# DATA ANALYSIS

# DISCRETE DATA

**Question 1**

The following set of data is given

$$11,\ 45,\ 47,\ 49,\ 50,\ 52,\ 53,\ 59,\ 60,\ 94,\ 117.$$

For this set of data, ...

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

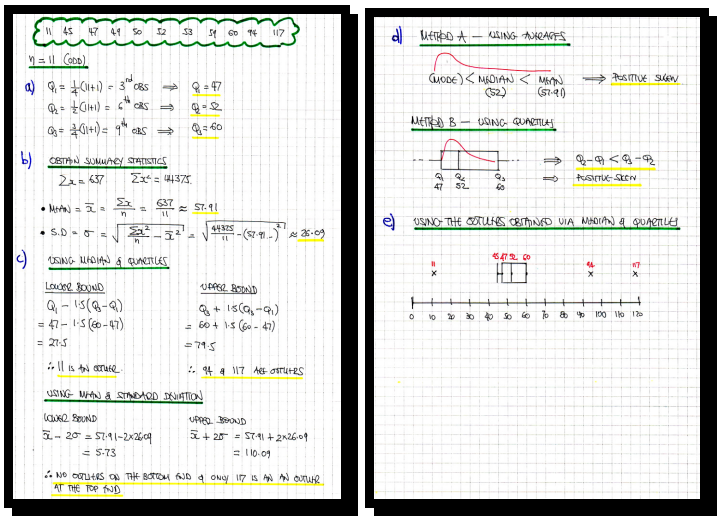
**b)** ... calculate the mean and the standard deviation.

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

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

**e)** ... draw a suitably labelled box plot.

$\square$, $\boxed{Q_1 = 47}$, $\boxed{Q_2 = 52}$, $\boxed{Q_3 = 60}$, $\boxed{\overline{x} = 57.91}$, $\boxed{\sigma \approx 26.09}$,
$\boxed{\text{(11, 94) and 117 are outliers depending on method}}$, $\boxed{\text{positive skew}}$



**Question 2**

The following set of data is given

$$25,\ 47,\ 47,\ 49,\ 50,\ 52,\ 54,\ 55,\ 56,\ 56,\ 56,\ 59.$$

For this set of data, ...

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

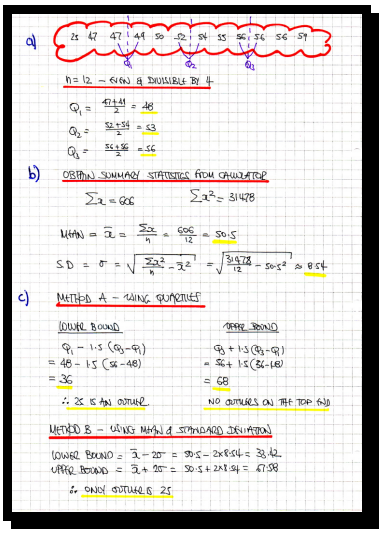
**b)** ... calculate the mean and the standard deviation.

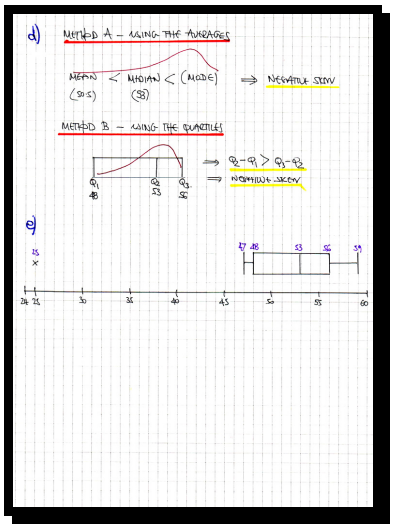
**c)** ... determine with justification whether there are any outliers.

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

**e)** ... draw a suitably labelled box plot.

$\square$, $\boxed{Q_1 = 48}$, $\boxed{Q_2 = 53}$, $\boxed{Q_3 = 56}$, $\boxed{\overline{x} = 50.5}$, $\boxed{\sigma \approx 8.54}$, $\boxed{25 \text{ is an outlier}}$, $\boxed{\text{negative skew}}$





**Question 3**

The following set of data is given

$$78,\ 79,\ 79,\ 79,\ 80,\ 82,\ 82,\ 85,\ 86,\ 88,\ 89,\ 92,\ 97.$$

For this set of data, ...

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

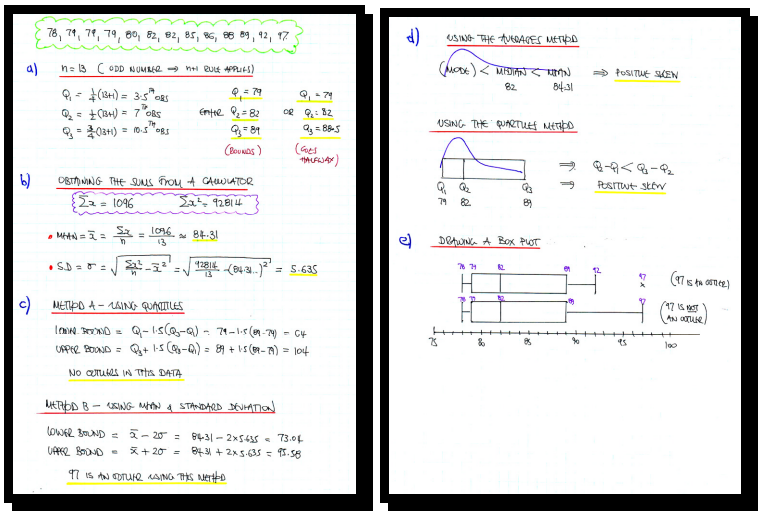
**b)** ... calculate the mean and the standard deviation.

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

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

**e)** ... draw a suitably labelled box plot.

$\square$, $\boxed{Q_1 = 79}$, $\boxed{Q_2 = 82}$, $\boxed{Q_3 = 89}$, $\boxed{\overline{x} = 84.31}$, $\boxed{\sigma \approx 5.635}$,
$\boxed{\text{no outliers or 97 is an outlier depenending on method}}$, $\boxed{\text{positive skew}}$



**Question 4**

The following set of data is given

$$50,\ 52,\ 52,\ 52,\ 53,\ 54,\ 56,\ 58,\ 61,\ 64,\ 65,\ 69,\ 75,\ 77.$$

For this set of data, ...

