DATA ANALYSIS EXAM QUESTIONS

Question 1 (**)

The number of phone text messages send by 11 different students is given below.

$$14,\ 25,\ 31,\ 36,\ 37,\ 41,\ 51,\ 52,\ 55,\ 79,\ 112.$$

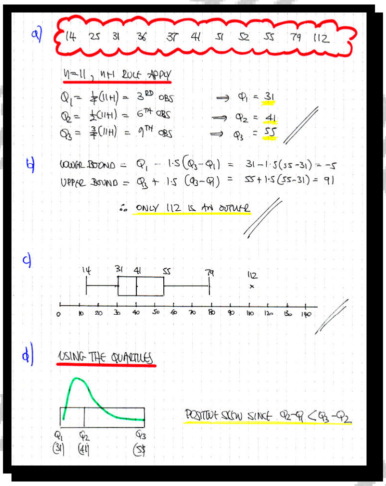
a) Find the lower quartile, the median and the upper quartile of the data.

b) Show clearly that there is only one outlier in the data.

c) Draw a suitably labelled box plot for this data, clearly indicating any outliers.

d) Determine with justification the skewness of the data.

$\boxed{Q_1 = 31}$, $\boxed{Q_2 = 41}$, $\boxed{Q_3 = 55}$, $\boxed{\text{112 is the only outlier}}$, $\boxed{\text{positive skew}}$



a) 14 25 31 36 37 41 51 52 55 79 112

N=11, N+1 RULE APPLY

$Q_1 = \frac{1}{4}(11+1) = 3^{RD}$ OBS $\Rightarrow Q_1 = 31$

$Q_2 = \frac{1}{2}(11+1) = 6^{TH}$ OBS $\Rightarrow Q_2 = 41$

$Q_3 = \frac{3}{4}(11+1) = 9^{TH}$ OBS $\Rightarrow Q_3 = 55$

b) LOWER BOUND $= Q_1 - 1.5(Q_3 - Q_1) = 31 - 1.5(55-31) = -5$

UPPER BOUND $= Q_3 + 1.5(Q_3 - Q_1) = 55 + 1.5(55-31) = 91$

$\therefore$ ONLY 112 IS AN OUTLIER

c)

d) USING THE QUARTILES

POSITIVE SKEW SINCE $Q_2 - Q_1 < Q_3 - Q_2$

Question 2 (**)

The number of bottles of red wine sold by a local supermarket over a two week period is shown below.

22, 14, 11, 33, 32, 45, 4, 12, 13, 20, 27, 44, 30, 15.

a) Display the above data in an ordered stem and leaf diagram.

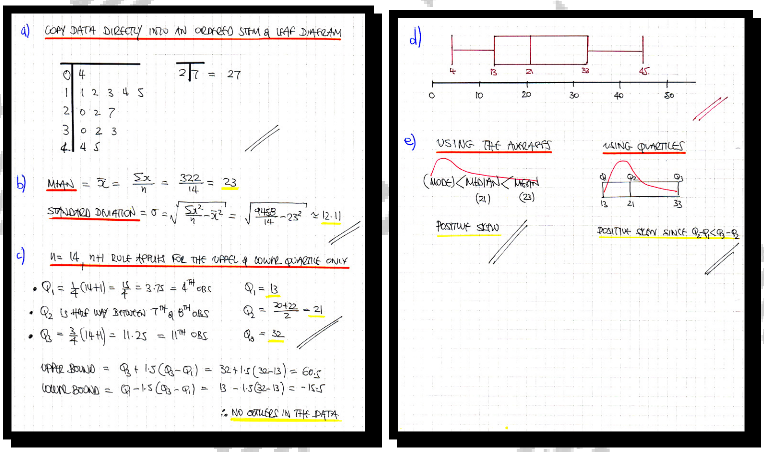
b) Calculate the mean and the standard deviation of the data.

c) Find the median and the quartiles of the data and use them to determine if there are any outliers.

d) Draw a suitably labelled box plot for this data.

e) Determine with justification the skewness of the data.

$\boxed{}$, $\boxed{\bar{x} = 23}$, $\boxed{\sigma = 12.11}$, $\boxed{Q_1 = 13}$, $\boxed{Q_2 = 21}$, $\boxed{Q_3 = 33}$, $\boxed{\text{no outliers}}$, $\boxed{\text{positive skew}}$



Question 3 (**+)

The concentration of lactic acid, in appropriate units, after a period of intense exercise was measured in the blood of 12 marathon runners.

Athlete	A	B	C	D	E	F	G	H	I	J	K	L
Lactic Acid Concentration	180	172	110	175	256	140	241	450	205	375	402	195

a) Find the mean and the standard deviation of the data.

b) Determine the value of the median and the quartiles.

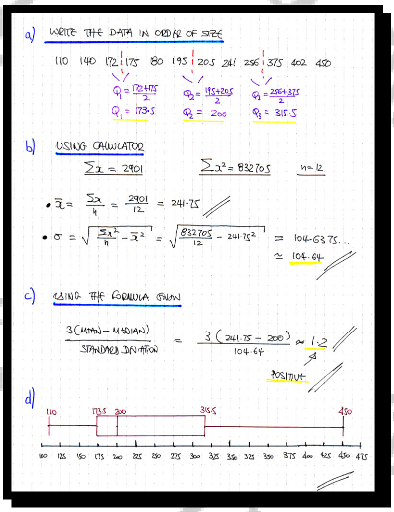
The skewness of data can be determined by the formula

$$\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}.$$

c) Evaluate this expression for this data and hence state its skew.

d) Draw a suitably labelled box plot for this data.
You may assume that there are no outliers in this data.

$\boxed{}$, $\boxed{\bar{x} = 241.75}$, $\boxed{\sigma \approx 104.64}$, $\boxed{Q_1 = 173.5}$, $\boxed{Q_2 = 200}$, $\boxed{Q_3 = 315.5}$, $\boxed{1.20}$, $\boxed{\text{positive skew}}$



Question 4 (**+)

The % marks, rounded to the nearest integer, of a recent Mathematics test taken by 16 students, were summarised in an ordered stem and leaf diagram.

4	7
5	2,3,8
6	0,3,4, a, b
7	3,6,c,d,8
8	1,9

where $5 \| 2 = 52$.

a) Determine the lower quartile of the data.

b) Given the median is 68 and $a \neq b$, find the value of a and the value of b.

It is further given that $c \neq d$.

c) Find the possible values of the upper quartile.

$\boxed{}$, $\boxed{Q_1 = 59}$, $\boxed{a = 7, \ b = 9}$, $\boxed{76.5, 77, 77.5}$

a)

4	7	(1)
5	2 3 8	(3)
6	0 3 4 a b	(5)
7	3 6 c d 8	(5)
8	1 9	(2)

TOTAL OF 16 OBSERVATIONS

Q_1 IS THE $4^{TH}/5^{TH}$ OBS

$\therefore Q_1 = \frac{58+60}{2} = 59$

b) MEDIAN IS $8^{TH}/9^{TH}$ OBSERVATION

IE $Q_2 = \frac{\text{"}6a\text{"} + \text{"}6b\text{"}}{2} = 68$

$\therefore a = 7 \ \& \ b = 9$

$\cancel{68} - \cancel{68}$ $a \neq b$

$67 - 69$

$\cancel{66} - \cancel{70}$

ETC

c) Q_3 IS $12^{TH}/13^{TH}$ OBSERVATION

IE $Q_3 = \frac{\text{"}7c\text{"} + \text{"}7d\text{"}}{2}$

$\cancel{76} - \cancel{76}$ $c \neq d$

$76 - 77 \Rightarrow Q_3 = 76.5$

$76 - 78 \Rightarrow Q_3 = 77$

$\cancel{77} - \cancel{77}$ $c \neq d$

$77 - 78 \Rightarrow Q_3 = 77.5$

$\therefore$ POSSIBLE VALUES OF Q_3 ARE 76.5, 77 & 77.5

Question 5 (+)**

A company decides to give their 23 employees a skills test in order to decide if any of these employees need to be retrained.

The maximum possible score in this test is 50 and the results are summarised in an ordered stem and leaf diagram.

0	5
1	9,9
2	1,6,8
3	3,4,5,7
4	2,3,4,4,8,9,9
5	0,0,0,0,0,0

where $2\|9 = 29$.

a) Find the median score of the test.

b) Determine the interquartile range of the scores.

The company decides to retrain any employee whose score is less than the **lower quartile minus the interquartile range**.

c) Show clearly that only one employee will undergo retraining.

d) Draw a suitably labelled box plot for this data, clearly indicating any outliers, as found in part (**c**).

e) Determine with justification the skewness of the scores.

$\boxed{}$, $\boxed{Q_2 = 43}$, $\boxed{IQR = 22}$, $\boxed{05 \text{ is the only outlier}}$, $\boxed{\text{negative skew}}$

a) LOOKING AT THE STEM & LEAF DIAGRAM

0	5	(1)
1	9 9	(2)
2	1 6 8	(3)
3	3 4 5 7	(4)
4	2 3 4 4 8 9 9	(7)
5	0 0 0 0 0	(5)

TOTAL OF 23 OBSERVATION

$Q_2 = \frac{1}{2}(23+1) = 12^{TH}$ OBS

$\therefore Q_2 = 43$

b) FIND THE QUARTILES

$Q_1 = \frac{1}{4}(23+1) = 6^{TH}$ OBS $\quad \therefore Q_1 = 28$

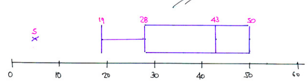
$Q_3 = \frac{3}{4}(23+1) = 18^{TH}$ OBS $\quad \therefore Q_3 = 50$

$\therefore IQR = Q_3 - Q_1 = 50 - 28 = 22$

c) LOWER BOUND $= Q_1 - IQR = 28 - 22 = 6$

$\therefore$ ONLY ONE EMPLOYEE WITH A SCORE OF 5 WILL UNDERGO RETRAINING

d)



e) LOOKING AT THE BOX PLOT

$Q_3 - Q_2 < Q_2 - Q_1 \Rightarrow$ NEGATIVE SKEW

Question 6 (**+)

The following set of data shows the number of posts made, in a given day, in a social media site by a group of individuals.

$$1,\ 12,\ 13,\ 14,\ 16,\ 17,\ 20,\ 21,\ 23,\ 24,\ 26,\ 39,\ 55.$$

For this set of data, ...

a) ... determine the value of the median and the quartiles.

b) ... calculate the mean and the standard deviation.

c) ... determine with justification whether there are any outliers.

d) ... state with justification if there is any type of skew.

