EXERCISES [MAA 4.1-4.3]

STATISTICS – BASIC CONCEPTS

Compiled by Christos Nikolaidis

O.   Practice questions

1.   [Maximum mark: 6]   *[without GDC]*

Indicate by the words "**discrete**" or "**continuous**" the type of data below:

[handwritten: Continuous data allows for infinitely small increments, while discrete data doesn't.]

Number of children in a family in Rome	*discrete*
Height of Greek men	*continuous*
Time spent in a supermarket	*discrete* *continuous*
Final grade in IB exams	*continuous* *discrete*
100m sprint time	*discrete* *continuous*
100m sprint time to the nearest second	*continuous* *discrete*

2.   [Maximum mark: 4]   *[without GDC]*

A population of 20 000 people consists of 15 000 men and 5 000 women. We need a sample of 100 people. Match the following:

	METHOD		EXAMPLE
1	**random sampling**	A	Select 75 men and 25 women.
2	**Systematic sampling**	B	Consider two groups: smokers and non-smokers. Select 50 from each group.
3	**Stratified sampling**	C	Select 100 people out of a hat
4	**Quota sampling**	D	Arrange names in a row. Pick every 200th person

....

....

....

3. [Maximum mark: 8]   *[with GDC]*

Consider the set of data   6, 2, 3, 8, 6, 5, 7, 6, 2.

$Q_1 = 2.5$
$Q_3 = 6.5$

(a) Complete the following tables

measures of central tendency	
mean	5
median	6
mode	6

measures of spread	
range	6
IQR	4
standard deviation	2.05...

[6]

(b) Find the variance. [2]

$\sigma^2 \cong 4.22$

....

....

4. [Maximum mark: 10]   *[without GDC]*

Consider the set of data   6, 2, 3, 8, 6, 5, 7, 6, 2.

By showing your working, find

(a) the **mean** [3]

(b) the **median** [2]

(c) the **mode** [1]

(d) (i) the lower quartile Q_1,   (ii) The upper quartile Q_3. [4]

....

....

....

....

....

....

....

....

....

....

....

....

5. [Maximum mark: 17]   *[with GDC]*

Consider the following frequency table for 200 entries of x:

x	f
4	20
5	34
6	60
7	16
8	20
9	30
10	20

(a) Complete the following tables

measures of central tendency	
mean	
median	
mode	

measures of spread	
range	
IQR	
standard deviation	

[6]

(b) Find the variance. [2]

....

....

6. [Maximum mark: 4]   *[without GDC]*

Consider the following frequency table for 100 entries of x:

x	f
4	10
5	30
6	60

By showing your working, find the mean.

....

....

....

....

....

7. [Maximum mark: 16]   *[without GDC]*

Consider the following frequency table for 200 entries of x:

x	f
4	20
5	34
6	60
7	16
8	20
9	30
10	20

54
114
130
150

By showing your working, find

(a)   the **median**   [2]

(b)   the **mode**   [1]

(c)   (i)   the lower quartile Q_1,   (ii)   The upper quartile Q_3.   [4]

(d)   Find the percentage of data that are less than or equal to 5.   [2]

(e)   Find   (i)   the 20th percentile.   (ii)   the 65th percentile.   [4]

(f)   Draw a box and whiskers plot.   [3]

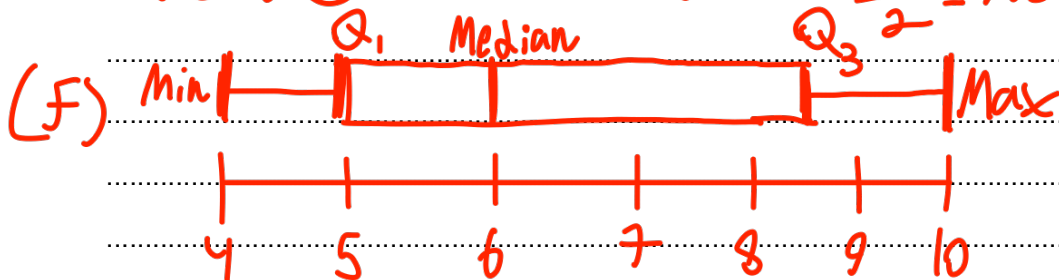(a) $\dfrac{201}{2} = 100.5$ (b) 6, highest frequency

$\dfrac{f(100) + f(101)}{2} = 6$

(c)(i) $\dfrac{1}{4}(200) = 50$; $Q_1 = f(50) = 5$

(ii) $\dfrac{3}{4}(200) = 150$; $Q_3 = f(150) = 8$

(d) $\dfrac{54}{200}$

(e)(i) $200 \cdot 0.20 = 40$,   (ii) $200 \cdot 0.65 = 130$

$f(40) = 5$ $f(130) = \dfrac{7+8}{2} = 7.5$

(f)

Min Q_1 Median Q_3 Max

4   5   6   7   8   9   10

8. [Maximum mark: 7]   *[with / without GDC]*

Consider the data presented in the **stem and leaf** diagram below:

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

Key 5 | 1 represents 51.

(a) Write down the first four entries (i.e. the smallest ones)   [1]

(b) Find the measures of central tendency: **mode** and **median,**   [2]

(c) Find the measures of spread: **range** and **interquartile range**   [4]

....

....

....

....

....

....

....

9. [Maximum mark: 4]   *[without GDC]*

The following results give the heights of sunflowers in centimetres.

180 184 195 177 175 173 169 167 197 173 166 183 161 195 177

192 161 165

Represent the data by a stem and leaf diagram.

....

....

....

....

....

....

....

....

10. [Maximum mark: 12]   *[without GDC]*

Consider the data   10, 20, 30, 40   with

mean = 25 standard deviation = $5\sqrt{5}$ variance = 125

Find the new mean, standard deviation and variance in the following cases

	mean	standard deviation	variance
if each number is increased by 2	27	$5\sqrt{5}$	125
if each number is multiplied by 2	50	$10\sqrt{5}$	500
if each number is increased by a	$25+a$	$5\sqrt{5}$	125
if each number is multiplied by a	$25a$	$5a\sqrt{5}$	$125a^2$

11. [Maximum mark: 6]   *[with GDC]*

Consider the following data

Class interval (number of words)	Frequency f
1–5	16
6–10	28
11–15	26
16–20	14

3
8
13
18

(a)   Find the mean.   [2]

(b)   Write down the modal group   [1]

(c)   Find the standard deviation and the variance   [3]

(a) ≈ 10.3

(b) $6-10$

(c) $\sigma = 4.9, \ \sigma^2 = 24.1$

A. Exam style questions (SHORT)

12. [Maximum mark: 6]   *[with GDC]*

A sample of discrete data is drawn from a population and given as

66, 72, 65 ,70, 69 ,73, 65, 71, 75.

Find

(a) the interquartile range;   [2]

(b) the mean of the population;   [1]

(c) the variance of the population.   [3]

....

....

....

....

....

13. [Maximum mark: 8]   *[with GDC]*

A random sample drawn from a large population contains the following data

6.2, 7.8, 12.1, 9.7, 5.2, 14.8, 16.2, 3.7 .

(a) Write down

    (i) the mean;   (ii) the median   [2]

(b) Find the variance.   [2]

(c) Find

    (i) the interquartile range   (ii) any outliers   [4]

(c) $Q_1 = 5.7$, $Q_3 = 13.45 \approx 13.5$

(i) $I_q = 13.45 - 5.7 = 7.75$

(ii) $LO = 5.7 - 1.5(7.75) = -5.925$

$UO = 25.075$

No values < -5.925 or > 25.075, hence no outliers.

....

....

....

14. [Maximum mark: 6]   *[without GDC]*

Three positive integers a, b, and c, where $a < b < c$, are such that their median is 11, their mean is 9 and their range is 10. Find the value of a.

$$b = 11$$
$$c - a = 10, \quad c = 10 + a$$
$$\frac{a + b + c}{3} = 9$$
$$a + 11 + 10 + a = 27, \quad 2a = 6, \quad a = 3$$

15. [Maximum mark: 6]   *[without GDC]*

Let a, b, c and d be integers such that $a < b$, $b < c$ and $c = d$.

The mode of these four numbers is 11. The range is 8. The mean is 8.

