Mean $= np$

$= 6(0.25)$

$= 1.5$

d) $P(X \geq 3) = 1 - P(X=0) - P(X=1) - P(X=2)$

$= 1 - 0.1780 - 0.3560 - 0.2966$

$= 0.1694$

e) $P(X=4) = (0.25)^3(0.75)$

$= 0.0117$

Question 6

a) Negative binomial distribution $X \sim NB(4, 0.7)$

b) $\sigma = \sqrt{\frac{nq}{p^2}}$

$= \sqrt{\frac{4 \times (1-0.7)}{0.7^2}}$

$= 1.5649$

Geometric

$\mu = \frac{1}{p}$ (first success)

nth success $= n(\frac{1}{p})$

$\sigma = \sqrt{\frac{1-p}{p^2}}$ (first success)

n^{th} success $= \sigma = \sqrt{\frac{n(1-p)}{p^2}}$

c) $P(X=6) = \binom{6-1}{4-1}(0.7)^4(0.3)^2$

$= 0.2161$

d) $P(X=7) = \binom{12}{7}(0.7)^7(0.3)^5$

$= 0.1585$