$\boxed{}$, $\boxed{(Q_1, Q_2, Q_3) = (14, 20, 26) \ \text{ or } \ (Q_1, Q_2, Q_3) = (13.5, 20, 25)}$, $\boxed{\bar{x} \approx 21.6}$, $\boxed{\sigma \approx 12.9}$, $\boxed{\text{55 is an outlier}}$, $\boxed{\text{no skew or positive skew depending on the method}}$

1, 12, 13, 14, 16, 17, 20, 21, 23, 24, 26, 39, 55

$n = 13$ (ODD)

a) MEDIAN & QUARTILES (n+1) RULE APPLIES

$Q_1 = \frac{1}{4}(13+1) = 3.5$ I.E $3^{RD}/4^{TH}$ OR 4^{TH}

$Q_2 = \frac{1}{2}(13+1) = 7^{th}$ OBS

$Q_3 = \frac{3}{4}(13+1) = 10.5$, I.E $10^{TH}/11^{TH}$ OR 11^{TH} OBS

$\therefore$ $Q_1 = 13.5$, $Q_2 = 20$, $Q_3 = 25$

OR

$Q_1 = 14$, $Q_2 = 20$, $Q_3 = 26$

b) USING CALCULATOR IN STAT MODE WE OBTAIN

$\sum x = 281$, $\sum x^2 = 8223$

- $\bar{x} = \frac{\sum x}{n} = \frac{281}{13} \simeq 21.6$
- $\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{8223}{13} - \left(\frac{281}{13}\right)^2} \approx 12.9$

c) USING THE MOST COMMON CRITERION

- LOWER BOUND $= Q_1 - \frac{3}{2}(IQR) = 14 - \frac{3}{2}(26-14) = -4$
- UPPER BOUND $= Q_3 + \frac{3}{2}(IQR) = 26 + \frac{3}{2}(26-14) = 44$

$\therefore$ 55 IS A OUTLIER

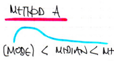
ALTERNATIVE METHOD FOR OUTLIERS

LOWER BOUND $= \bar{x} - 2\sigma = 21.6 - 2 \times 12.9 \simeq -4$

UPPER BOUND $= \bar{x} + 2\sigma = 21.6 + 2 \times 12.9 \simeq 47$

$\therefore$ 55 IS AN OUTLIER

d) METHOD A

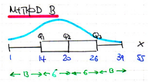


(MODE) < MEDIAN < MEAN

20 21.6

$\therefore$ POSITIVE SKEW

METHOD B



$Q_2 - Q_1 = Q_3 - Q_2$

$\therefore$ SYMMETRICAL (NO SKEW)

THE TWO METHODS HERE DISAGREE (SOMETIMES THIS HAPPENS)
THIS IS DUE TO THE OUTLIER AT 55.

IF THE OUTLIER IS CONSIDERED THERE IS SMALL POSITIVE SKEW; NOTE THAT MEAN & MEDIAN ARE STILL CLOSE.

IF THE OUTLIER IS NOT CONSIDERED THE DISTRIBUTION IS VERY SYMMETRICAL

Question 7 (***)

A farmer keeps chicken and sells most of the eggs they lay.

The table below summarizes information about the number of eggs laid by his chickens every week, for a period of 47 weeks.

Total number of eggs laid in a week	Number of weeks
52	1
53	4
54	7
55	10
56	11
57	8
58	5
59	1

a) Calculate the mean and the standard deviation of the eggs laid per week.

b) Determine the median and the quartiles for these data.

c) If the farmer only sells 45 eggs per week and keeps the rest for his family, find the mean and the standard deviation of the eggs he keeps for his family.

d) Use the median and mean to determine the skew of the above data, and hence determine whether this data can be modelled by a Normal distribution.

$\boxed{}$, $\boxed{\bar{x} \approx 55.6}$, $\boxed{\sigma_x \approx 1.59}$, $\boxed{Q_1 = 54}$, $\boxed{Q_2 = 56}$, $\boxed{Q_3 = 57}$, $\boxed{\bar{y} \approx 10.6}$, $\boxed{\sigma_y \approx 1.59}$

EGGS LAID IN A WEEK — NUMBER OF WEEKS

$52 \longrightarrow 1 \ (1)$
$53 \longrightarrow 4 \ (5)$
$54 \longrightarrow 7 \ (12)$
$55 \longrightarrow 10 \ (22)$
$56 \longrightarrow 11 \ (33)$
$57 \longrightarrow 8 \ (41)$
$58 \longrightarrow 5 \ (46)$
$59 \longrightarrow 1 \ (47)$

a) FROM CALCULATOR IN STAT MODE

$\Sigma x = 2613 \qquad \Sigma x^2 = 145391 \qquad n = 47$

MEAN, $\bar{x} = \frac{\Sigma x}{n} = \frac{2613}{47} \approx 55.6$

S.D, $\sigma = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2} = \sqrt{\frac{145391}{47} - 55.6...^2} \approx 1.59$

b) $n = 47$ (ODD, SO $n+1$ RULE APPLIES)

- $Q_1 = \frac{1}{4}(47+1) = 12^{TH}$ OBS
- $Q_2 = \frac{1}{2}(47+1) = 24^{TH}$ OBS
- $Q_3 = \frac{3}{4}(47+1) = 36^{TH}$ OBS

$\therefore Q_1 = 54$, $Q_2 = 56$, $Q_3 = 57$

c) THIS IS CODING $y = x - 45$

- MEAN, $\bar{y} = \bar{x} - 45 = 10.6$
- S.D, $\sigma_y = \sigma_x = 1.59$ (UNCHANGED)

d)

$\left.\begin{array}{l}\text{MEDIAN} = 54 \\ \text{MEAN} = 55.6 \\ \text{MODE} = 56\end{array}\right\}$ APPROXIMATELY EQUAL, SO VERY LITTLE SKEW

DATA CANNOT BE MODELLED BY A NORMAL DISTRIBUTION AS THE DATA IS DISCRETE AND NOT GROUPED

Question 8 **(***)**

The number of hours worked in a given week by a group of 64 individuals is summarized in the table below.

Hours (nearest hour)	Frequency
1 – 10	5
11 – 20	16
21 – 25	14
26 – 30	17
31 – 40	10
41 – 59	2

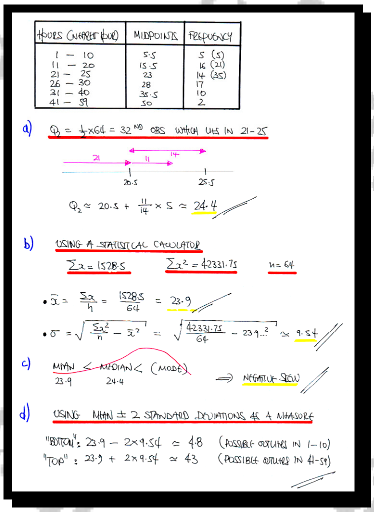
a) Estimate, by linear interpolation, the value of the median.

b) Estimate the mean and the standard deviation of these data.

c) Establish, with justification, the skewness of the data.

d) Determine the possibility whether the data contain any outliers.

$\boxed{}$, $\boxed{Q_2 \approx 24.4}$, $\boxed{\bar{x} \approx 23.88}$, $\boxed{\sigma \approx 9.54}$, $\boxed{\text{negative skew}}$



HOURS (NEAREST HOUR)	MIDPOINTS	FREQUENCY
1 – 10	5.5	5 (5)
11 – 20	15.5	16 (21)
21 – 25	23	14 (35)
26 – 30	28	17
31 – 40	35.5	10
41 – 59	50	2

a) $Q_2 = \frac{1}{2} \times 64 = 32^{ND}$ OBS WHICH LIES IN 21 – 25

21, 11, 14, 20.5, 25.5

$Q_2 \approx 20.5 + \frac{11}{14} \times 5 \approx 24.4$

b) USING A STATISTICAL CALCULATOR

$\sum x = 1528.5$ $\qquad \sum x^2 = 42331.75$ $\qquad n = 64$

• $\bar{x} = \frac{\sum x}{n} = \frac{1528.5}{64} = 23.9$

• $\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{42331.75}{64} - 23.9...^2} \simeq 9.54$

c) MEAN < MEDIAN < (MODE) $\Longrightarrow$ NEGATIVE SKEW

23.9 $\quad$ 24.4

d) USING MEAN ± 2 STANDARD DEVIATIONS AS A MEASURE

"BOTTOM": $23.9 - 2 \times 9.54 \simeq 4.8$ (POSSIBLE OUTLIERS IN 1 – 10)

"TOP": $23.9 + 2 \times 9.54 \simeq 43$ (POSSIBLE OUTLIERS IN 41 – 59)

Question 9 (***)

A group of patients with a certain respiratory condition were asked to hold their breath for as long as they could.

The results are summarized in the table below.