Calculate the value of each of the integers a, b, c, d.

$$a < b < c \qquad\qquad c = a + 8$$
$$c = d = 11 \qquad\qquad a + 8 = 11$$
$$a - c = a - d = 8 \qquad a = 3$$
$$\frac{a + b + c + d}{4} = 8 \qquad \frac{3 + b + 11 + 11}{4} = 8$$
$$b = 7$$

16. [Maximum mark: 6]   *[without GDC]*

Consider the four numbers a, b, c, d with $a \leq b \leq c \leq d$, where $a, b, c, d \in \mathbb{Z}$.

The mean of the four numbers is 4. The mode is 3. The median is 3. The range is 6.

Find the value of a, of b, of c and of d.

....

....

....

....

....

....

....

....

17. [Maximum mark: 6]   *[with GDC]*

The population below is listed in ascending order.

$$5, 6, 7, 7, 9, 9, r, 10, s, 13, 13, t$$

The median of the population is 9.5. The upper quartile Q_3 is 13.

(a)   Write down the value of   (i) r (ii) s [4]

(b)   The mean of the population is 10. Find the value of t.   [2]

....

....

....

....

18. [Maximum mark: 6]   *[without GDC]*

A set of data is

$$18, 18, 19, 19, 20, 22, 22, 23, 27, 28, 28, 31, 34, 34, 36.$$

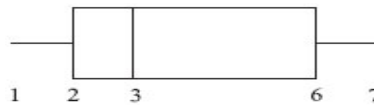The box and whisker plot for this data is shown below.

(a)   Write down the values of   A, B, C, D, E   [5]

(b)   Find the interquartile range.   [1]

....

....

....

19. [Maximum mark: 6]   *[without GDC]*

The box and whisker diagram shown below represents the marks received by 32 students.

(a)   Write down the value of the median mark.   [1]

(b)   Write down the value of the upper quartile.   [2]

(c)   Estimate the number of students who received a mark greater than 6.   [3]

....

....

....

....

20. [Maximum mark: 5]   *[without GDC]*

The following diagram is a box and whisker plot for a set of data.
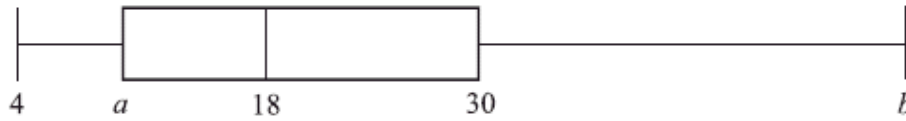
The interquartile range is 20 and the range is 40.

(a)   Write down the median value.   [1]

(b)   Find the value of   (i) a;   (ii) b.   [4]

....

....

....

21. [Maximum mark: 5]   *[without GDC]*

A scientist has 100 female fish and 100 male fish. She measures their lengths to the nearest cm. These are shown in the following box and whisker diagrams.

Female fish

Male fish
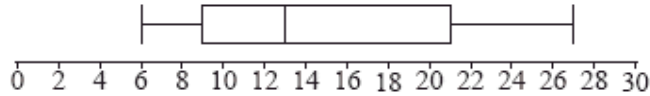
(a)   Find the range of the lengths of **all** 200 fish.   [3]

(b)   Four cumulative frequency graphs are shown below.
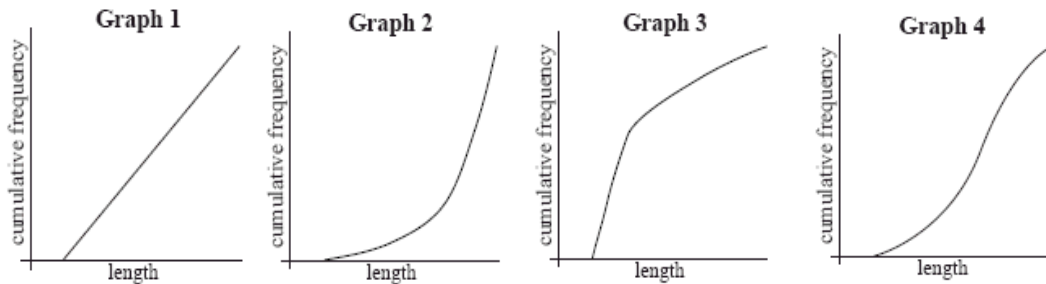
Which graph is the best representation of the lengths of the **female** fish?   [2]

....

....

....

22. [Maximum mark: 6]   *[with GDC]*

The 45 students in a class each recorded the number of whole minutes, x, spent doing experiments on Monday. The results are $\Sigma x = 2230$.

(a)   Find the mean number of min the students spent doing experiments on Monday.   [2]

Two new students joined the class and reported that they spent 37 minutes and 30 minutes respectively.

(b)   Calculate the new mean including these two students.   [4]

....

....

....

....

....

....

....

....

23. [Maximum mark: 4]   *[without GDC]*

The mean of the population $x_1, x_2, ..., x_{25}$ is m

(a)   Given that $\sum_{i=1}^{25} x_i = 300$, find the value of m .   [2]

(b)   Given that $m = 10$, find the value of $\sum_{i=1}^{25} x_i$.   [2]

....

....

....

....

....

....

....

....

....

24. [Maximum mark: 5]   *[without GDC]*

Consider the data set $\{k-2,\ k,\ k+1,\ k+4\}$, where $k \in \mathbb{R}$.

(a)   Find the mean of this data set in terms of k.   [3]

Each number in the above data set is now **decreased** by 3.

(b)   Find the mean of this **new** data set in terms of k.   [2]

....

....

....

....

....

....

....

....

25. [Maximum mark: 4]   *[with GDC]*

At a conference of 100 mathematicians there are 72 men and 28 women. The men have a mean height of 1.79 m and the women have a mean height of 1.62 m. Find the mean height of the 100 mathematicians.

....

....

....

....

....

26. [Maximum mark: 6]   *[with GDC]*

From January to September, the mean number of car accidents per month was 630. From October to December, the mean was 810 accidents per month.

What was the mean number of car accidents per month for the whole year?

....

....

....

....

....

27. [Maximum mark: 4]   *[with GDC]*

A machine produces packets of sugar. The weights in grams of thirty packets chosen at random are shown below.

Weight (g)	29.6	29.7	29.8	29.9	30.0	30.1	30.2	30.3
Frequency	2	3	4	5	7	5	3	1

Find

(a)   the mean of this sample;   [2]

(b)   the variance of this sample.   [2]

....

....

....

....

....

28. [Maximum mark: 6]   *[with / without GDC]*

In a sample of 50 boxes of light bulbs, the number of defective light bulbs per box is shown below.

Number of defective light bulbs per box	0	1	2	3	4	5	6
Number of boxes	7	3	15	11	6	5	3

(a)   Calculate the median number of defective light bulbs per box.   [3]

(b)   Calculate the mean number of defective light bulbs per box.   [3]

....

....

....

....

....

....

....

....

....

....

29. [Maximum mark: 7]   *[with / without GDC]*

A box contains 100 cards. Each card has a number between one and six written on it.

The following table shows the frequencies for each number.

Number	1	2	3	4	5	6
Frequency	26	10	20	k	29	11

(a) Calculate the value of k. [2]

(b) Find

(i) the median; (ii) the interquartile range. [5]

....

....

....

....

....

....

30. [Maximum mark: 4]   *[with / without GDC]*

Given the following frequency distribution,

Number (x)	1	2	3	4	5	6
Frequency (f)	5	9	16	18	20	7

find

(a) the median; [2]

(b) the mean. [2]

....

....

....

....

....

....

....

....

....

31. [Maximum mark: 6]   *[with GDC]*

A standard die is rolled 36 times. The results are shown in the following table.

Score	1	2	3	4	5	6
Frequency	3	5	4	6	10	8

(a)   Write down the standard deviation.   [2]

(b)   Write down the median score.   [1]

(c)   Find the interquartile range.   [3]

....

....

....

....

32. [Maximum mark: 6]   *[with / without GDC]*

The number of hours of sleep of 21 students are shown in the frequency table below.

Hours of sleep	Number of students
4	2
5	5
6	4
7	3
8	4
10	2
12	1

Find

(a)   the median;   [2]

(b)   the lower quartile;   [2]

(c)   the interquartile range.   [2]

....