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

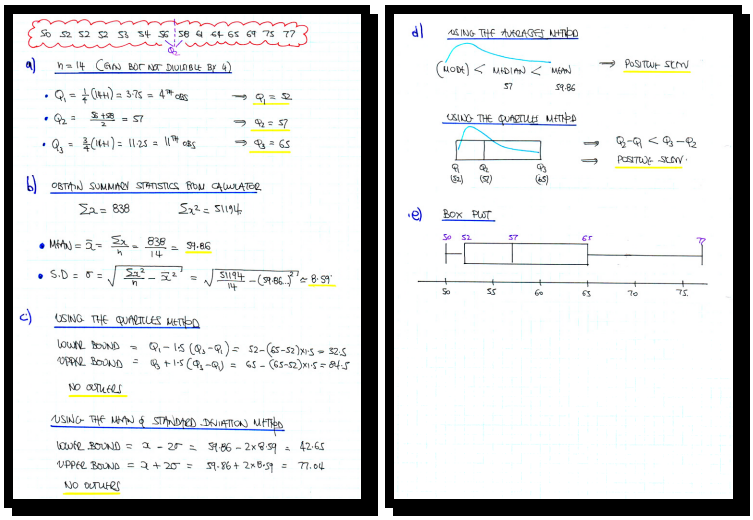
**b)** ... calculate the mean and the standard deviation.

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

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

**e)** ... draw a suitably labelled box plot.

$\square$, $\boxed{Q_1 = 52}$, $\boxed{Q_2 = 57}$, $\boxed{Q_3 = 65}$, $\boxed{\overline{x} = 59.86}$, $\boxed{\sigma \approx 8.59}$, $\boxed{\text{no outliers}}$, $\boxed{\text{positive skew}}$



**Question 5**

The following set of data is given

$$7,\ 8,\ 10,\ 12,\ 16,\ 19,\ 19,\ 20,\ 23,\ 24,\ 30.$$

For this set of data, ...

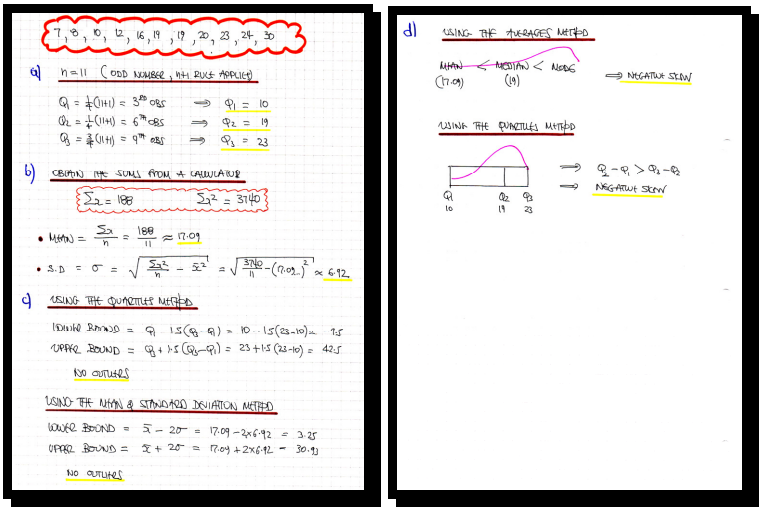
**a)** ... determine the value of the median and the quartiles.

**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.

$\square$, $\boxed{Q_1 = 10}$, $\boxed{Q_2 = 19}$, $\boxed{Q_3 = 23}$, $\boxed{\overline{x} \approx 17.09}$, $\boxed{\sigma \approx 6.92}$, $\boxed{\text{no outliers}}$, $\boxed{\text{negative skew}}$



**Question 6**

The following set of data is given

$$5,\ 10,\ 11,\ 11,\ 14,\ 17,\ 19,\ 24,\ 30,\ 36,\ 40,\ 45.$$

For this set of data, ...

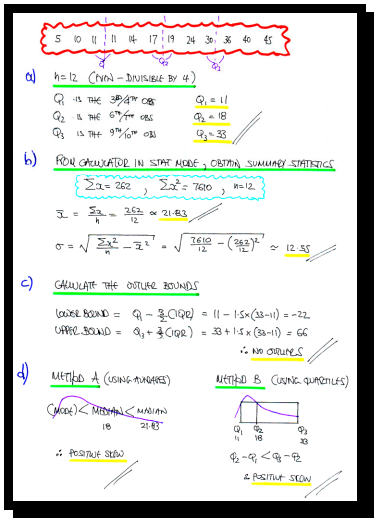
**a)** ... determine the value of the median and the quartiles.

**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.

$\square$, $\boxed{Q_1 = 11}$, $\boxed{Q_2 = 18}$, $\boxed{Q_3 = 33}$, $\boxed{\overline{x} \approx 21.83}$, $\boxed{\sigma \approx 12.55}$, $\boxed{\text{no outliers}}$, $\boxed{\text{positive skew}}$



## Question 7

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, ...

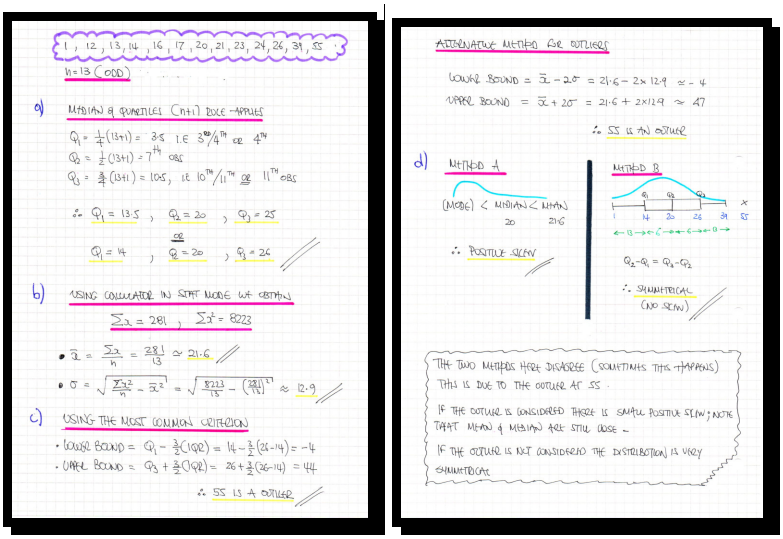
**a)** ... determine the value of the median and the quartiles.

**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.

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



**Question 8**

The following set of data is given

$$21,\ 50,\ 51,\ 53,\ 54,\ 57,\ 58,\ 60,\ 62,\ 63,\ 64,\ 68,\ 82,\ 97.$$

For this set of data, ...