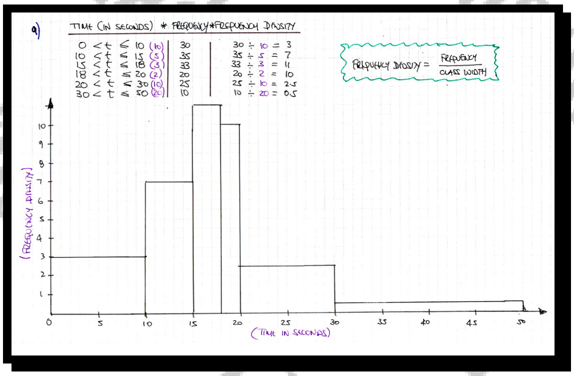
Time t (in seconds)	Frequency
$0 < t \le 10$	30
$10 < t \le 15$	35
$15 < t \le 18$	33
$18 < t \le 20$	20
$20 < t \le 30$	25
$30 < t \le 50$	10

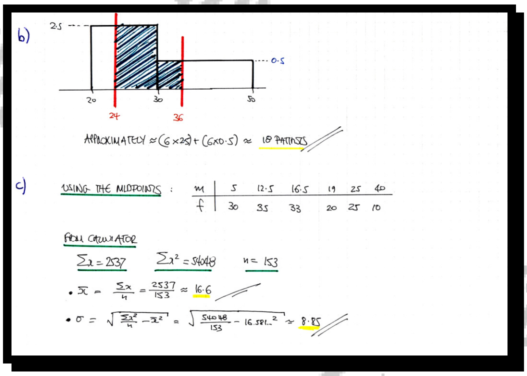
a) Draw an accurate histogram to represent this data.

b) Use the histogram to estimate the number of patients that managed to hold their breath between 24 and 36 seconds.

c) Calculate estimates for the mean and standard deviation of this data.

$\boxed{}$, $\boxed{\approx 18}$, $\boxed{\bar{x} \approx 16.6}$, $\boxed{\sigma \approx 8.85}$





Question 10 (***)

The daily commuting distances of 125 individuals, rounded to the nearest mile, is summarised in the table below.

Distance (nearest mile)	Frequency
0 – 9	12
10 – 19	22
20 – 29	48
30 – 39	26
40 – 49	8
50 – 59	5
60 – 69	3
70 – 79	1

a) Estimate the mean and the standard deviation of these commuting distances.

b) Use linear interpolation to estimate the value of the median.

c) Determine with justification the skewness of the data.

d) Explain which out of the mean and standard deviation or the median and the interquartile range are more appropriate measures to summarize this data.

$\boxed{}$, $\boxed{\bar{x} \approx 26.74}$, $\boxed{\sigma \approx 13.85}$, $\boxed{Q_2 = 25.3 - 25.5}$, $\boxed{\text{positive skew}}$, $\boxed{\text{median \& IQR}}$

DISTANCE (NEAREST MILE)	MIDPOINT	FREQUENCY
0 – 9	4.5	12 (12)
10 – 19	14.5	22 (34)
20 – 29	24.5	48 (82)
30 – 39	34.5	26 (108)
40 – 49	4.5	8 (116)
50 – 59	54.5	5 (121)
60 – 69	64.5	3 (124)
70 – 79	74.5	1 (125)

a) OBTAIN SUMMARY STATISTICS FROM CALCULATOR

- $\sum fx = 3342.5$
- $\sum fx^2 = 113351.25$
- $\sum f = 125$

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{3342.5}{125} = 26.74$$

$$\sigma = \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2} = \sqrt{\frac{113351.25}{125} - 26.74^2} \simeq 13.85$$

b) BY LINEAR INTERPOLATION – Q_2 IS $\frac{1}{2} \times 125 = 62.5$ OBS IN 20–29

34, 28.5, 48, 19.5, 29.5

$$\therefore Q_2 = 19.5 + \frac{28.5}{48} \times 10 \simeq 25.44$$

c) USING THE AVERAGES

(MODE) < MEDIAN < MEAN

25.44 26.74

$\therefore$ POSITIVE SKEW

d) WE TEND TO USE MEAN & STANDARD DEVIATION, IF THERE ARE NO EXTREME VALUES & NEGLIGIBLE SKEW, OTHERWISE WE TEND TO USE MEDIAN & IQR

HERE WE HAVE LITTLE SKEW, AS MEDIAN IS NUMERICALLY CLOSE TO THE MEAN, BUT THERE ARE EXTREME VALUES AT THE "TOP END"

$\therefore$ MEDIAN & IQR ARE PREFERABLE

Question 11 (***)

The ages of the residents of Arnold Street are denoted by x the ages of the residents of Benedict Street are denoted by y.

These are summarized in the following back to back stem and leaf diagram.

x		y
5	0	
5, 5, 3, 3	1	
9, 9, 1	2	5
9, 8, 6, 5, 5, 4, 3, 2, 2, 2, 1	3	6, 7, 8
6, 4, 1, 0, 0, 0, 0	4	1, 2, 2, 3, 4, 8
9	5	1, 4, 4, 4, 4, 5, 8, 8
	6	1, 3, 4, 4, 5, 9, 9
	7	2, 6, 9

where 2|3|9 = 32 in Arnold Street and 39 in Benedict Street.

a) Find separately for the residents of Arnold Street and Benedict Street, ...

i. ... the mode.

ii. ... the lower quartile, the median and the upper quartile.

iii. ... the mean and the standard deviation.

You may assume $\sum x = 866$, $\sum x^2 = 31514$, $\sum y = 1516$, $\sum y^2 = 86880$.

[continues overleaf]

[continued from overleaf]

A coefficient of skewness is defined as

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}.$$

b) Evaluate this coefficient for the ages in each street.

c) Compare the distribution of the ages between the two streets.

$$\boxed{\begin{array}{l} \text{mode} = 40 \\ Q_1 = 29 \\ Q_2 = 34 \\ Q_3 = 40 \\ \bar{x} \approx 32.07 \\ \sigma_x \approx 11.77 \\ \text{skew} \approx -0.67 \end{array}}, \quad \boxed{\begin{array}{l} \text{mode} = 54 \\ Q_1 = 42.5 \\ Q_2 = 54 \\ Q_3 = 64 \\ \bar{y} \approx 54.14 \\ \sigma_y \approx 13.09 \\ \text{skew} \approx 0.01 \end{array}}$$

	x ARNOLD ST		BENEDICT ST y	
(1)	5	0		
(4)	5 5 3 3	1		
(3)	9 9 1	2	5	(1)
(11)	9 8 6 5 5 4 3 2 2 2 1	3	6 7 8	(3)
(7)	6 4 1 0 0 0 0	4	1 2 2 3 4 8	(6)
(1)	9	5	1 4 4 4 4 5 8 8	(8)
		6	1 3 4 4 5 9 9	(7)
		7	2 6 9	(3)
27				28

$13 = 1|3|6 = 36$

a)

ARNOLD STREET	BENEDICT STREET
MODE = 40	MODE = 54
$Q_1 = \frac{1}{4}(27+1) = 7^{\text{TH}}$ OBS	$Q_1 = 7^{\text{TH}}/8^{\text{TH}}$ OBS
$Q_1 = 29$	$Q_1 = 42.5$
$Q_2 = \frac{1}{2}(27+1) = 14^{\text{TH}}$ OBS	$Q_2 = 14^{\text{TH}}/15^{\text{TH}}$ OBS
$Q_2 = 34$	$Q_2 = 54$
$Q_3 = \frac{3}{4}(27+1) = 21^{\text{ST}}$ OBS	$Q_3 = 21^{\text{ST}}/22^{\text{ND}}$ OBS
$Q_3 = 40$	$Q_3 = 64$
$\bar{x} = \frac{\Sigma x}{n} = \frac{866}{27} = 32.07$	$\bar{y} = \frac{\Sigma y}{n} = \frac{1516}{28} = 54.14$
$\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2} = 11.77$	$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - \bar{y}^2} = 13.09$

b) USING THE FORMULA GIVEN

FOR ARNOLD STREET $= \frac{32.07 - 40}{11.77} = -0.67$

FOR BENEDICT STREET $= \frac{54.14 - 54}{13.09} = 0.01$

c)

- MEDIAN & MEAN IS HIGHER IN BENEDICT STREET, INDICATING OLDER PEOPLE LIVING THERE
- BENEDICT STREET AGES ARE SLIGHTLY MORE VARIED, AS INDICATED BY THE STANDARD DEVIATION
- DATA IN ARNOLD STREET IS NEGATIVELY SKEWED AS INDICATED BY PART (b), WHILE DATA IN BENEDICT STREET IS PRACTICALLY SYMMETRICAL (SLIGHT POSITIVE SKEW) AS INDICATED BY (b)

Question 12 (***)

The number of hours worked in a given week by a group of 64 freelance electricians is summarized in the table below.

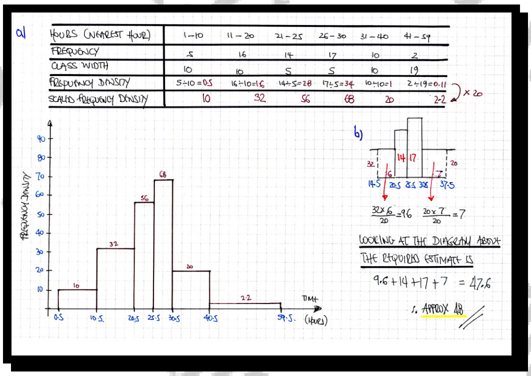
Hours (nearest hour)	Frequency
1 – 10	5
11 – 20	16
21 – 25	14
26 – 30	17
31 – 40	10
41 – 59	2

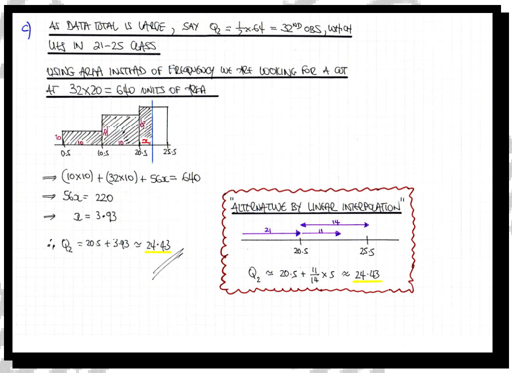
a) Draw an accurate histogram to represent this data.

b) Use the histogram to estimate the number of freelance electricians that worked between 15 and 37 hours during that week.

c) Estimate the median of the data.