....

....

....

....

....

....

33. [Maximum mark: 6]]   *[without GDC]*

The table below shows the marks gained in a test by a group of students.

Mark	1	2	3	4	5
Number of students	5	10	p	6	2

The median is 3 and the mode is 2. Find the **two** possible values of p.

$$15 < \frac{23+p+1}{2} \ , \ p > 6$$

$$p < 10$$

$$p = 8 \text{ or } 9$$

34. [Maximum mark: 7]   *[with GDC]*

The following table gives the examination grades for 120 students.

Grade	Number of students	Cumulative frequency
1	9	9
2	25	34
3	35	p
4	q	109
5	11	120

(a) Find the value of

(i) p;   (ii) q.   [4]

(b) Find the mean grade.   [2]

(c) Write down the standard deviation.   [1]

35. [Maximum mark: 7]   *[with GDC]*

In a school with 125 girls, each student is tested to see how many sit-up exercises (sit-ups) she can do in one minute. The results are given in the table below.

Number of sit-ups	Number of students	Cumulative number of students
15	11	11
16	21	32
17	33	p
18	q	99
19	18	117
20	8	125

(a) (i)   Write down the value of p (ii)   Find the value of q.   [3]

(b)   Find the median number of sit-ups.   [2]

(c)   Find the mean number of sit-ups.   [2]

....

....

....

....

36. [Maximum mark: 6]   *[with GDC]*

The following table shows the mathematics marks scored by students.

Mark	1	2	3	4	5	6	7
Frequency	0	4	6	k	8	6	6

The mean mark is 4.6.

(a)   Find the value of k.   [5]

(b)   Write down the mode.   [1]

....

....

....

....

....

....

....

....

37. [Maximum mark: 4]   *[with GDC]*

The table shows the scores of competitors in a competition.

Score	10	20	30	40	50
Number of competitors with this score	1	2	5	k	3

The mean score is 34. Find the value of k.

....

....

....

....

....

....

38. [Maximum mark: 6]   *[without GDC]*

The following stem and leaf diagram gives the heights in cm of 39 schoolchildren.

Stem	Leaf	Key 13 │ 2 represents 132 cm.
13	2, 3, 3, 5, 8,	
14	1, 1, 1, 4, 5, 5, 9,	
15	3, 4, 4, 6, 6, 7, 7, 7, 8, 9, 9,	
16	1, 2, 2, 5, 6, 6, 7, 8, 8,	
17	4, 4, 4, 5, 6, 6,	
18	0,	

(a)   State for the height

(i)   the lower quartile,   (ii) the median   (iii) the upper quartile.   [3]

(b)   Draw a box-and-whisker plot of the data on the diagram below.

120   130   140   150   160   170   180   190

height in cm   [3]

....

....

....
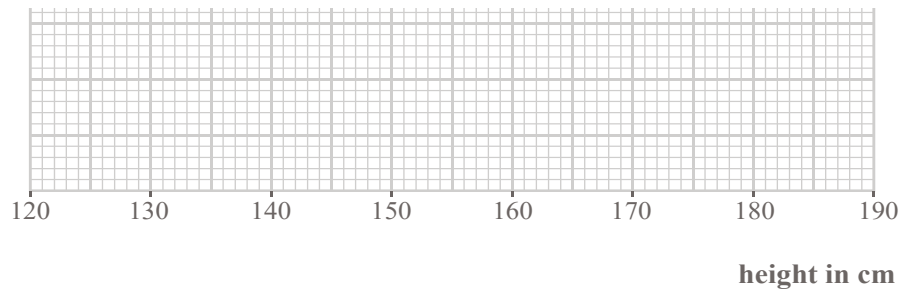
39. [Maximum mark: 4]   *[with GDC]*

A sample of 70 batteries was tested to see how long they last. The results were:

Time (hours)	Number of batteries (frequency)
$0 \le t \le 10$	2
$10 \le t \le 20$	4
$20 \le t \le 30$	8
$30 \le t \le 40$	9
$40 \le / \le 50$	12
$50 \le t \le 60$	13
$60 \le t \le 70$	8
$70 \le t \le 80$	7
$80 \le t \le 90$	6
$90 \le t \le 100$	1
Total	**70**

Find   (i)   the mean;   (ii)   the standard deviation.

....

....

40. [Maximum mark: 5]   *[without GDC]*

The 80 applicants for a Sports Science course were required to run 800 metres and their times were recorded. The results were used to produce the following cumulative frequency graph.
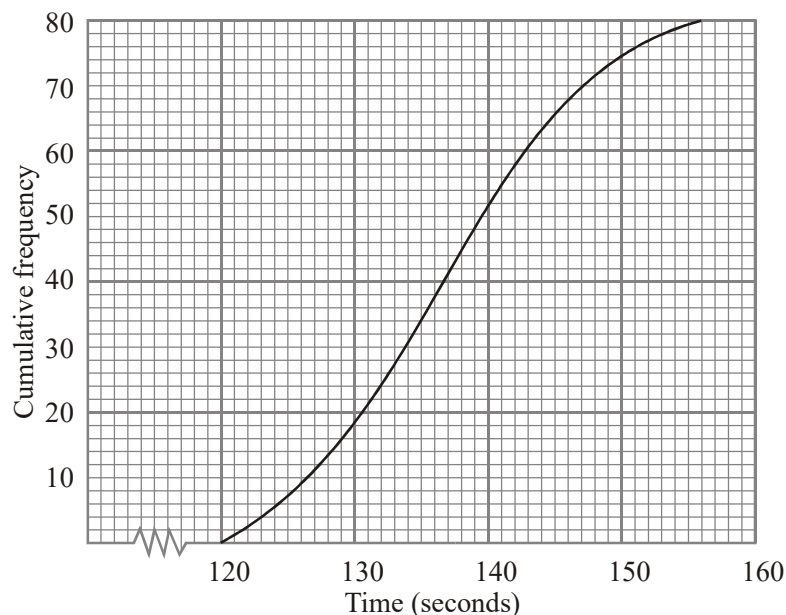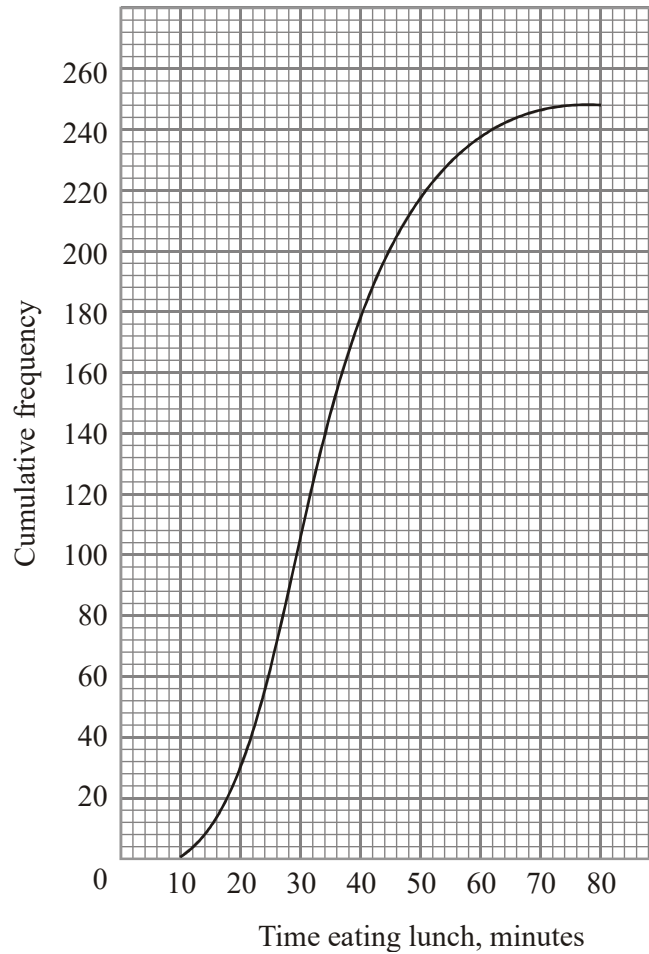

(a)   Estimate the median;   [2]

(b)   Estimate the interquartile range.   [3]

....