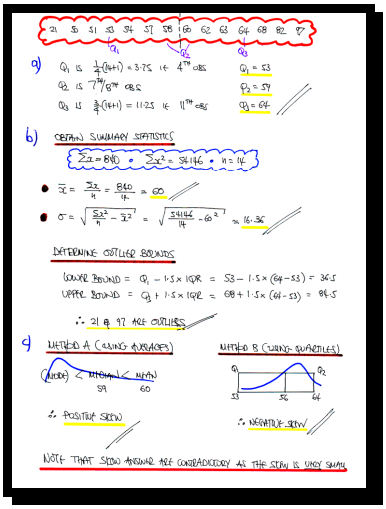
**a)** ... determine the value of the median and the quartiles.

**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.

$\square$, $\boxed{Q_1 = 53}$, $\boxed{Q_2 = 59}$, $\boxed{Q_3 = 64}$, $\boxed{\overline{x} = 60}$, $\boxed{\sigma \approx 16.36}$, $\boxed{\text{21 and 97 are outliers}}$, $\boxed{\text{positive/negative skew}}$



**Question 9**

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 $5\|2 = 52$.

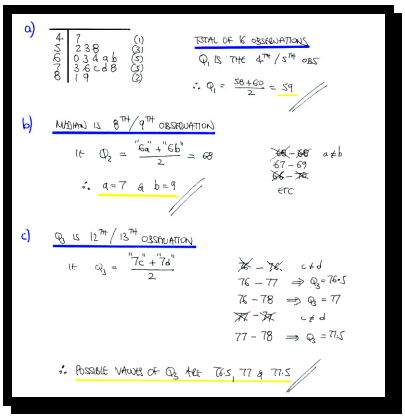
**a)** Determine the lower quartile of the data.

**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.

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



## Question 10

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

**b)** Find the mean and the standard deviation of the data.

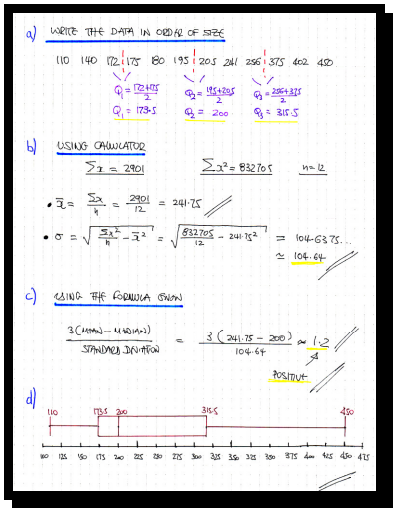
The skewness of data can be determined by the formula

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

**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.

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



**Question 11**

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

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

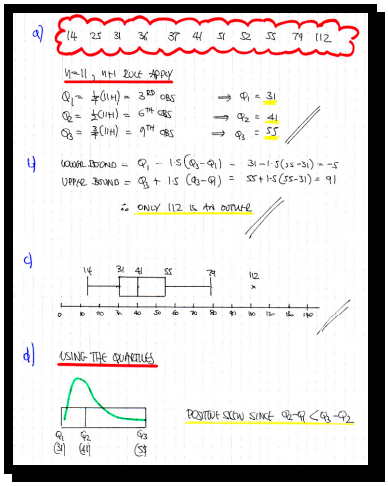
**a)** Find the lower quartile, the median and the upper quartile of the data.

**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.

$\boxed{Q_1 = 3}$, $\boxed{Q_2 = 41}$, $\boxed{Q_3 = 55}$, $\boxed{112 \text{ is the only outlier}}$, $\boxed{\text{positive skew}}$



## Question 12

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.

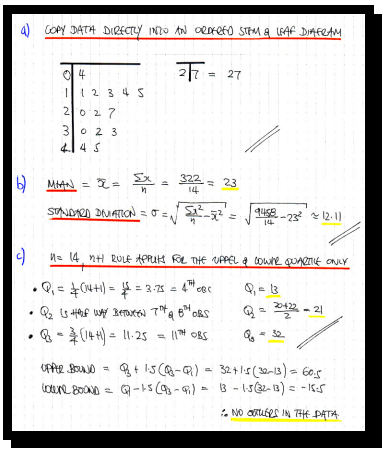
**b)** Calculate the mean and the standard deviation of the data.

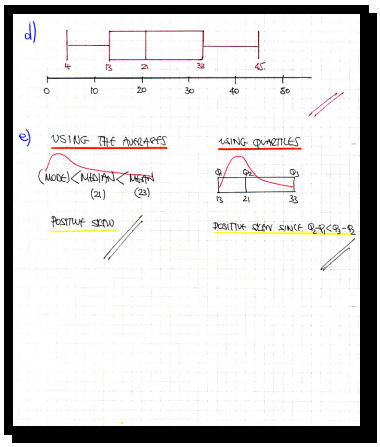
**c)** Find the median and the quartiles of the data and use them to determine if there are any outliers.

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

**e)** Determine with justification the skewness of the data.

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



**Question 13**

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 $\overline{2\|9} = 29$.

**a)** Find the median score of the test.

**b)** Determine the interquartile range of the scores.

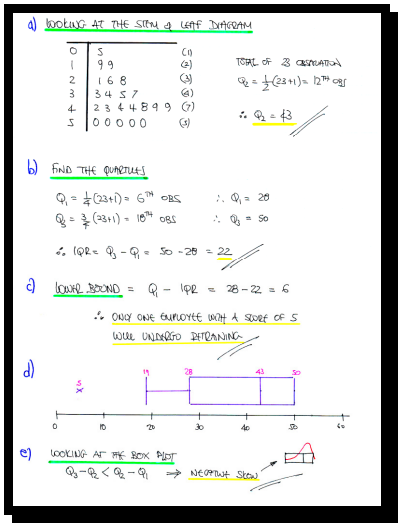
The company decides to retrain any employee whose score is less than the **lower quartile minus the interquartile range**.