$\boxed{}$, $\boxed{\approx 48}$, $\boxed{Q_2 \approx 24.4}$





Question 13 (***)

The number of hours spent on homework by 70 students, in a particular week, is summarized in the table below.

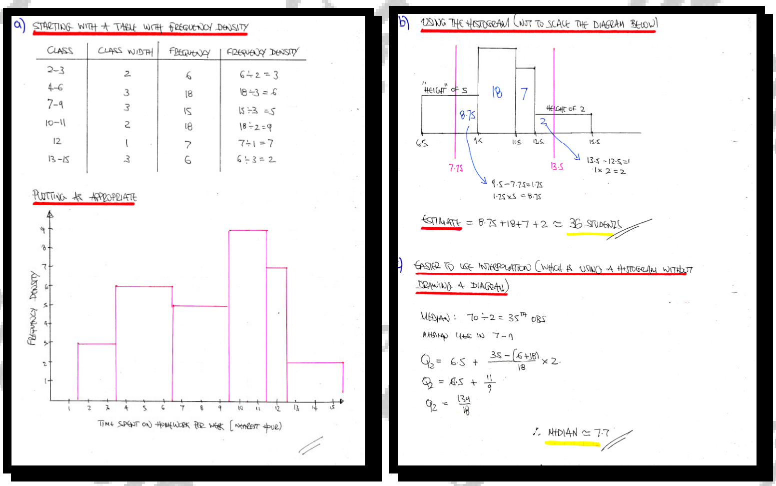
Hours (nearest hour)	Frequency
2 – 3	6
4 – 6	18
7 – 9	15
10 – 11	18
12	7
13 – 15	6

a) Draw an accurate histogram to represent this data.

b) Use the histogram to estimate the number of students that spent between 7.75 and 13.5 hours during that week.

c) Estimate the median of the data.

$\boxed{}$, $\boxed{\approx 36}$, $\boxed{Q_2 \approx 7.72}$



Question 14 (***)

The times taken to complete a 3 mile run, in minutes, by the members of a jogging club are summarized in the table below.

Times (nearest hour)	Frequency
11 – 14	24
15 – 17	24
18 – 19	19
20	11
21 – 23	21
24 – 28	15

a) Estimate the mean and standard deviation of this data.

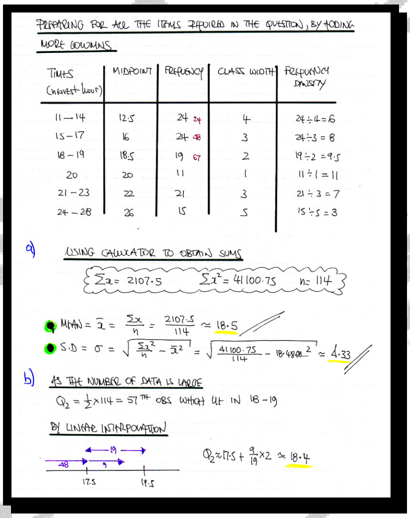
b) Estimate, by linear interpolation, the median of this data.

c) Draw an accurate histogram to represent this data.

d) Find the proportion of data which lies within 3 standard deviations of the mean.

e) Discuss briefly whether this data could be modelled by a Normal distribution.

$\boxed{}$, $\boxed{\bar{x} \approx 18.5}$, $\boxed{\sigma \approx 4.33}$, $\boxed{Q_2 \approx 18.4}$, $\boxed{100\%}$


PREPARING FOR ALL THE ITEMS REQUIRED IN THE QUESTION, BY ADDING MORE COLUMNS

TIMES (nearest hour)	MIDPOINT	FREQUENCY		CLASS WIDTH	FREQUENCY DENSITY
11 – 14	12.5	24	24	4	$24 \div 4 = 6$
15 – 17	16	24	48	3	$24 \div 3 = 8$
18 – 19	18.5	19	67	2	$19 \div 2 = 9.5$
20	20	11		1	$11 \div 1 = 11$
21 – 23	22	21		3	$21 \div 3 = 7$
24 – 28	26	15		5	$15 \div 5 = 3$

a) USING CALCULATOR TO OBTAIN SUMS

$\sum x = 2107.5 \qquad \sum x^2 = 41100.75 \qquad n = 114$

- MEAN $= \bar{x} = \frac{\sum x}{n} = \frac{2107.5}{114} \approx 18.5$
- S.D $= \sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{41100.75}{114} - 18.4868...^2} \approx 4.33$

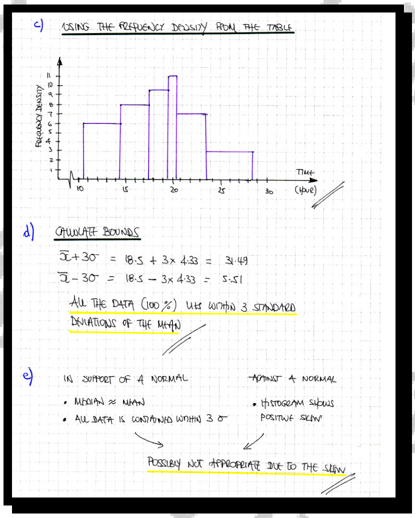
b) AS THE NUMBER OF DATA IS LARGE

$Q_2 = \frac{1}{2} \times 114 = 57^{\text{TH}}$ OBS WHICH LIE IN 18 – 19

BY LINEAR INTERPOLATION

48 9 19 17.5 19.5

$Q_2 \approx 17.5 + \frac{9}{19} \times 2 \approx 18.4$


c) USING THE FREQUENCY DENSITY FROM THE TABLE

d) CALCULATE BOUNDS

$\bar{x} + 3\sigma = 18.5 + 3 \times 4.33 = 31.49$

$\bar{x} - 3\sigma = 18.5 - 3 \times 4.33 = 5.51$

ALL THE DATA (100 %) LIES WITHIN 3 STANDARD DEVIATIONS OF THE MEAN

e) IN SUPPORT OF A NORMAL

- MEDIAN ≈ MEAN
- ALL DATA IS CONTAINED WITHIN 3 σ

AGAINST A NORMAL

- HISTOGRAM SHOWS POSITIVE SKEW

POSSIBLY NOT APPROPRIATE DUE TO THE SKEW

Question 15 (***)

The monthly mileages of a sales rep are summarised in the table below.

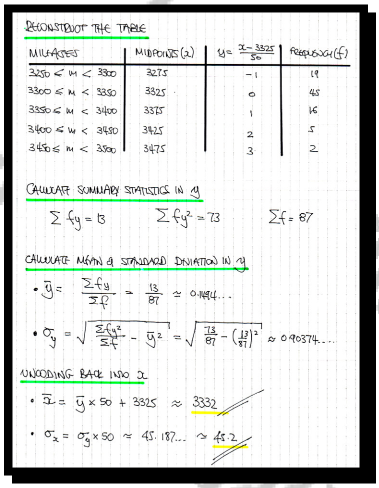
Mileages (m)	Frequency
$3250 \le m < 3300$	19
$3300 \le m < 3350$	45
$3350 \le m < 3400$	16
$3400 \le m < 3450$	5
$3450 \le m < 3500$	2

By using the coding

$$y = \frac{x - 3325}{50},$$

where x represents the midpoint of each class, estimate the mean and the standard deviation of this data.

$\boxed{\bar{x} \approx 3332}$, $\boxed{\sigma \approx 45.2}$


RECONSTRUCT THE TABLE

MILEAGES	MIDPOINTS (x)	$y = \frac{x-3325}{50}$	FREQUENCY (f)
$3250 \le m < 3300$	3275	-1	19
$3300 \le m < 3350$	3325	0	45
$3350 \le m < 3400$	3375	1	16
$3400 \le m < 3450$	3425	2	5
$3450 \le m < 3500$	3475	3	2

CALCULATE SUMMARY STATISTICS IN y

$\sum fy = 13 \qquad \sum fy^2 = 73 \qquad \sum f = 87$

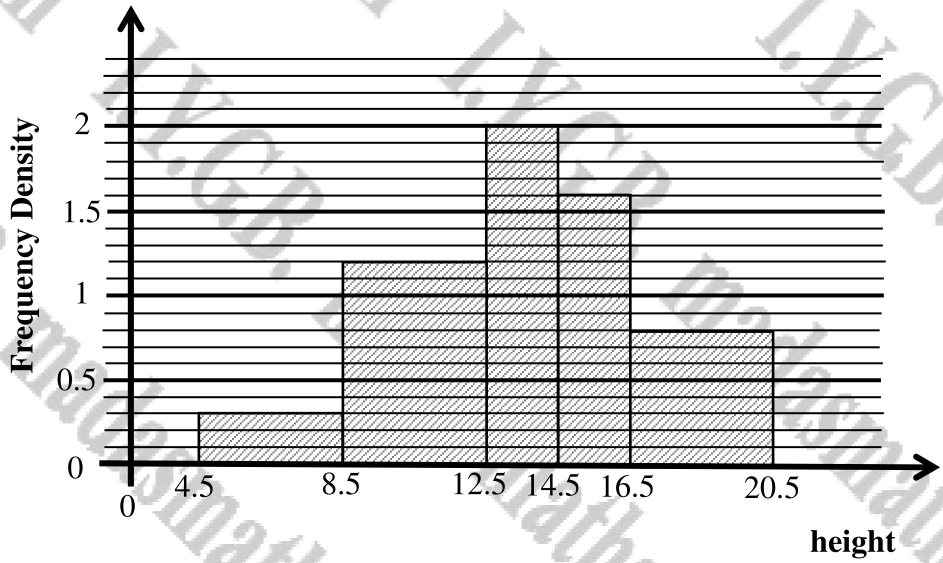
CALCULATE MEAN & STANDARD DEVIATION IN y