....

41. [Maximum mark: 6]   *[with / without GDC]*

The cumulative frequency curve below indicates the amount of time 250 students spend eating lunch.

(a) Estimate the number of students who spend between 20 and 40 minutes eating lunch.   [2]

(b) Estimate the 20th percentile.   [2]

(c) If 20 % of the students spend more than x minutes eating lunch, estimate the value of x.   [2]

....

....

....

....

....

....

....
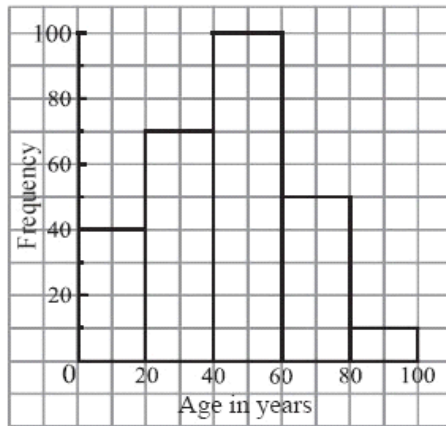
42. [Maximum mark: 6]   *[with GDC]*

The histogram below represents the ages of 270 people in a village.

(a)   Use the histogram to complete the table below.   [2]

Age range	Frequency	Mid-interval value
0 < age < 20	40	10
20 ≤ age < 40		
40 ≤ age < 60		
60 ≤ age < 80		
80 ≤ age ≤100		

(b)   Hence, calculate an estimate of the mean age.   [4]

....

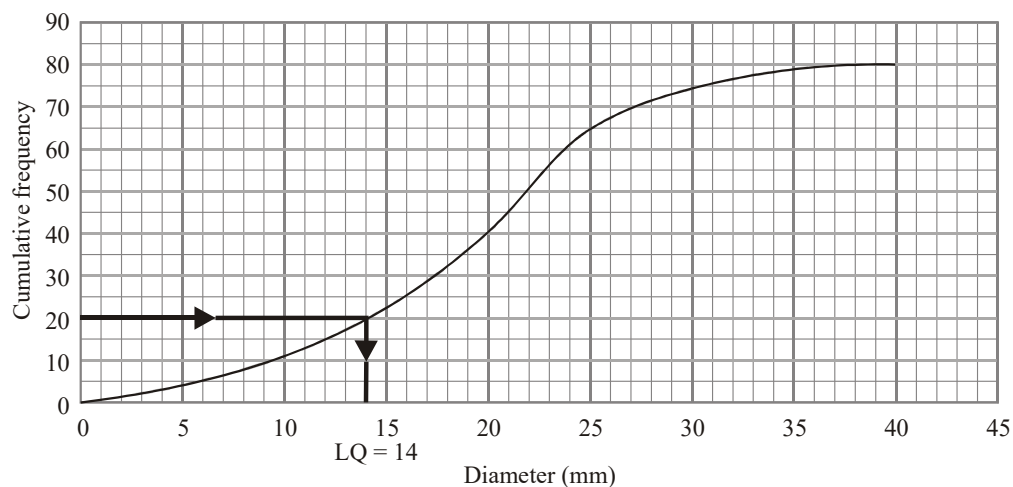...

43. [Maximum mark: 4]   *[without GDC]*

A student measured the diameters of 80 snail shells. His results are shown in the following cumulative frequency graph. The lower quartile (LQ) is 14 mm and is marked clearly on the graph.

(a)   On the graph, mark clearly in the same way and write down the value of

(i)   the median; (ii)   the upper quartile.   [2]

(b)   Write down the interquartile range.   [2]

....

....

44. [Maximum mark: 6]   *[without GDC]*

The cumulative frequency curve below shows the heights of 120 basketball players in centimetres.

Use the curve to estimate

(a)   the median height;   [2]

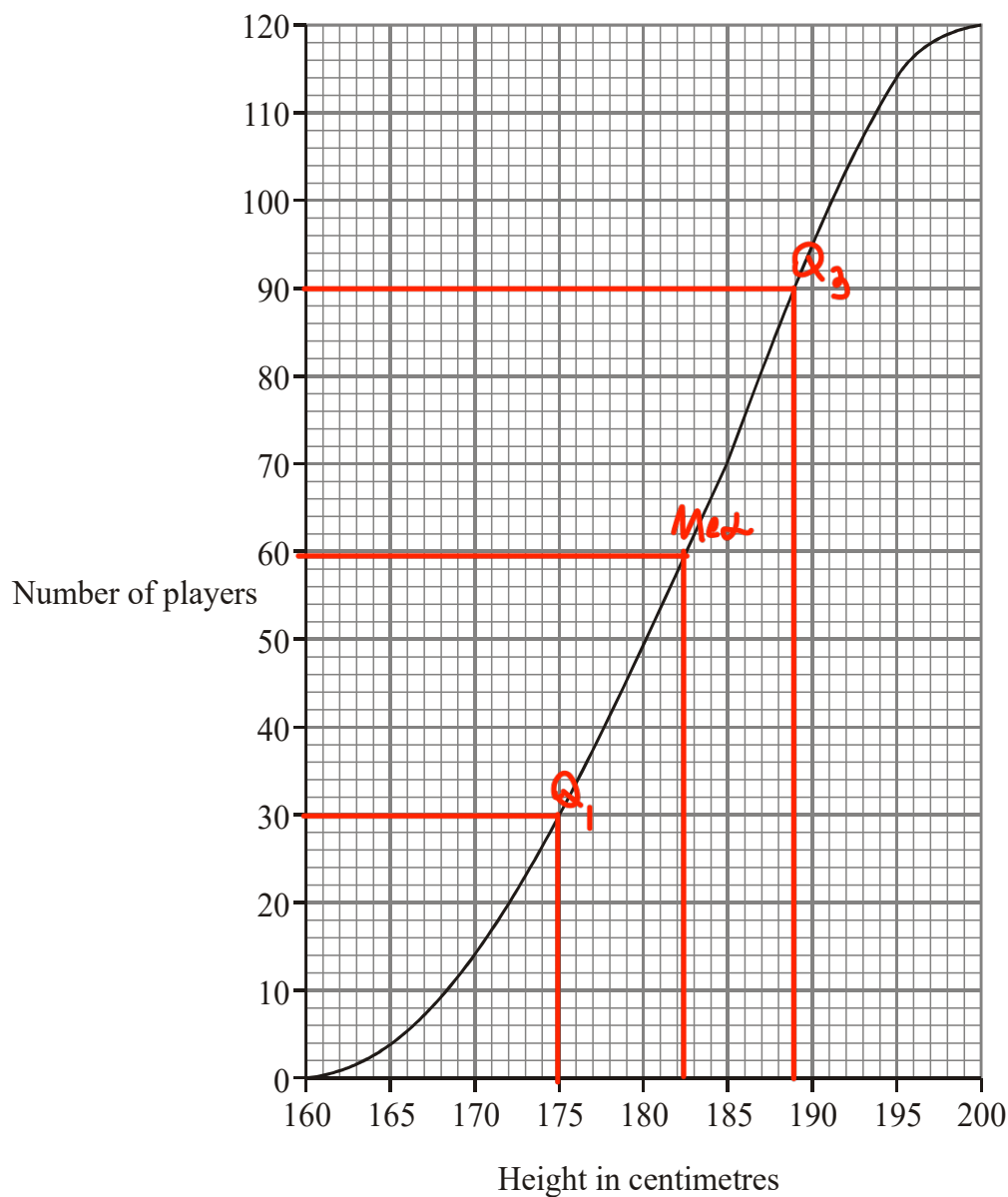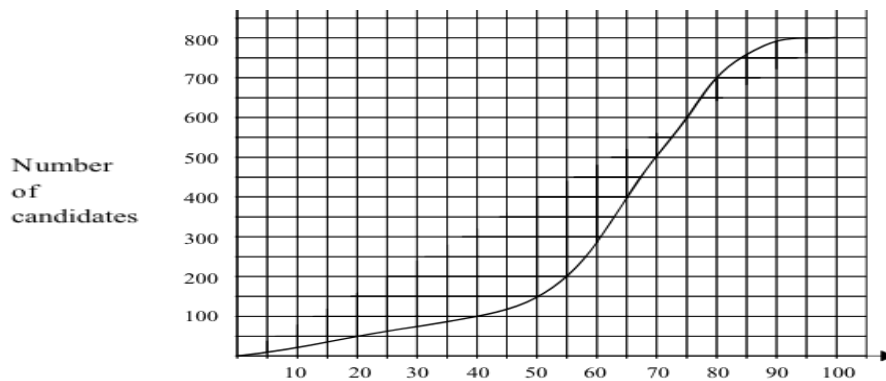(b)   the interquartile range.   [4]

(a) $182.5 \approx 183$ cm

(b) $Q_3 = 189, Q_1 = 175$

$189 - 175 = I_R = 14$

45. [Maximum mark: 6]   *[without GDC]*

A test marked out of 100 is written by 800 students. The cumulative frequency graph for the marks is given below.