**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.

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



**Question 14**

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 $\overline{2}|\overline{3}|\overline{9} = 32$ 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]**

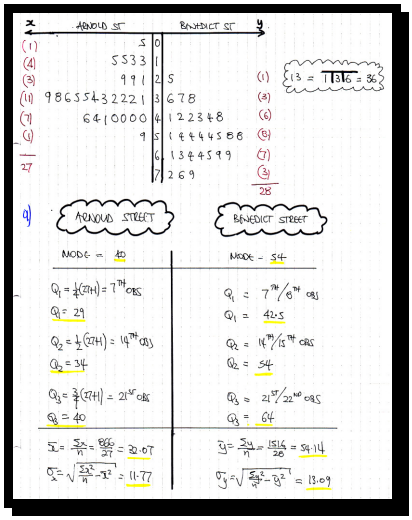
A coefficient of skewness is defined as

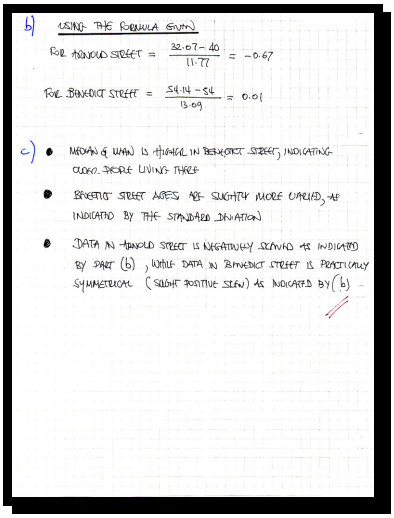
$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}.$$

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

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

| | |
|---|---|
| $\text{mode} = 40$ | $\text{mode} = 54$ |
| $Q_1 = 29$ | $Q_1 = 42.5$ |
| $Q_2 = 34$ | $Q_2 = 54$ |
| $Q_3 = 40$ | $Q_3 = 64$ |
| $\overline{x} \approx 32.07$ | $\overline{y} \approx 54.14$ |
| $\sigma_x \approx 11.77$ | $\sigma_y \approx 13.09$ |
| $\text{skew} \approx -0.67$ | $\text{skew} \approx 0.01$ |



**Question 15**

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

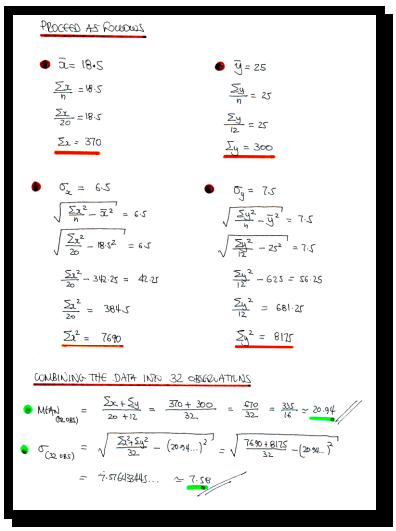
$$\overline{x} = 18.5 \quad \text{and} \quad \sigma_x = 6.5 .$$

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

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

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

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



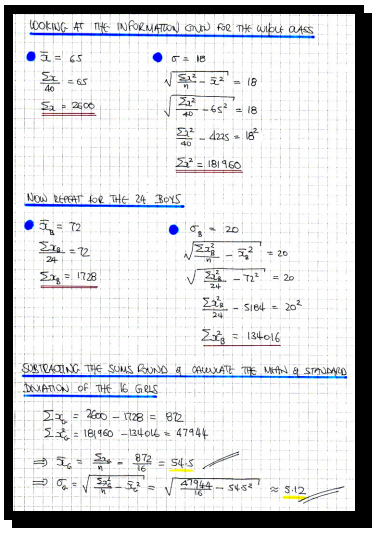
**Question 16**

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.

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



**Question 17**

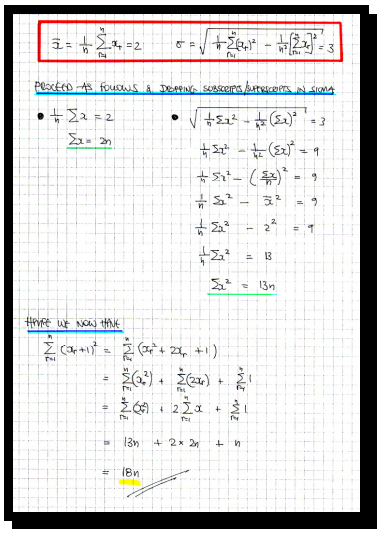
It is given that for a sample of data $x_1,\ x_2,\ x_3, x_4, x_5, \ldots x_n$ the mean $\overline{x}$ and standard deviation $\sigma$ are

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

Determine, in terms of $n$, the value of

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

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



# CONTINUOUS DATA

## Question 1

The distances achieved by a group of javelin throwers, rounded to the nearest metre, is summarized in the table below.

| Distance (nearest metre) | Frequency |
|---|---|
| 30 – 35 | 12 |
| 36 – 38 | 18 |
| 39 – 40 | 21 |
| 41 | 37 |
| 42 – 49 | 10 |
| 50 – 60 | 2 |

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

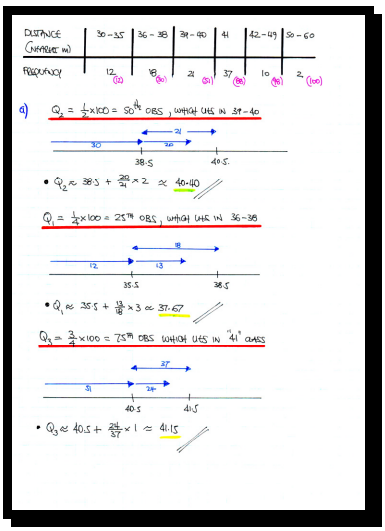
**b)** Estimate the mean and the standard deviation of these data.

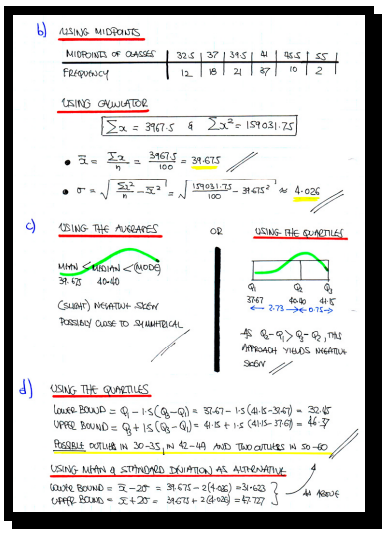
**c)** Determine with justification the skewness of the data.