- $\bar{y} = \frac{\sum fy}{\sum f} = \frac{13}{87} \approx 0.1494\ldots$
- $\sigma_y = \sqrt{\frac{\sum fy^2}{\sum f} - \bar{y}^2} = \sqrt{\frac{73}{87} - \left(\frac{13}{87}\right)^2} \approx 0.90374\ldots$

UNCODING BACK INTO x

- $\bar{x} = \bar{y} \times 50 + 3325 \approx 3332$
- $\sigma_x = \sigma_y \times 50 \approx 45.187\ldots \approx 45.2$

Question 16 (***+)

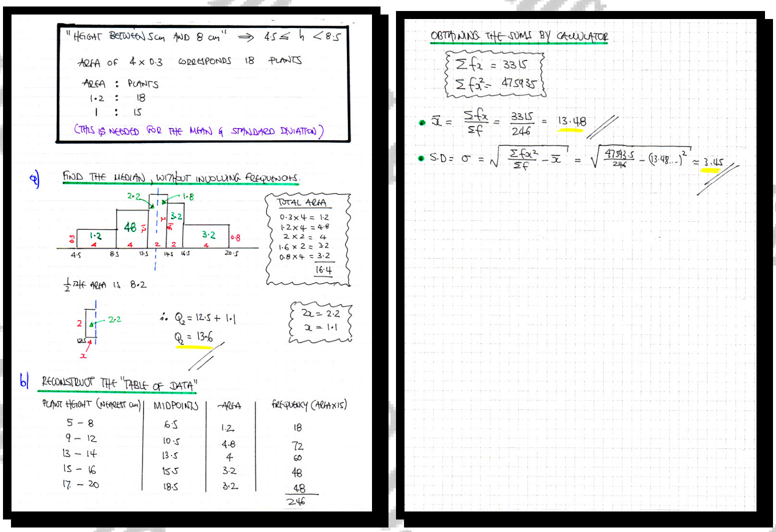


The histogram above shows the distribution of the heights, to the nearest cm, of some plants in a garden centre. It is further given that there were 18 plants with a height between 5 cm and 8 cm, rounded to the nearest cm.

a) Use the histogram to estimate the median.

b) Estimate, by calculation, the mean and the standard deviation of the heights of these plants.

$\boxed{\text{MMS-H}}$, $\boxed{\text{median} \approx 13.6}$, $\boxed{\bar{x} \approx 13.48}$, $\boxed{\sigma \approx 3.45}$



Question 17 (***+)

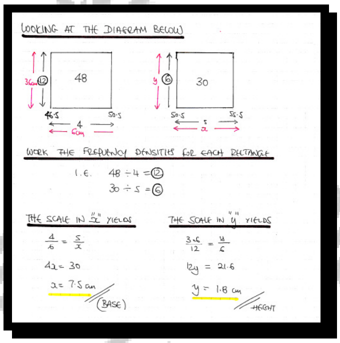
In a histogram the commuting times of a group of individuals, correct to the nearest minute, are plotted on the x axis.

In this histogram the class $47-50$ has a frequency of 48 and is represented by a rectangle of base 6 cm and height 3.6 cm.

In the same histogram the class $51-55$ has a frequency of 30.

Determine the measurements, in cm, of the rectangle that represents the class $51-55$.

$\boxed{}$, $\boxed{\text{base} = 7.5 \text{ cm}}$, $\boxed{\text{height} = 1.8 \text{ cm}}$



Question 18 **(***+)**

The diameters of fine sand particles, in mm, are summarised in the table below.

Diameters (d)	Frequency
$0.02 < d \le 0.04$	25
$0.04 < d \le 0.06$	76
$0.06 < d \le 0.08$	111
$0.08 < d \le 0.10$	255
$0.10 < d \le 0.12$	33

a) By using the coding

$$y = 50(x - 0.09),$$

where x represents the midpoint of each class, estimate the mean and the standard deviation of this data.

b) Estimate, by linear interpolation, the median diameter of these sand particles.

c) Describe, with justification, the skewness of the data.

$\boxed{}$, $\boxed{\bar{x} \approx 0.0778}$, $\boxed{\sigma \approx 0.0197}$, $\boxed{Q_2 = 0.08298}$

a) RECONSTRUCT THE TABLE

DIAMETERS (mm)	MIDPOINTS (x)	$y = 50(x - 0.09)$	FREQUENCY (f)
$0.02 < d \le 0.04$	0.03	−3	25 (25)
$0.04 < d \le 0.06$	0.05	−2	76 (101)
$0.06 < d \le 0.08$	0.07	−1	111 (212)
$0.08 < d \le 0.10$	0.09	0	255 (467)
$0.10 < d \le 0.12$	0.11	1	33 (500)

CALCULATE SUMMARY STATISTICS IN y

$\sum fy = -305$ $\quad \sum fy^2 = 673$ $\quad \sum f = 500$

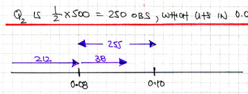
CALCULATE THE MEAN & STANDARD DEVIATION IN y

- $\bar{y} = \dfrac{\sum fy}{\sum f} = \dfrac{-305}{500} = -0.61$
- $\sigma_y = \sqrt{\dfrac{\sum fy^2}{\sum f} - \bar{y}^2} = \sqrt{\dfrac{673}{500} - (-0.61)^2} \approx 0.986863719\ldots$

UNCODE BACK INTO x

- $\bar{x} = \bar{y} \div 50 + 0.09 = -0.61 \div 50 + 0.09 = 0.0778$
- $\sigma_x = \sigma_y \div 50 = 0.986863\ldots \div 50 \approx 0.0197$

b) Q_2 IS $\frac{1}{2} \times 500 = 250$ OBS, WHICH LIES IN $0.08 < d \le 0.10$



$\Rightarrow Q_2 = 0.08 + \frac{38}{255} \times 0.02 \approx 0.0830$

c) USING THE AVERAGES

MEAN < MEDIAN < (MODE) $\Rightarrow$ POSITIVE SKEW

0.0778 $\quad$ 0.0830

Question 19 (***+)

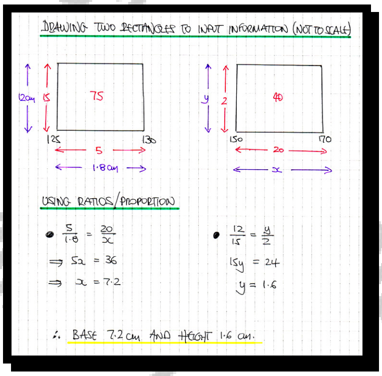
In a histogram the weights of apples, W grams, are plotted on the x axis.

In this histogram the class $125 \le W < 130$ has a frequency of 75 and is represented by a rectangle of base 1.8 cm and height 12 cm.

In the same histogram the class $150 \le W < 170$ has a frequency of 40.

Find the measurements, in cm, of the rectangle that represents the class $150 \le W < 170$.

$\boxed{}$, $\boxed{\text{base} = 7.2 \text{ cm}}$, $\boxed{\text{height} = 1.6 \text{ cm}}$



Question 20 (***+)

The masses, x kg, of 40 students were measured and the results were summarized using the notation below.

$$\sum_{n=1}^{40}(x_n - 50) = 140 \quad \text{and} \quad \sum_{n=1}^{40}(x_n - 50)^2 = 4490.$$

Calculate the mean and standard deviation of the masses of these 40 students.

$\boxed{\bar{x} = 53.5}$, $\boxed{\sigma = 10}$

LOOKING AT THE CODED SUMMARY STATISTICS

$$\sum_{n=1}^{40}(x_n - 50) = 140 \qquad \sum_{n=1}^{40}(x_n - 50)^2 = 4490$$

LET $y = x - 50$

$$\sum y = 140 \qquad \sum y^2 = 4490 \qquad n = 40$$

CALCULATE THE MEAN & STANDARD DEVIATION IN y

$$\bar{y} = \frac{\sum y}{n} = \frac{140}{40} = 3.5$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2} = \sqrt{\frac{4490}{40} - 3.5^2} = 10$$

UNCODE BACK INTO x

- $\bar{x} = \bar{y} + 50$

  $\bar{x} = 3.5 + 50$

  $\bar{x} = 53.5$

- $\sigma_x = \sigma_y$

  $\sigma_x = 10$

  (STANDARD DEVIATION DOES NOT GET AFFECTED BY ADDITION/SUBTRACTION)

Question 21 (***+)

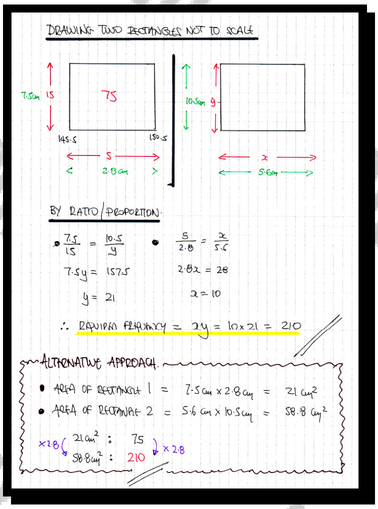
In a histogram the weights of peaches, correct to the nearest gram, are plotted on the x axis.