Number of candidates

(a) Write down the number of students who scored 40 marks or less on the test.   [2]

(b) The middle 50 % of test results lie between marks a and b, $(a < b)$. Find a and b.   [4]

....

....

46. [Maximum mark: 6]   *[without GDC]*

The following is a cumulative frequency diagram for the time t, in minutes, taken by 80 students to complete a task.

cumulative frequency
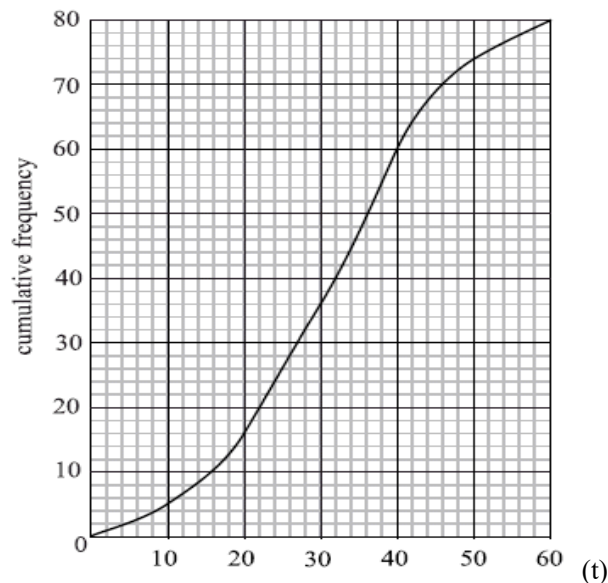
Time (minutes)	Number of students
$0 \leq t < 10$	5
$10 \leq t < 20$	11
$20 \leq t < 30$	20
$30 \leq t < 40$	24
$40 \leq t < 50$	14
$50 \leq t < 60$	6

(a) Write down the median.   [1]

(b) Find the interquartile range.   [3]

(c) Complete the frequency table next to the diagram.   [2]

....

....

47. [Maximum mark: 6]   *[with GDC]*

The cumulative frequency graph below shows the heights of 120 girls in a school.
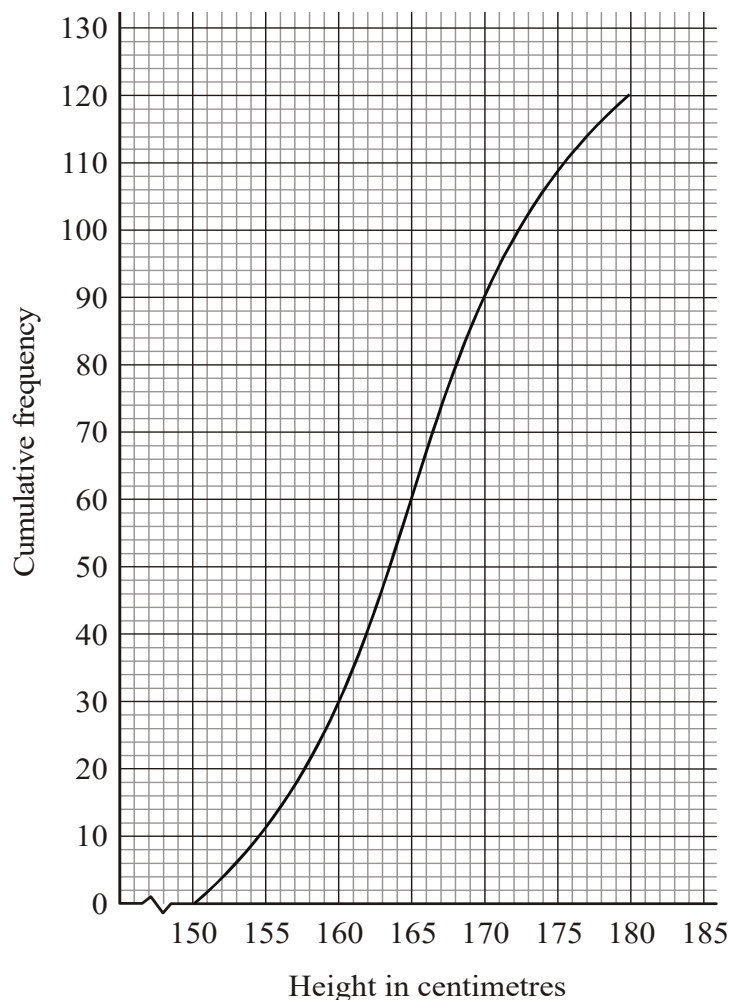
(a) Find
(i) the median;   (ii) the interquartile range.   [4]

(b) Given that 60% of the girls are taller than a cm, find the value of a.   [2]

....

....

....

....

....

....

....

48. [Maximum mark: 6]   *[without GDC]*

The cumulative frequency curve below shows the marks obtained in an examination by a group of 200 students.
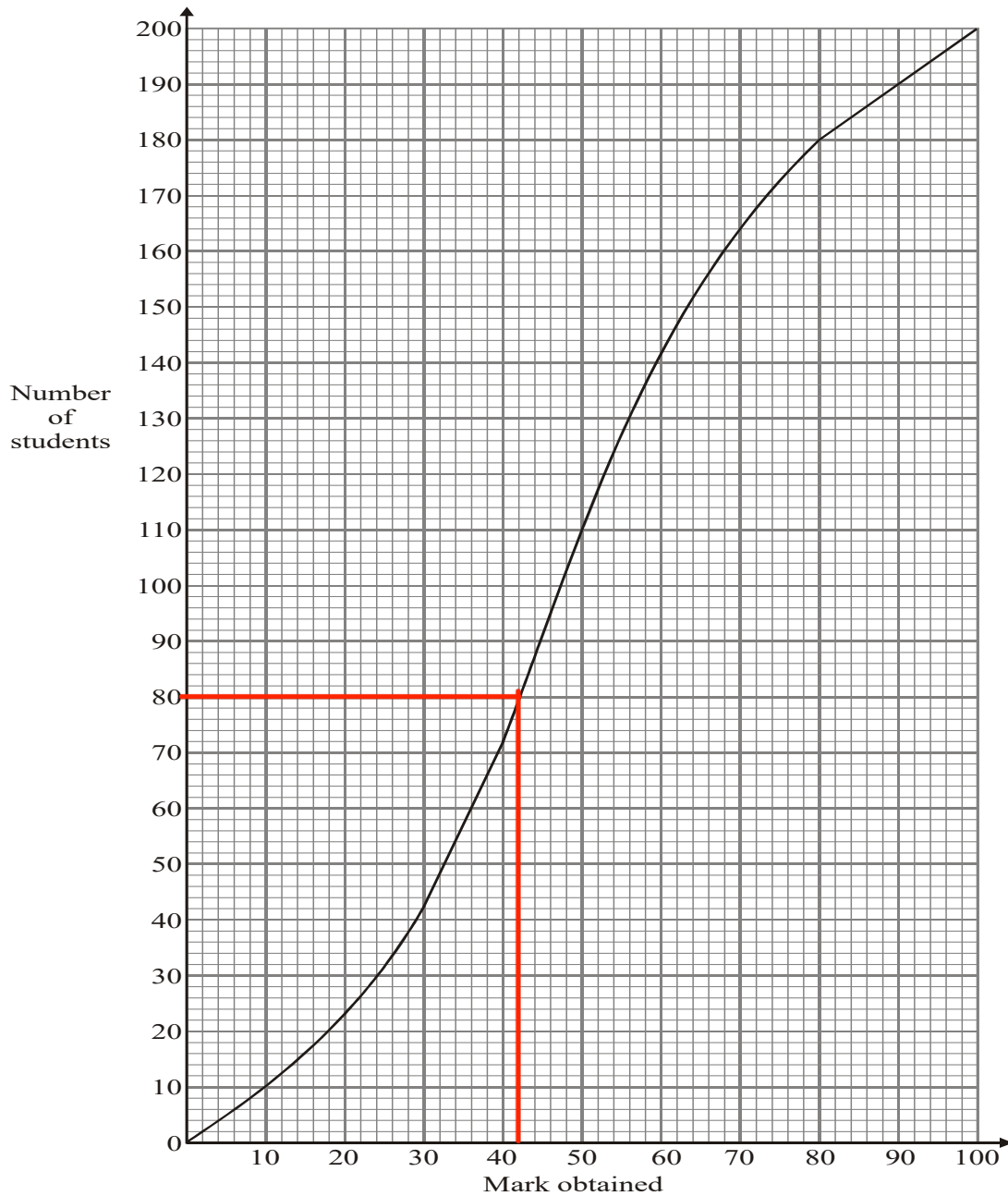
(a)   Use the cumulative frequency curve to complete the frequency table below.   [3]