**d)** Investigate the possibility of any outliers

$\square$, $\boxed{Q_1 \approx 37.67}$, $\boxed{Q_2 \approx 40.40}$, $\boxed{Q_3 \approx 41.15}$, $\boxed{\overline{x} = 39.675}$, $\boxed{\sigma \approx 4.026}$,

$\boxed{\text{negative skew}}$





**Question 2**

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 |

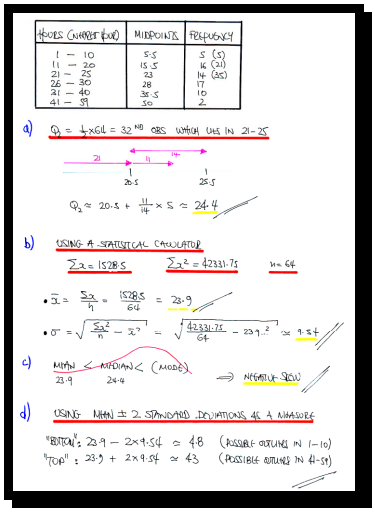
**a)** Estimate, by linear interpolation, the value of the median.

**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.

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



**Question 3**

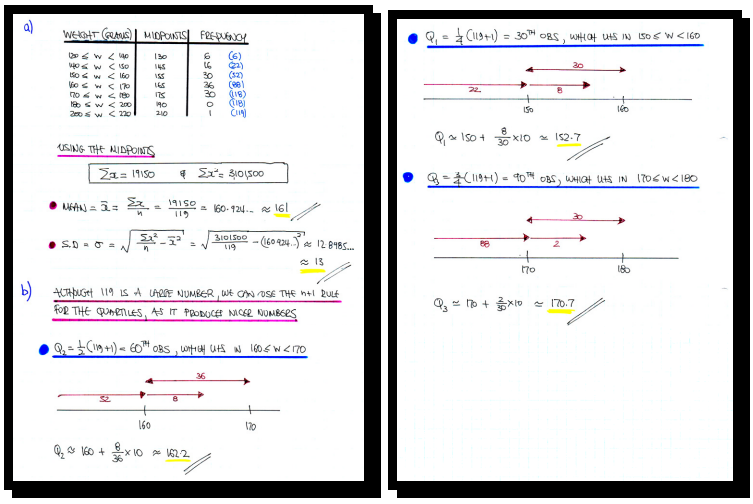
The weights of a random sample of a variety of apples, in grams, is summarised in the table below.

| Weight (grams) | Frequency |
|---|---|
| $120 \leq w < 140$ | 6 |
| $140 \leq w < 150$ | 16 |
| $150 \leq w < 160$ | 30 |
| $160 \leq w < 170$ | 36 |
| $170 \leq w < 180$ | 30 |
| $180 \leq w < 200$ | 0 |
| $200 \leq w < 220$ | 1 |

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

**b)** Estimate the median and the quartiles of these weights.

$\square$, $\boxed{\bar{x} \approx 161}$, $\boxed{\sigma \approx 13}$, $\boxed{Q_1 \approx 152.6 - 152.7}$, $\boxed{Q_2 \approx 162.1 - 162.2}$, $\boxed{Q_3 \approx 170.4 - 170.7}$



## Question 4

The mileages of 120 journeys covered by a minicab driver over a monthly period are summarized in the table below.

| Mileages | Frequency |
|---|---|
| $10 \leq m < 12$ | 2 |
| $12 \leq m < 17$ | 54 |
| $17 \leq m < 19$ | 28 |
| $19 \leq m < 21$ | 16 |
| $21 \leq m < 23$ | 13 |
| $23 \leq m < 25$ | 7 |

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

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

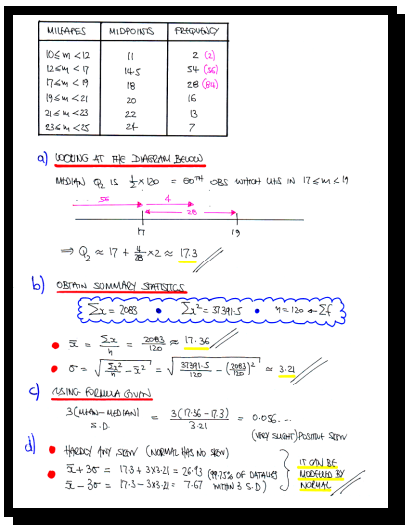
A skewness coefficient can be determined by

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

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

**d)** Determine whether the data could be modelled by a Normal distribution.

$\square$, $\boxed{Q_2 \approx 17.3}$, $\boxed{\overline{x} \approx 17.36}$, $\boxed{\sigma \approx 3.21}$, $\boxed{\text{slight positive skew/no skew}}$



## Question 5

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 |

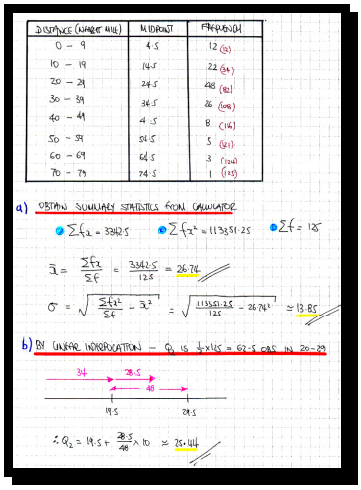
**a)** Estimate the mean and the standard deviation of these commuting distances.

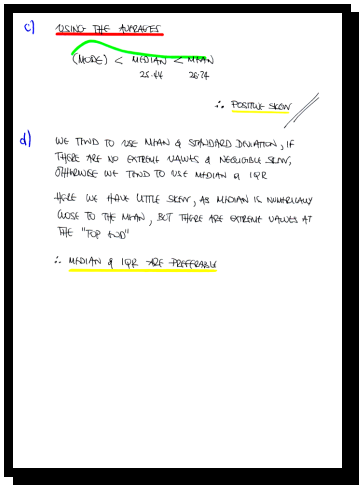
**b)** Use linear interpolation to estimate the value of the median.

**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.

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





# HISTOGRAMS

**Question 1**

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.

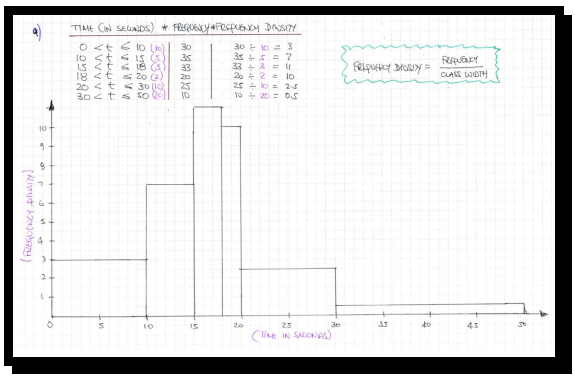
| Time $t$ (in seconds) | Frequency |
|---|---|
| $0 < t \leq 10$ | 30 |
| $10 < t \leq 15$ | 35 |
| $15 < t \leq 18$ | 33 |
| $18 < t \leq 20$ | 20 |
| $20 < t \leq 30$ | 25 |
| $30 < t \leq 50$ | 10 |

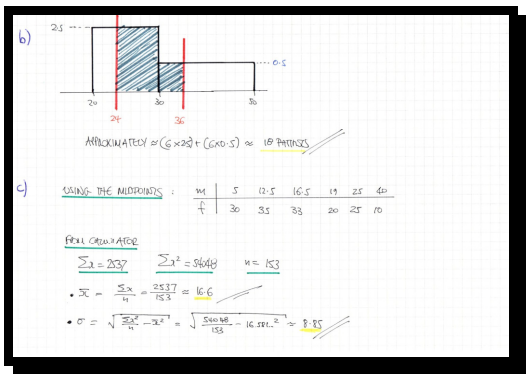
**a)** Draw an accurate histogram to represent this data.