In this histogram the class $146-150$ has a frequency of 75 and is represented by a rectangle of base 2.8 cm and height 7.5 cm.

In the same histogram a different class is represented by a rectangle of base 5.6 cm and height 10.5 cm.

Determine the frequency of this class.

$\boxed{}$, $\boxed{f = 210}$



Question 22 (***+)

The following information about 5 observations of x is shown below.

$$\sum_{i=1}^{5}\left(\frac{x_i-255}{2}\right) = 50 \qquad \text{and} \qquad \sum_{i=1}^{5}\left(\frac{x_i-255}{2}\right)^2 = 1650 .$$

Calculate the mean and standard deviation of x.

$\boxed{}$, $\boxed{\bar{x}=275}$, $\boxed{\sigma = 2\sqrt{230} \approx 30.3}$

LOOKING AT THE SUMMARY STATISTICS

$$\sum_{i=1}^{5}\left(\frac{x_i-255}{2}\right) = 50 \qquad \sum_{i=1}^{5}\left(\frac{x_i-255}{2}\right)^2 = 1650$$

LET $y = \dfrac{x-255}{2}$

$$\sum y = 50 \qquad \sum y^2 = 1650 \qquad n=5$$

CALCULATE THE MEAN & STANDARD DEVIATION IN y

$$\bar{y} = \frac{\sum y}{n} = \frac{50}{5} = 10$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2} = \sqrt{\frac{1650}{5} - 10^2} = \sqrt{230}$$

UNCODE THE MEAN & STANDARD DEVIATION BACK INTO x.

- $\bar{x} = \bar{y} \times 2 + 255$
  $\bar{x} = 10 \times 2 + 255$
  $\bar{x} = 275$

- $\sigma_x = \sigma_y \times 2$
  $\sigma_x = 2\sqrt{230}$
  $\sigma_x \approx 30.3$

Question 23 (***+)

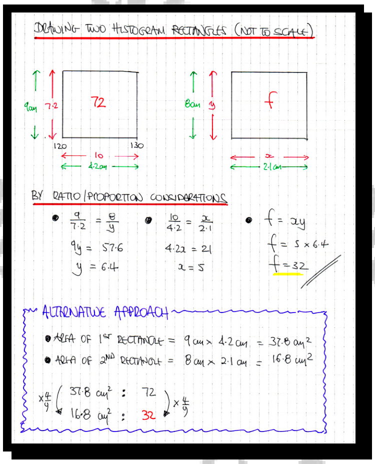
In a histogram the heights, h cm, of primary school pupils are plotted on the x axis.

In this histogram the class $120 \le h < 130$ has a frequency of 72 and is represented by a rectangle of base 4.2 cm and height 9 cm.

In the same histogram a different class is represented by a rectangle of base 2.1 cm and height 8 cm.

Determine the frequency of this class.

$\boxed{}$, $\boxed{f = 32}$



Question 24 (***+)

The table below shows the length of time, rounded to the nearest minute, spent by a group of patients for their dentist's check up visit.

One of the frequencies is given as a positive constant k.

Time (nearest minute)	2 - 6	7 - 11	12 - 16	17 - 31	32 - 36
Number of Patients	6	15	k	24	12

Determine the standard deviation of these times, given that the mean of these times is 18.6 minutes.

$\boxed{}$, $\boxed{\sigma \approx 9.37}$

TIME (NEAREST MIN)	2 – 6	7 – 11	12 – 16	17 – 31	32 – 36
	4	9	14	24	34
NO OF PATIENTS	6	15	k	24	12

- FINDING THE MIDPOINTS (ABOVE TABLE IN GREEN)
- THEN SET AN EQUATION FOR THE MEAN, GIVEN TO BE 18.6

$$\Longrightarrow \frac{(4\times6)+(9\times15)+(14k)+(24\times24)+(34\times12)}{6+15+k+24+12} = 18.6$$

$$\Longrightarrow \frac{24+135+14k+576+408}{k+57} = \frac{93}{5}$$

$$\Longrightarrow \frac{1143+14k}{k+57} = \frac{93}{5}$$

$$\Longrightarrow 93k + 5301 = 5715 + 70k$$

$$\Longrightarrow 23k = 414$$

$$\Longrightarrow k = 18$$

- FINALLY GETTING AN EXPRESSION FOR THE STANDARD DEVIATION

$$\Longrightarrow \sigma = \sqrt{\frac{\Sigma f x^2}{\Sigma f} - \bar{x}^2} \quad \text{OR} \quad \sigma = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

$$\Longrightarrow \sigma = \sqrt{\frac{(4^2\times6)+(9^2\times15)+(14^2\times18)+(24^2\times24)+(34^2\times12)}{6+15+18+24+12} - (18.6)^2}$$

$$\Longrightarrow \sigma = \sqrt{\frac{32535}{75} - (18.6)^2}$$

$$\Longrightarrow \sigma = \sqrt{87.84} \approx 9.37$$

Question 25 (***+)

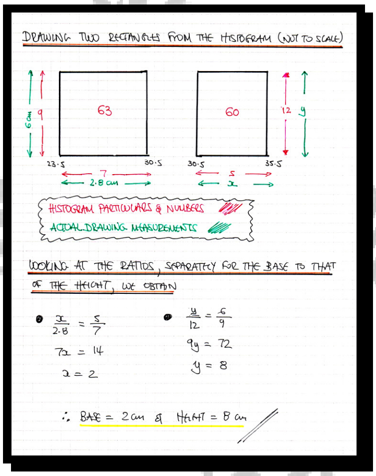
In a histogram the weights of baby hamsters, correct to the nearest gram, are plotted on the x axis.

In this histogram the class $24-30$ has a frequency of 63 and is represented by a rectangle of base 2.8 cm and height 6 cm.

In the same histogram the class $31-35$ has a frequency of 60.

Determine the measurements, in cm, of the rectangle that represents the class $31-35$.

$\boxed{}$, $\boxed{\text{base} = 2 \text{ cm}}$, $\boxed{\text{height} = 8 \text{ cm}}$



Question 26 (***+)

The distances rounded to the nearest mile, of 64 journeys covered by a taxi driver during a given week, is summarized in the table below.

Distance (nearest mile)	Frequency
3 – 5	12
6 – 7	14
8	19
9 – 11	13
12 – 17	6

a) Estimate the mean and the standard deviation of these weekly distances.

b) Estimate, by linear interpolation, the median value.

In a histogram drawn for the above data, the class 3 – 5 is represented by a rectangle of base length 1.2 cm and height 5 cm.

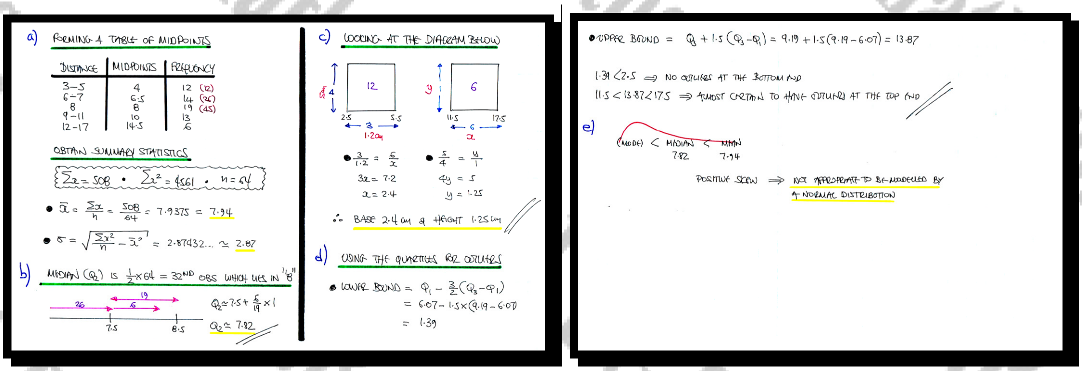
c) Find the base length and height of the rectangle representing the class 12 – 17 in the same histogram.

It is further given that the lower and upper quartiles of these distances are 6.07 and 9.19, respectively.

d) Investigate the possibility of any outliers.

e) By considering the skewness using the averages, discuss briefly whether the above set of data can be modelled by a normal distribution.

$\boxed{}$, $\boxed{\bar{x} \approx 7.94}$, $\boxed{\sigma \approx 2.87}$, $\boxed{Q_2 \approx 7.82}$, $\boxed{\text{base} = 2.4 \text{ cm}}$, $\boxed{\text{height} = 1.25 \text{ cm}}$



Question 27 (***+)

Weight in kg (w)	Frequency
$1 \le w < 3$	15
$3 \le w < 5$	31
$5 \le w < 6$	45
$6 \le w < 6.5$	37
$6.5 \le w < 7$	21
$7 \le w < 10$	15

The weights, in kg, of the 164 bags packed by supermarket customers is summarized in the table above.

a) Estimate the mean and the standard deviation of these weights.

b) Estimate, by linear interpolation, the median value and hence determine with justification, the skewness of the data.

In a histogram drawn for the above data, the $1 \le w < 3$ class is represented by a rectangle of base length 2.4 cm and height 2.5 cm.

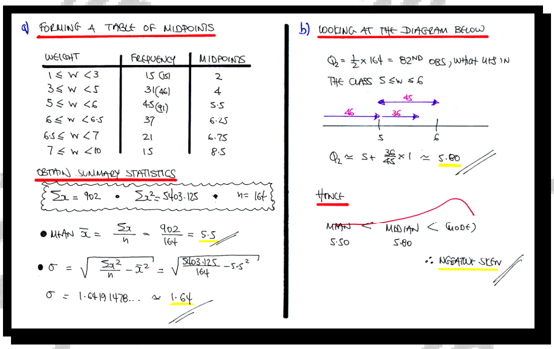
c) Find the base length and height of the rectangle representing the $6.5 \le w < 7$ class in the same histogram.