Mark (x)	$0 \leq x < 20$	$20 \leq x < 40$	$40 \leq x < 60$	$60 \leq x < 80$	$80 \leq x < 100$
Number of students	22				20

(b)   Forty percent of the students fail. Find the pass mark.   [3]

Pass Mark = 42

49. [Maximum mark: 9]   *[with GDC]*

In the research department of a university, 300 mice were timed as they each ran through a maze. The results are shown in the cumulative frequency diagram below.

time t (seconds)
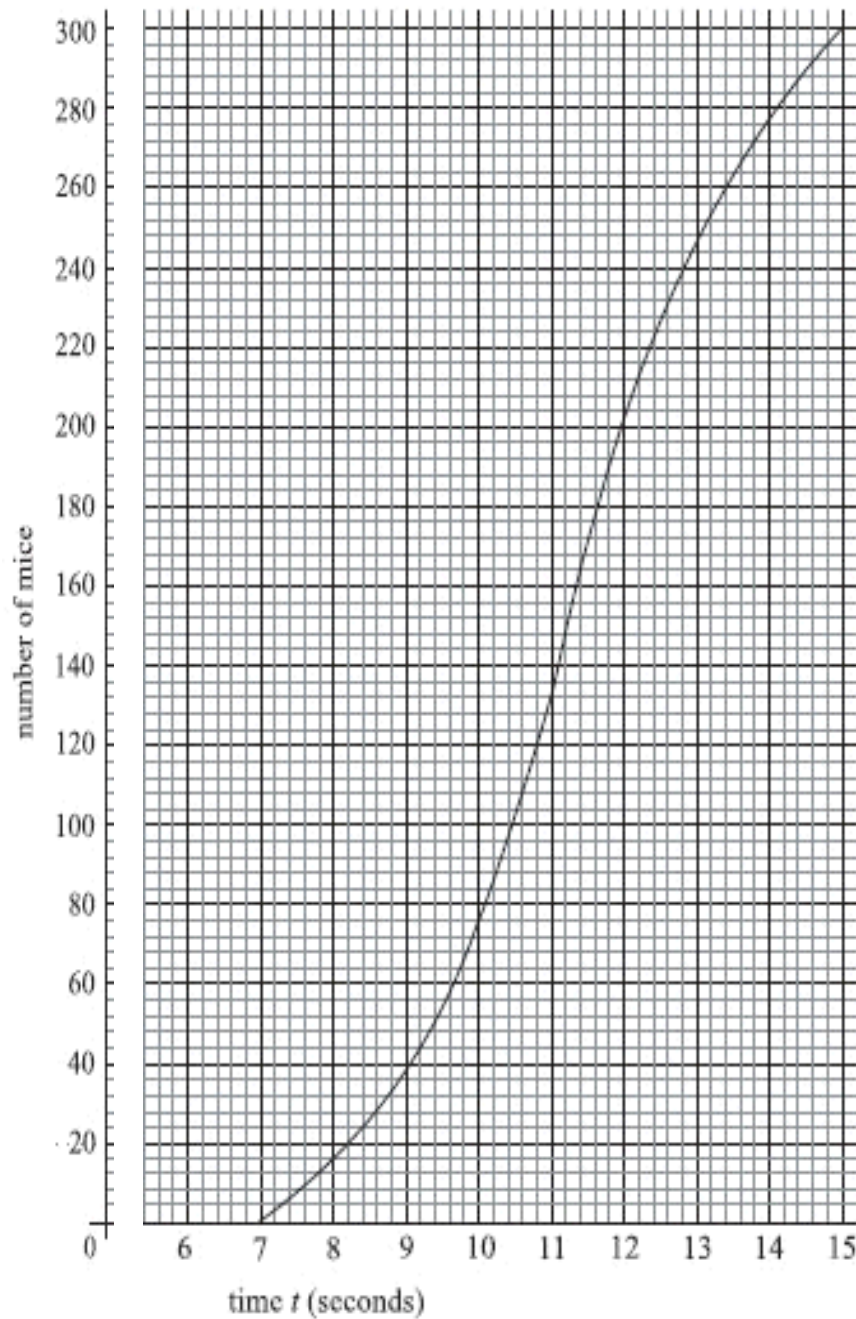
(a)   How many mice complete the maze

(i)   in less than 10 seconds?

(ii)   In more than 12 seconds?   [2]

(b)   Estimate

(i)   the median time.

(ii)   the 20th percentile.   [3]

(c)   Another way of showing the results is the frequency table below.

Time t (seconds)	Number of mice
$t < 7$	0
$7 \leq t < 8$	16
$8 \leq t < 9$	22
$9 \leq t < 10$	p
$10 \leq t < 11$	q
$11 \leq t < 12$	70
$12 \leq t < 13$	44
$13 \leq t < 14$	31
$14 \leq t \leq 15$	23

(i)   Find the value of p and the value of q.

(ii)   Calculate an estimate of the mean time.   [4]

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

50. [Maximum mark: 5]   *[without GDC]*

A recruitment company tests the aptitude of 100 applicants applying for jobs in engineering. Each applicant does a puzzle and the time taken, t, is recorded. The cumulative frequency curve for these data is shown below.
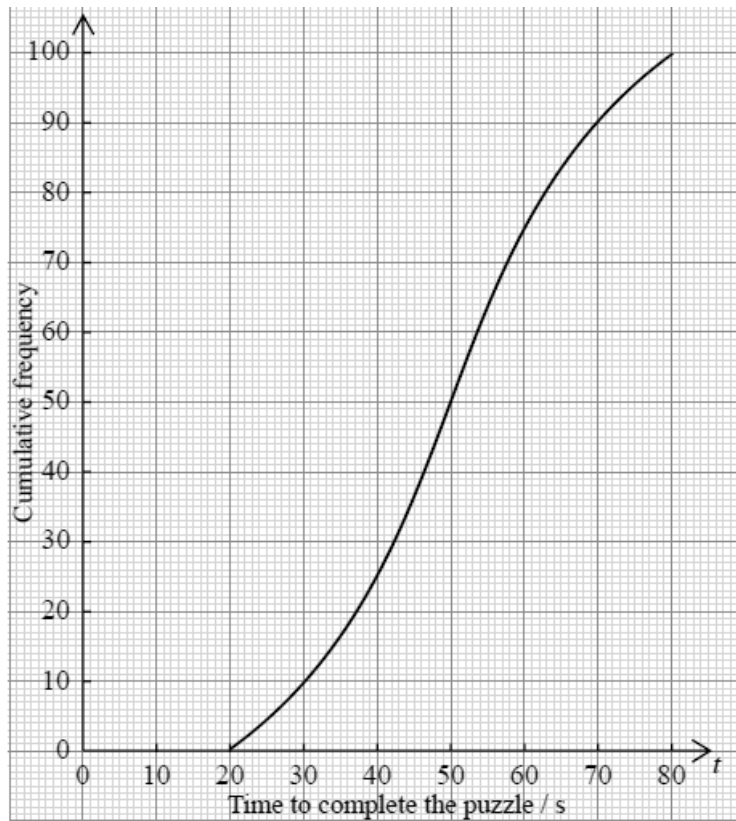
Using the cumulative frequency curve,

(a)   write down the value of the median;   [1]

(b)   determine the interquartile range;   [2]

(c)   complete the frequency table below.