**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.

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



## Question 2

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

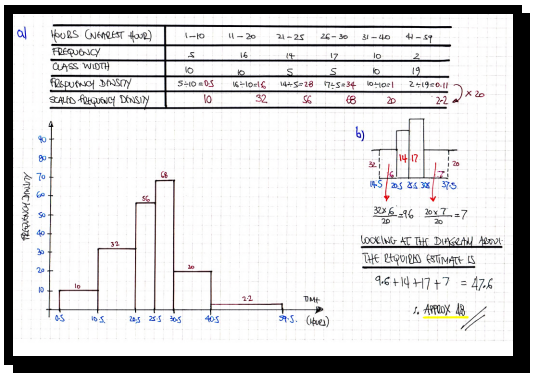
| Hours (nearest hour) | Frequency |
|---|---|
| 1 – 10 | 5 |
| 11 – 20 | 16 |
| 21 – 25 | 14 |
| 26 – 30 | 17 |
| 31 – 40 | 10 |
| 41 – 59 | 2 |

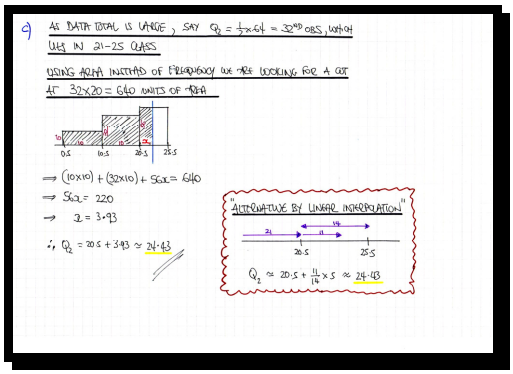
**a)** Draw an accurate histogram to represent this data.

**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.

$\approx 48$, $Q_2 \approx 24.4$





## Question 3

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.

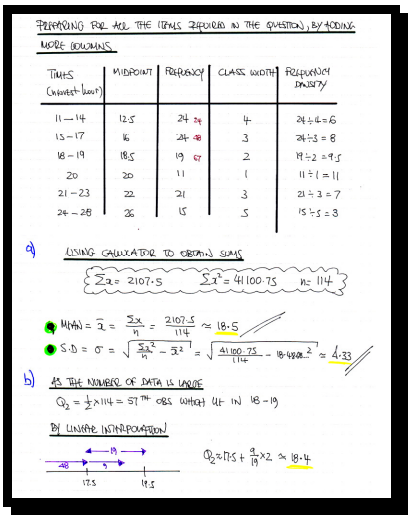
**b)** Estimate, by linear interpolation, the median of this data.

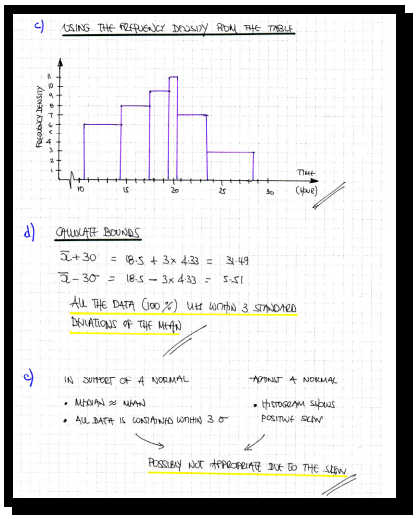
**c)** Draw an accurate histogram to represent this data.

**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.

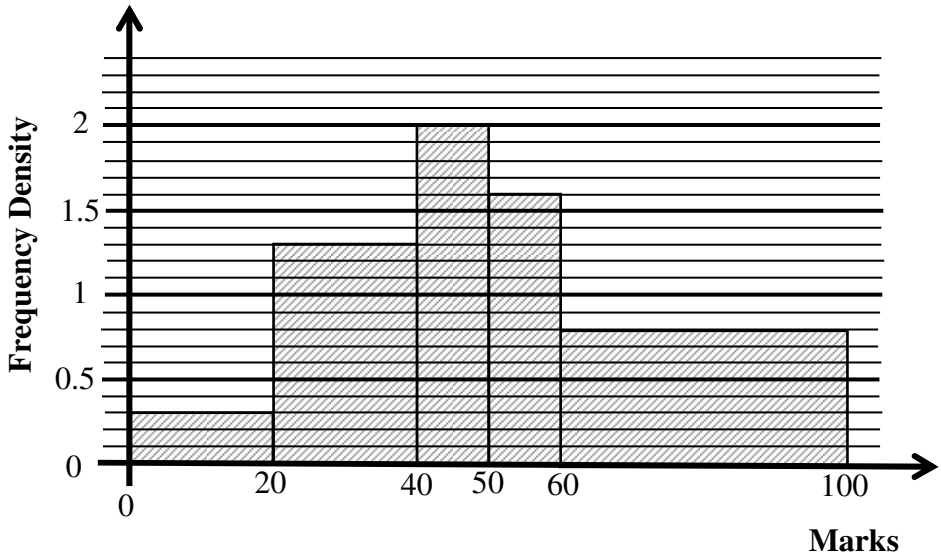
$\boxed{\overline{x} \approx 18.5}$, $\boxed{\sigma \approx 4.33}$, $\boxed{Q_2 \approx 18.4}$, $\boxed{100\%}$