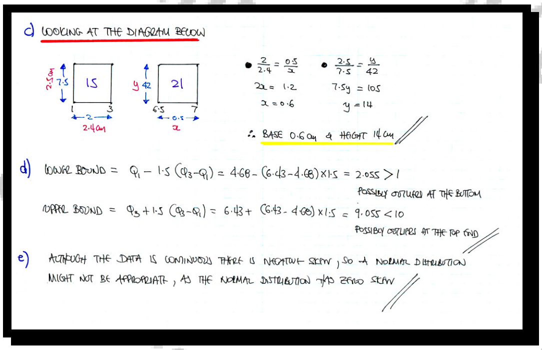
It is further given that the lower and upper quartiles of these distances are 4.68 and 6.43, respectively.

d) Investigate the possibility of any outliers.

e) Discuss briefly whether the above set of data can be modelled by a normal distribution.

$\boxed{}$, $\boxed{\bar{x} = 5.5}$, $\boxed{\sigma \approx 1.64}$, $\boxed{Q_2 \approx 5.80}$, $\boxed{\text{negative skew}}$, $\boxed{\text{base} = 0.6 \text{ cm}}$, $\boxed{\text{height} = 14 \text{ cm}}$





Question 28 (***+)

The masses of 68 cows, in kg, are summarised in the table below.

Mass (m)	Frequency
$600 < m \le 625$	11
$625 < m \le 650$	14
$650 < m \le 675$	28
$675 < m \le 700$	7
$700 < m \le 725$	5
$725 < m \le 750$	2
$750 < m \le 775$	1

a) By using the coding

$$y = \frac{x - 662.5}{25},$$

where x represents the midpoint of each class, estimate the mean and standard deviation of this data.

b) Estimate, by the method of linear interpolation, the median mass of these cows.

$\boxed{}$, $\boxed{\bar{x} \approx 659.19}$, $\boxed{\sigma \approx 32.91}$, $\boxed{Q_2 = 658.0}$

a)

"WEIGHTS" kg	FREQUENCY f	MIDPOINTS x	$y = \frac{x-662.5}{25}$	fy	fy^2
$600 < m \le 625$	11 (11)	612.5	−2	−22	44
$625 < m \le 650$	14 (25)	637.5	−1	−14	14
$650 < m \le 675$	28 (53)	662.5	0	0	0
$675 < m \le 700$	7	687.5	1	7	7
$700 < m \le 725$	5	712.5	2	10	20
$725 < m \le 750$	2	737.5	3	6	18
$750 < m \le 775$	1	762.5	4	4	16

$\Sigma fy = -9$ $\quad \Sigma fy^2 = 119$

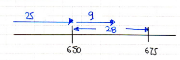
$$\bar{y} = \frac{\Sigma fy}{\Sigma f} = \frac{-9}{68}$$

$$\sigma_y = \sqrt{\frac{\Sigma fy^2}{\Sigma f} - \bar{y}^2} = \sqrt{\frac{119}{68} - \left(-\frac{9}{68}\right)^2} \simeq 1.316238\ldots$$

$$\bar{x} = (\bar{y} \times 25) + 662.5 \simeq 659.2$$

$$\sigma_x = 1.316238\ldots \times 25 \approx 32.9$$

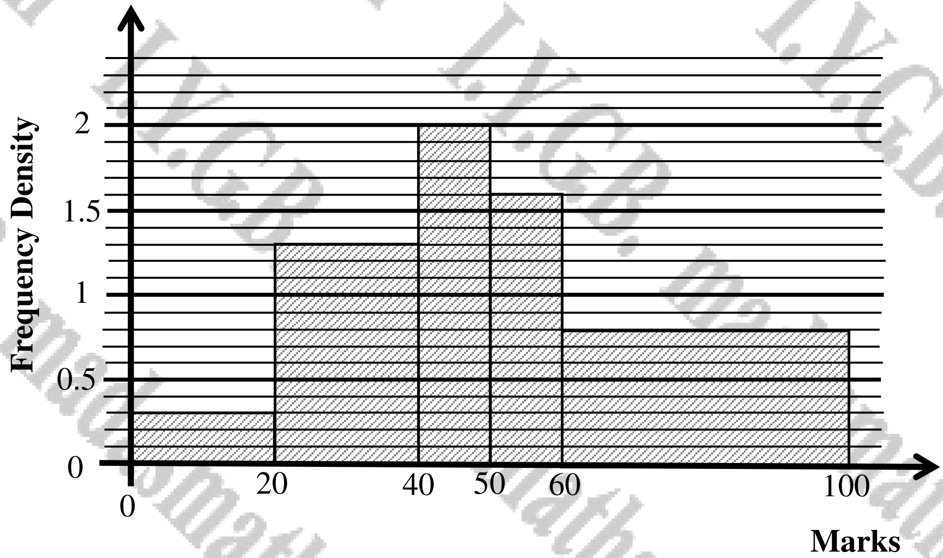
b) Q_2 IS $\frac{1}{2} \times 68 = 34^{\text{TH}}$ OBSERVATION WHICH LIES IN $650 < m \le 675$



$$\therefore Q_2 = 650 + \frac{9}{28} \times 25 \approx 658.0$$

Question 29 (****)

The histogram below shows the distribution of the marks of 250 students.

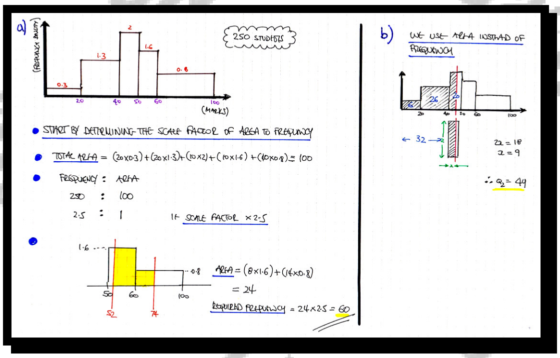


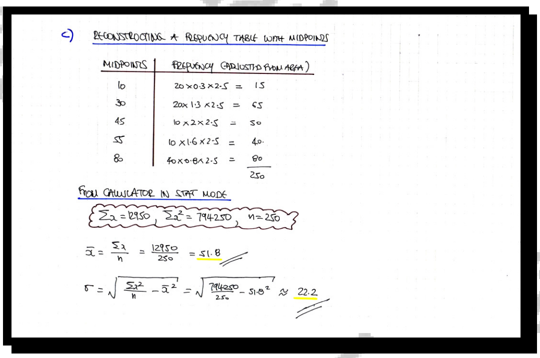
a) Estimate how many students scored between 52 and 74 marks.

b) Use the histogram estimate the median.

c) Calculate estimates for the mean and standard deviation of the marks of these students.

$\boxed{}$, $\boxed{60}$, $\boxed{49}$, $\boxed{\bar{x} \approx 51.8}$, $\boxed{\sigma \approx 22.22}$





Question 30 (****)

The mean and standard deviation of 20 observations $x_1, x_2, x_3, \ldots, x_{20}$ are

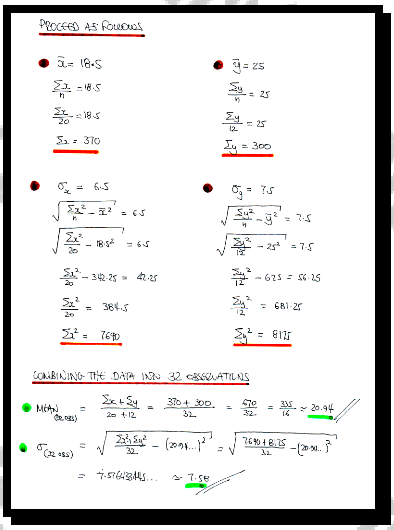
$$\bar{x} = 18.5 \quad \text{and} \quad \sigma_x = 6.5 .$$

The mean and standard deviation of 12 observations $y_1, y_2, y_3, \ldots, y_{12}$ are

$$\bar{y} = 25 \quad \text{and} \quad \sigma_y = 7.5 .$$

Determine the mean and the standard deviation of all 32 observations.

$\boxed{}$, $\boxed{\text{mean} \approx 20.94}$, $\boxed{\text{standard deviation} \approx 7.58}$


PROCEED AS FOLLOWS

- $\bar{x} = 18.5$
  $\frac{\Sigma x}{n} = 18.5$
  $\frac{\Sigma x}{20} = 18.5$
  $\Sigma x = 370$

- $\bar{y} = 25$
  $\frac{\Sigma y}{n} = 25$
  $\frac{\Sigma y}{12} = 25$
  $\Sigma y = 300$

- $\sigma_x = 6.5$
  $\sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2} = 6.5$
  $\sqrt{\frac{\Sigma x^2}{20} - 18.5^2} = 6.5$
  $\frac{\Sigma x^2}{20} - 342.25 = 42.25$
  $\frac{\Sigma x^2}{20} = 384.5$
  $\Sigma x^2 = 7690$

- $\sigma_y = 7.5$
  $\sqrt{\frac{\Sigma y^2}{n} - \bar{y}^2} = 7.5$
  $\sqrt{\frac{\Sigma y^2}{12} - 25^2} = 7.5$
  $\frac{\Sigma y^2}{12} - 625 = 56.25$
  $\frac{\Sigma y^2}{12} = 681.25$
  $\Sigma y^2 = 8175$

COMBINING THE DATA INTO 32 OBSERVATIONS

- $\text{MEAN}_{(32\text{ OBS})} = \frac{\Sigma x + \Sigma y}{20+12} = \frac{370+300}{32} = \frac{670}{32} = \frac{335}{16} \simeq 20.94$
- $\sigma_{(32\text{ OBS})} = \sqrt{\frac{\Sigma x^2 + \Sigma y^2}{32} - (20.94\ldots)^2} = \sqrt{\frac{7690+8175}{32} - (20.94\ldots)^2}$
  $= 7.57643344\ldots \simeq 7.58$

Question 31 (****)

The mean and standard deviation of the test marks of 40 pupils in a Mathematics class are 65 and 18, respectively.