Time to complete puzzle in seconds	Number of applicants
$20 < t \leq 30$	
$30 < t \leq 35$	
$35 < t \leq 40$	
$40 < t \leq 45$	
$45 < t \leq 50$	
$50 < t \leq 60$	
$60 < t \leq 80$	

[2]

....

....

....

51. [Maximum mark: 5]   *[with GDC]*

The following is the cumulative frequency diagram for the heights of 30 plants given in centimetres.
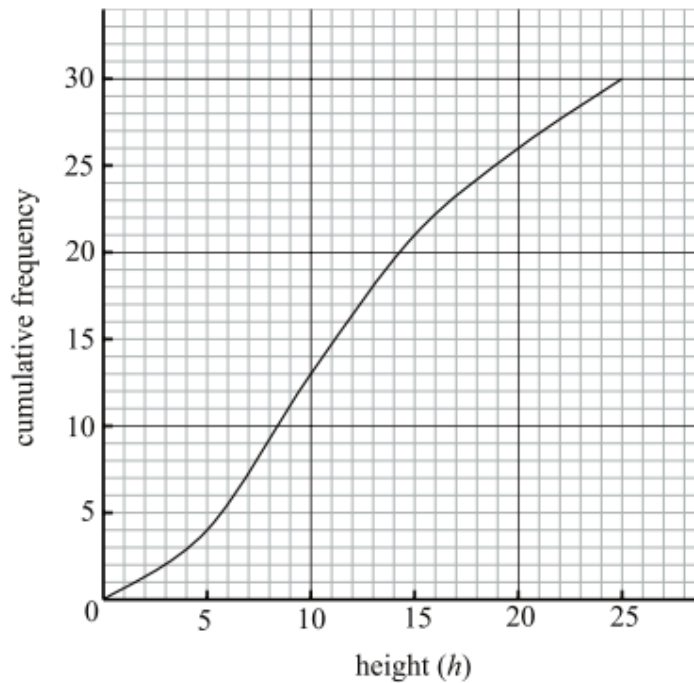
(a)   Use the diagram to estimate the median height.   [2]

(b)   Complete the following frequency table.

Height (h)	Frequency
$0 \le h < 5$	4
$5 \le h < 10$	9
$10 \le h < 15$	
$15 \le h < 20$	
$20 \le h < 25$	

[2]

(c)   Hence estimate the mean height.   [2]

....

....

....

....

....

....

....

52. [Maximum mark: 6]   *[without GDC]*

The four populations A, B, C and D are the same size and have the same range.

Frequency histograms for the four populations are given below.
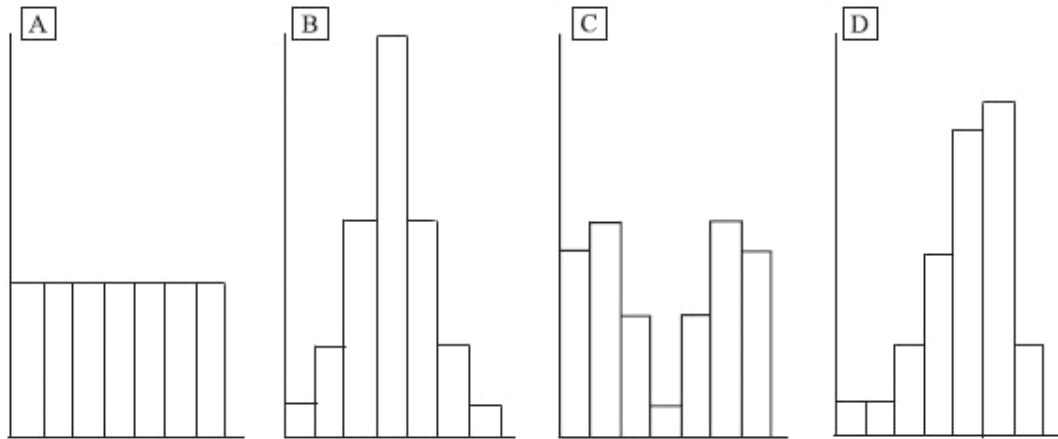
(a)   Each of the three box and whisker plots below corresponds to one of the four populations. Write the letter of the correct population under each plot.

D   B   C
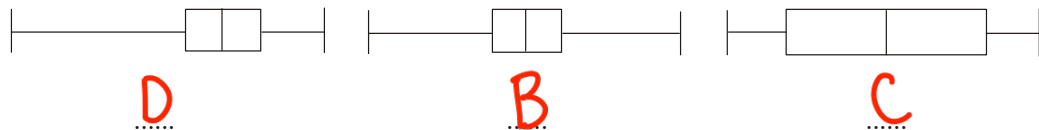
[3]

(b)   Each of the three cumulative frequency diagrams below corresponds to one of the four populations. Write the letter of the correct population under each diagram.

B   A   C
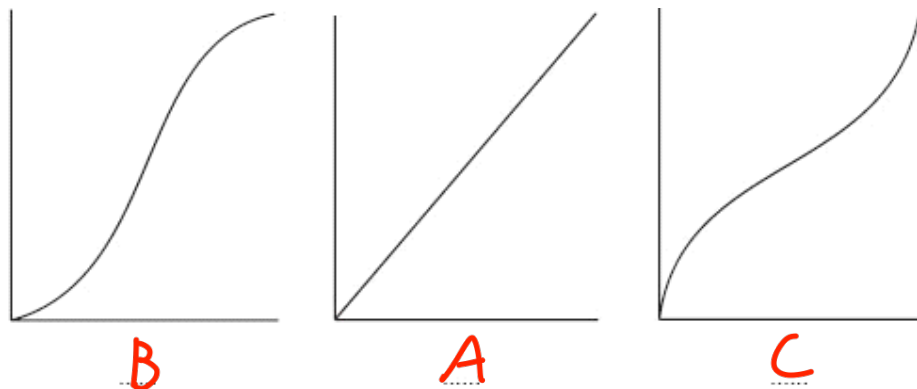
[3]

53. [Maximum mark: 4]   *[without GDC]*

The box-and-whisker plots shown represent the heights of female students and the heights of male students at a certain school.
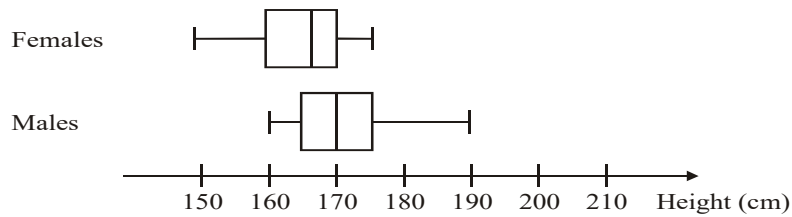
(a)   What percentage of female students are shorter than any male students?   [2]

(b)   What percentage of male students are shorter than some female students?   [2]

....

....

....

54. [Maximum mark: 10]   *[with GDC]*

The following diagram represents the lengths, in cm, of 80 plants grown in a laboratory.
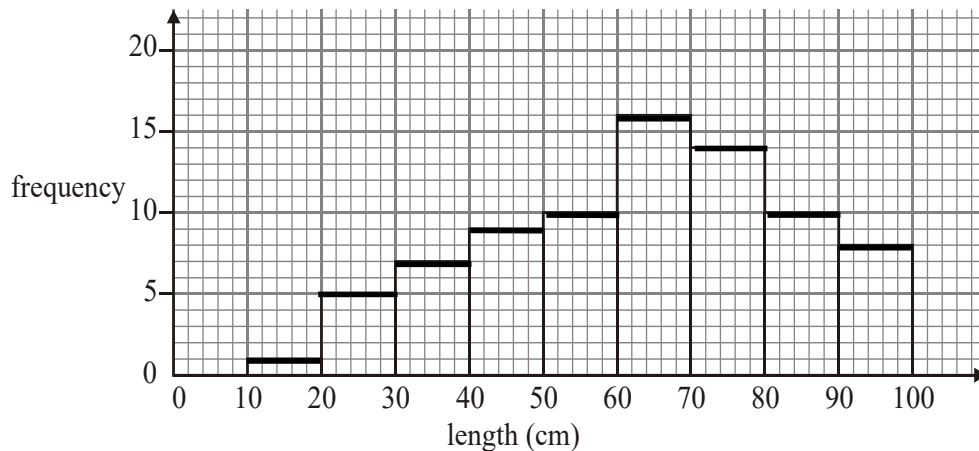
(a)   How many plants have lengths in cm between

(i)   50 and 60?   (ii)   70 and 90?   [2]

(b)   Calculate estimates for the mean and the st.deviation of the lengths of the plants.   [4]

(c)   Explain what feature suggests that the median is different than the mean.   [1]

....