**Question 4**

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

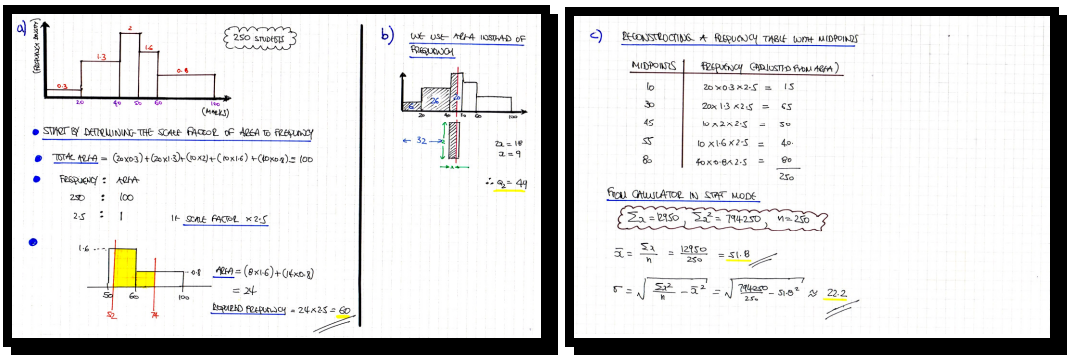



**a)** Estimate how many students scored between $52$ and $74$ marks.

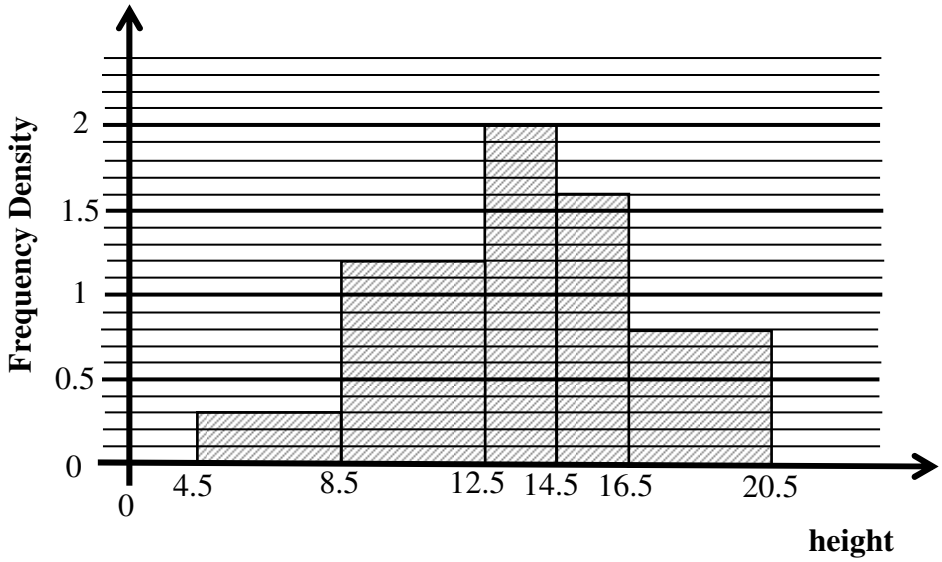
**b)** Use the histogram estimate the median.

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

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



**Question 5**

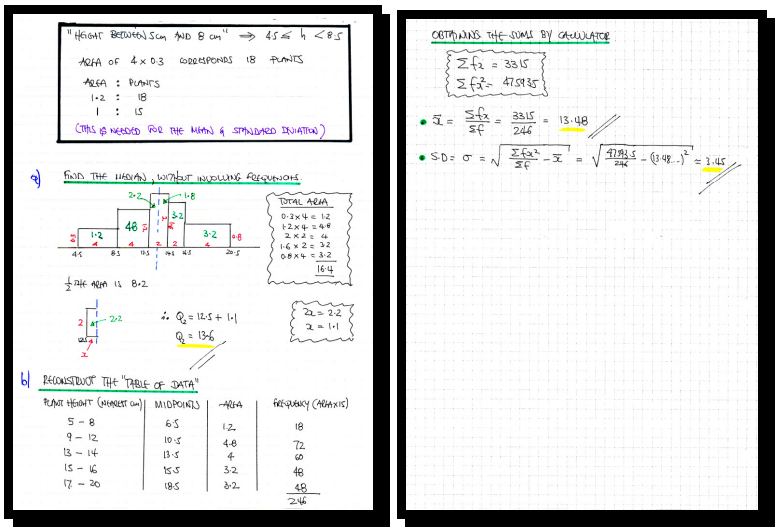



The histogram above shows the distribution of the heights, to the nearest cm, of some plants in a garden centre. It is further given that there were 18 plants with a height between 5 cm and 8 cm, rounded to the nearest cm.

**a)** Use the histogram to estimate the median.

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

$\boxed{\text{median} \approx 13.6}$, $\boxed{\overline{x} \approx 13.48}$, $\boxed{\sigma \approx 3.45}$



**Question 6**

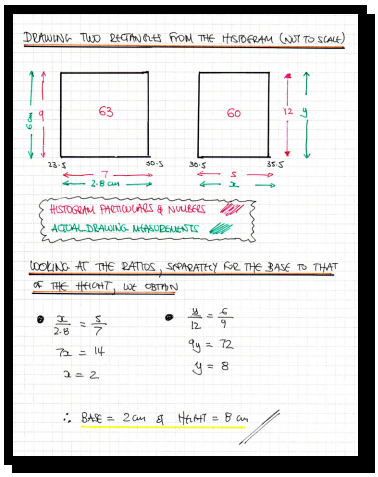
In a histogram the weights of baby hamsters, correct to the nearest gram, are plotted on the $x$ axis.

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 $31-35$.

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



**Question 7**

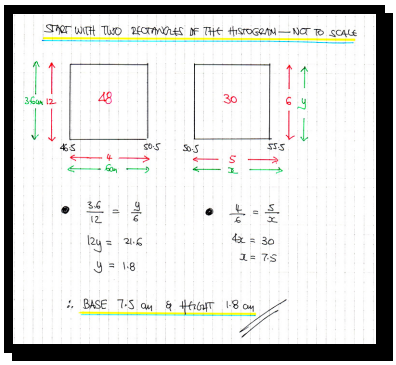
In a histogram the commuting times of a group of individuals, correct to the nearest minute, are plotted on the $x$ axis.

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 $51-55$.

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



**Question 8**

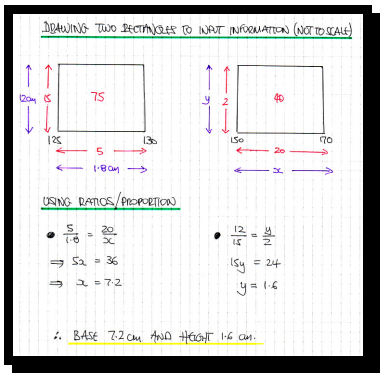
In a histogram the weights of apples, $W$ grams, are plotted on the $x$ axis.