The mean and standard deviation of the test marks of the 24 boys of the class are 72 and 20, respectively.

Find the mean and standard deviation of the test marks of the 16 girls of the class.

$\boxed{}$, $\boxed{\text{mean} = 54.5}$, $\boxed{\text{standard deviation} \approx 5.12}$

LOOKING AT THE INFORMATION GIVEN FOR THE WHOLE CLASS

- $\bar{x} = 65$

$$\frac{\sum x}{40} = 65$$

$$\sum x = 2600$$

- $\sigma = 18$

$$\sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = 18$$

$$\sqrt{\frac{\sum x^2}{40} - 65^2} = 18$$

$$\frac{\sum x^2}{40} - 4225 = 18^2$$

$$\sum x^2 = 181960$$

NOW REPEAT FOR THE 24 BOYS

- $\bar{x}_B = 72$

$$\frac{\sum x_B}{24} = 72$$

$$\sum x_B = 1728$$

- $\sigma_B = 20$

$$\sqrt{\frac{\sum x_B^2}{n} - \bar{x}_B^2} = 20$$

$$\sqrt{\frac{\sum x_B^2}{24} - 72^2} = 20$$

$$\frac{\sum x_B^2}{24} - 5184 = 20^2$$

$$\sum x_B^2 = 134016$$

SUBTRACTING THE SUMS FOUND & CALCULATE THE MEAN & STANDARD DEVIATION OF THE 16 GIRLS

$$\sum x_G = 2600 - 1728 = 872$$

$$\sum x_G^2 = 181960 - 134016 = 47944$$

$$\Rightarrow \bar{x}_G = \frac{\sum x_G}{n} = \frac{872}{16} = 54.5$$

$$\Rightarrow \sigma_G = \sqrt{\frac{\sum x_G^2}{n} - \bar{x}_G^2} = \sqrt{\frac{47944}{16} - 54.5^2} \approx 5.12$$

Question 32 (****)

The masses, x kg, of 40 students were measured and the results were summarized using the notation below.

$$\sum_{n=1}^{40}(x_n - 50) = 150 \quad \text{and} \quad \sum_{n=1}^{40}(x_n - 50)^2 = 4650 .$$

Determine the value of $\displaystyle\sum_{n=1}^{40}(x_n)^2$.

$$\boxed{}, \quad \boxed{\sum_{n=1}^{40}(x_n)^2 = 119650}$$

DEFINE $y = x - 50$

MEAN $\bar{y} = \dfrac{\sum(x-50)}{40} = \dfrac{150}{40} = 3.75$

VARIANCE $\sigma_y^2 = \dfrac{\sum(x-50)^2}{40} - \bar{y}^2 = \dfrac{4650}{40} - 3.75^2 = \dfrac{1635}{16} = 102.1875$

UNCODING AND NOTING THAT VARIANCE IS NOT AFFECTED BY "SUBTRACTION"

$\bar{x} = \bar{y} + 50 \qquad \sigma_x^2 = \sigma_y^2$

$\bar{x} = 53.75 \qquad \& \qquad \sigma_x^2 = 102.1875$

FINALLY USING THE VARIANCE FORMULA IN x

$\sigma_x^2 = \dfrac{\sum x^2}{n} - \bar{x}^2$

$102.1875 = \dfrac{\sum x^2}{40} - 53.75^2$

$4087.5 = \sum x^2 - 115562.5$

$\sum x^2 = 119650$

ALTERNATIVE USING SERIES

$\Rightarrow \sum (x-50)^2 = 4650$

$\Rightarrow \sum (x^2 - 100x + 2500) = 4650$

$\Rightarrow \sum x^2 - 100\sum x + 2500\sum 1 = 4650$

$\Rightarrow \sum x^2 - 100 \times 40 \times \dfrac{\sum x}{40} + 2500 \times 40 = 4650$

$\Rightarrow \sum x^2 - 4000\bar{x} + 100000 = 4650$

$\Rightarrow \sum x^2 = 4000\bar{x} - 95350$

NOW WE HAVE $\bar{x} = \dfrac{150}{40} + 50 = 53.75$ (AS BEFORE)

$\Rightarrow \sum x^2 = 4000 \times 53.75 - 95350$

$\Rightarrow \sum x^2 = 215000 - 95350$

$\Rightarrow \sum x^2 = 119650$

AS BEFORE

Question 33 (****+)

It is given that for a sample of data x_1, x_2, x_3, x_4, x_5, … x_n the mean $\bar{x}$ and standard deviation σ are

$$\bar{x} = \frac{1}{n}\sum_{r=1}^{n} x_r = 2 \qquad \text{and} \qquad \sigma = \sqrt{\frac{1}{n}\sum_{r=1}^{n}(x_r)^2 - \frac{1}{n^2}\left(\sum_{r=1}^{n} x_r\right)^2} = 3.$$

Determine, in terms of n, the value of

$$\sum_{r=1}^{n}(x_r+1)^2.$$

$$\boxed{}, \quad \boxed{\sum_{r=1}^{n}(x_r+1)^2 = 18n}$$

$$\boxed{\bar{x} = \frac{1}{n}\sum_{r=1}^{n} x_r = 2 \qquad \sigma = \sqrt{\frac{1}{n}\sum_{r=1}^{n}(x_r)^2 - \frac{1}{n^2}\left[\sum_{r=1}^{n} x_r\right]^2} = 3}$$

PROCEED AS FOLLOWS & DROPPING SUBSCRIPTS/SUPERSCRIPTS IN SIGMA

- $\frac{1}{n}\sum x = 2$

  $\sum x = 2n$

- $\sqrt{\frac{1}{n}\sum x^2 - \frac{1}{n^2}\left(\sum x\right)^2} = 3$

  $\frac{1}{n}\sum x^2 - \frac{1}{n^2}\left(\sum x\right)^2 = 9$

  $\frac{1}{n}\sum x^2 - \left(\frac{\sum x}{n}\right)^2 = 9$

  $\frac{1}{n}\sum x^2 - \bar{x}^2 = 9$

  $\frac{1}{n}\sum x^2 - 2^2 = 9$

  $\frac{1}{n}\sum x^2 = 13$

  $\sum x^2 = 13n$

HENCE WE NOW HAVE

$$\begin{aligned}\sum_{r=1}^{n}(x_r+1)^2 &= \sum_{r=1}^{n}(x_r^2 + 2x_r + 1) \\ &= \sum_{r=1}^{n}(x_r^2) + \sum_{r=1}^{n}(2x_r) + \sum_{r=1}^{n}1 \\ &= \sum_{r=1}^{n}(x_r^2) + 2\sum_{r=1}^{n}x + \sum_{r=1}^{n}1 \\ &= 13n + 2 \times 2n + n \\ &= 18n\end{aligned}$$

Question 34 (****+)

The test marks, x, of 20 students were coded and their results were summarized as

$$\sum (x-10) = 220 \quad \text{and} \quad \sum (x-10)^2 = 2720.$$

a) Use a detailed method to show that

$$\sum x^2 = 9120.$$

b) Calculate the mean and standard deviation of the test marks of these students.

$\boxed{}$, $\boxed{\bar{x} = 21}$, $\boxed{\sigma = \sqrt{15} \approx 3.87}$

a) $\sum_{r=1}^{20} (x_r - 10) = 220 \quad \sum_{r=1}^{20} (x_r - 10)^2 = 2720 \quad n = 20$

$$\sum_{r=1}^{20} (x_r - 10)^2 = \sum_{r=1}^{20} \left[x_r^2 - 20x_r + 100\right]$$

$$2720 = \sum_{r=1}^{20} x_r^2 - 20\sum_{r=1}^{20} x_r + 100\sum_{r=1}^{20} 1$$

$$2720 = \sum_{r=1}^{20} x_r^2 - 20\sum_{r=1}^{20} x_r + 100 \times 20$$

BY INSPECTION $\sum_{r=1} x = 220 + 20 \times 10 = 420$ OR BY USING A DETAILED METHOD

$$\sum_{r=1}^{20} (x_r - 10) = 220$$

$$\sum_{r=1}^{20} x_r - 10\sum_{r=1}^{20} 1 = 220$$

$$\sum_{r=1}^{20} x_r - 10 \times 20 = 220$$

$$\sum_{r=1}^{20} x_r = 420$$

RETURNING TO THE MAIN LINE

$$\Rightarrow 2720 = \sum_{r=1}^{20} x_r^2 - 20 \times 420 + 2000$$

$$\Rightarrow \sum_{r=1}^{20} x_r^2 = 9120$$

b)

- $\bar{x} = \frac{\Sigma x}{n} = \frac{420}{20} = 21$
- $\sigma = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2} = \sqrt{\frac{9120}{20} - 21^2} = \sqrt{15}$ ≈ 3.87

ALTERNATIVE USING THE CODED VALUES

$y = x - 10$ so $\Sigma y = 220$ & $\Sigma y^2 = 2720$

$\bar{y} = \frac{\Sigma y}{n} = \frac{220}{20} = 11$

$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - \bar{y}^2} = \sqrt{\frac{2720}{20} - 11^2} = \sqrt{15}$

UNCODING

$\bar{x} = \bar{y} + 10 = 21$

$\sigma_x = \sigma_y = \sqrt{15}$ (UNAFFECTED BY SUBTRACTION)