In this histogram the class $125 \leq W < 130$ 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 $150 \leq W < 170$ has a frequency of $40$.

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

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



**Question 9**

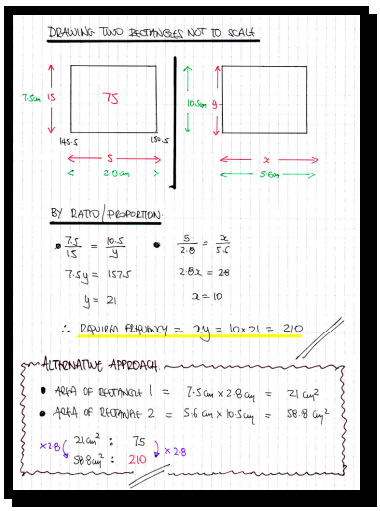
In a histogram the weights of peaches, correct to the nearest gram, are plotted on the $x$ axis.

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.

$$\boxed{f=210}$$



**Question 10**

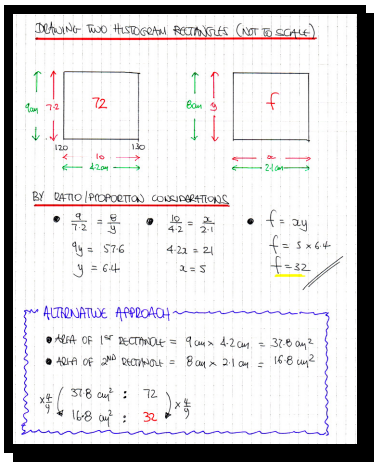
In a histogram the heights, $h$ cm, of primary school pupils are plotted on the $x$ axis.

In this histogram the class $120 \leq h < 130$ 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.

$$\boxed{f = 32}$$



# DATA CODING

**Question 1**

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

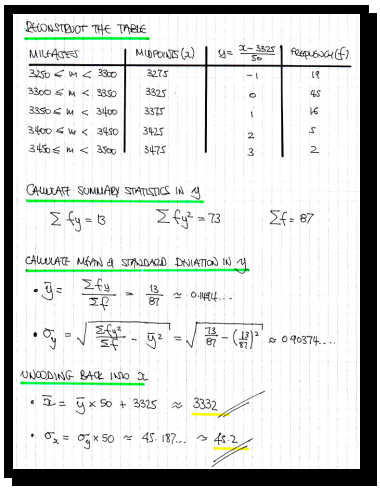
| Mileages ($m$) | Frequency |
|---|---|
| $3250 \leq m < 3300$ | 19 |
| $3300 \leq m < 3350$ | 45 |
| $3350 \leq m < 3400$ | 16 |
| $3400 \leq m < 3450$ | 5 |
| $3450 \leq m < 3500$ | 2 |

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.

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



**Question 2**

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

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

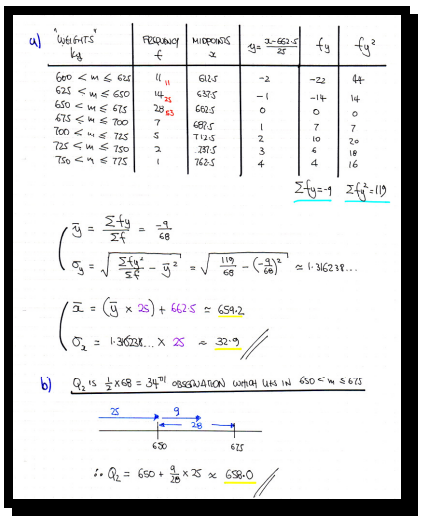
**a)** By using the coding

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

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.

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



**Question 3**

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

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

**a)** By using the coding

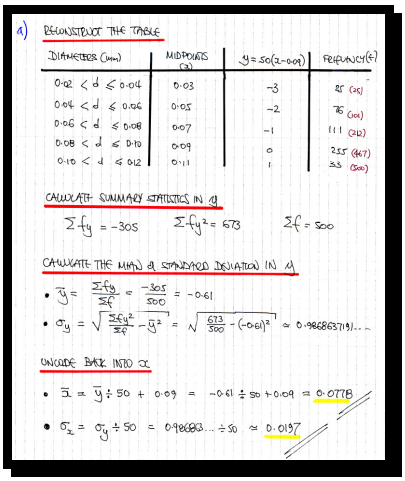
$$y = 50(x - 0.09),$$

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.

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



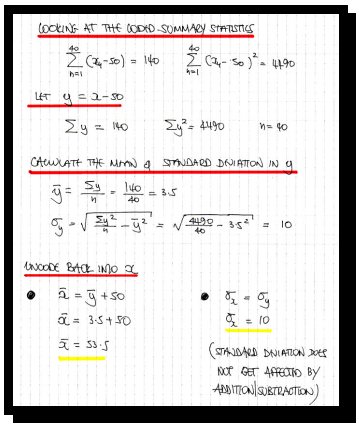
**Question 4**

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

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

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

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



**Question 5**

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

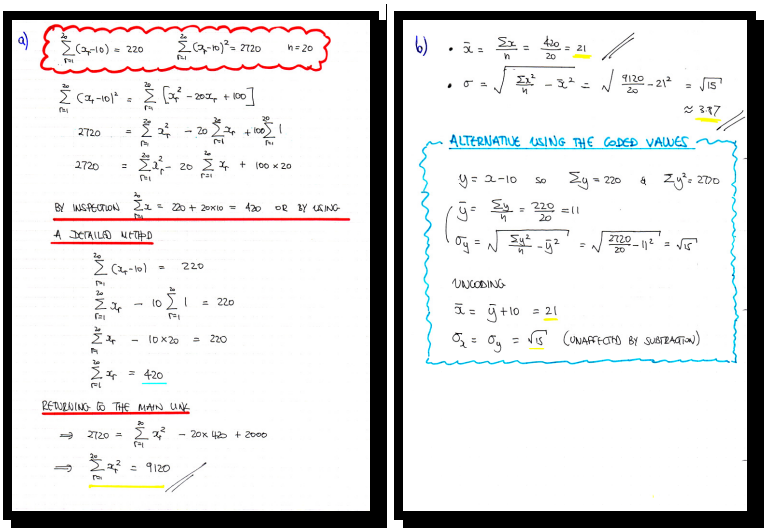
$$\sum (x-10) = 220 \qquad \text{and} \qquad \sum (x-10)^2 = 2720 .$$

**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.

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



**Question 6**

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 $x$.

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