....

....

....

....

B.   Exam style questions (LONG)

55.   [Maximum mark: 14]   *[with GDC]*

The following is the cumulative frequency curve for the time, t minutes, spent by 150 people in a store on a particular day.

time (t)
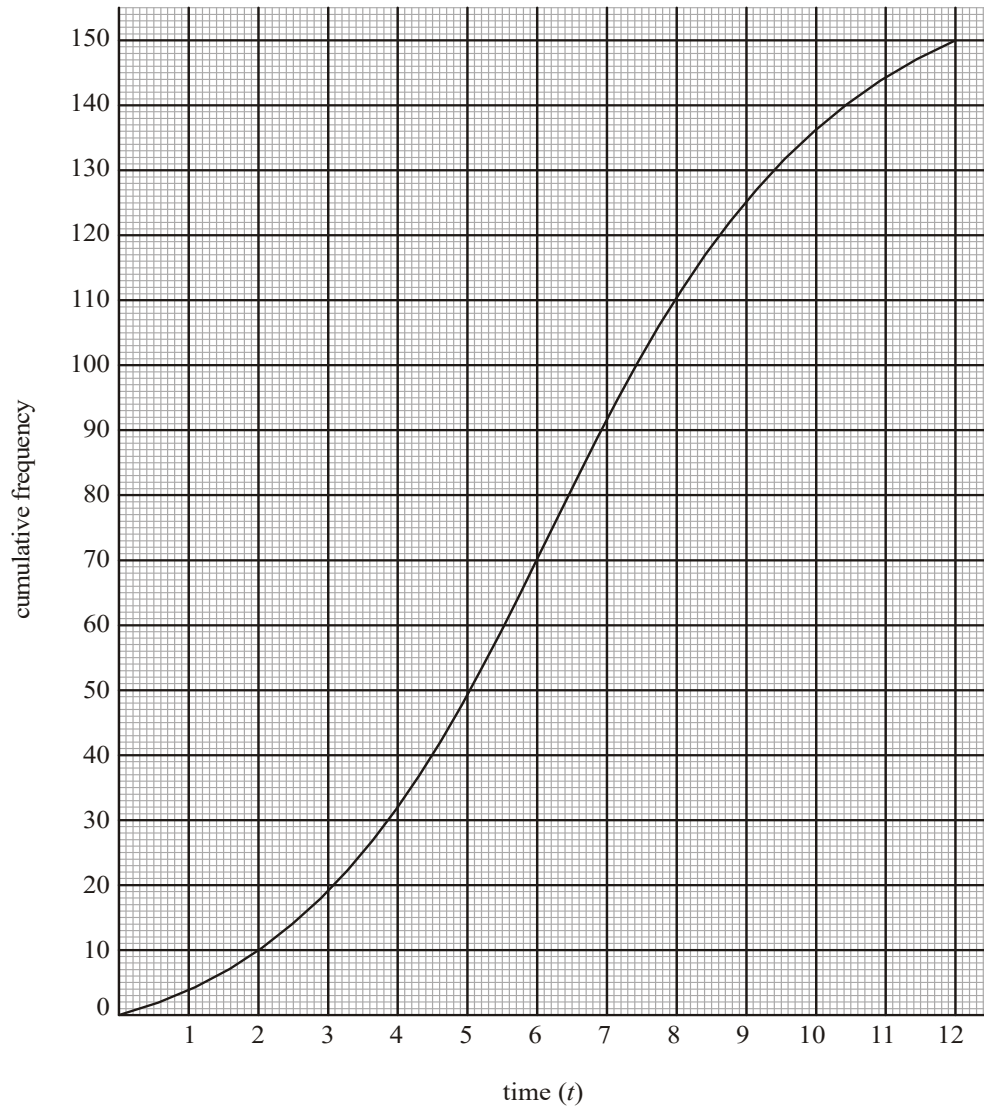
(a)   (i)   How many people spent less than 5 minutes in the store?

(ii)   Find the number of people who spent between 5 and 7 minutes in the store.

(iii)   Find the median time spent in the store.   [6]

(b)   Given that 40% of the people spent longer than k minutes, find the value of k.   [3]

(c)   (i) Complete the following frequency table.

t (minutes)	$0 \le t \le 2$	$2 < t \le 4$	$4 < t \le 6$	$6 < t \le 8$	$8 < t \le 10$	$10 < t \le 12$
Frequency	10	23				15

(ii)   Hence, calculate an estimate for the mean time spent in the store.   [5]

56. [Maximum mark: 11]   *[with GDC]*

The speeds in km h^{-1} of cars passing a point on a highway are recorded in the first table below.

Speed v	Number of cars
$v \le 60$	0
$60 < v \le 70$	7
$70 < v \le 80$	25
$80 < v \le 90$	63
$90 < v \le 100$	70
$100 < v \le 110$	71
$110 < v \le 120$	39
$120 < v \le 130$	20
$130 < v \le 140$	5
$v > 140$	0

Speed v	Cumulative frequency
$v \le 60$	0
$v \le 70$	7
$v \le 80$	32
$v \le 90$	95
$v \le 100$	a
$v \le 110$	236
$v \le 120$	b
$v \le 130$	295
$v \le 140$	300

(a) Calculate an estimate of the mean speed of the cars.   [2]

(b) The second table gives some of the cumulative frequencies.

   (i) Write down the values of a and b.

   (ii) On graph paper, construct a cumulative frequency **curve** to represent this information. Use a scale of 1 cm for 10 km h^{-1} on the horizontal axis and a scale of 1 cm for 20 cars on the vertical axis.   [5]

(c) Use your graph to determine

   (i) the percentage of cars travelling at a speed in excess of 105 km h^{-1};

   (ii) the speed which is exceeded by 15% of the cars.   [4]

....

....

....

....

....

....

....

....

....

....

....

57. [Maximum mark: 10]   *[without GDC]*

A survey is carried out to find the waiting times for 100 customers at a supermarket.

waiting time (seconds)	number of customers
0–30	5
30– 60	15
60– 90	33
90 –120	21
120–150	11
150–180	7
180–210	5
210–240	3

(a) Calculate an estimate for the mean of the waiting times, by using an appropriate approximation to represent each interval. [2]

(b) Construct a cumulative frequency table for these data [1]

(c) Use the cumulative frequency table to draw, on graph paper, a cumulative frequency graph, using a scale of 1 cm per 20 seconds waiting time for the horizontal axis and 1 cm per 10 customers for the vertical axis. [4]

(d) Use the cumulative frequency graph to find estimates for the median and the lower and upper quartiles. [3]

....

....

....

....

....

....

....

....

....

....

....

....

....

....

58. [Maximum mark: 15]   *[with GDC]*

In a suburb of a large city, 100 houses were sold in a three-month period. The following **cumulative frequency table** shows the distribution of selling prices (in thousands of dollars).

Selling price P ($1000)	$P \leq 100$	$P \leq 200$	$P \leq 300$	$P \leq 400$	$P \leq 500$
Total number of houses	12	58	87	94	100

(a) Represent this information on a cumulative frequency **curve**, using a scale of 1 cm to represent $50000 on the horizontal axis, 1 cm to represent 5 houses on the vertical axis. [4]

(b) Use your curve to find the interquartile range. [3]

The information above is represented in the following frequency distribution.

Selling price P ($1000)	$0 < P \leq 100$	$100 < P \leq 200$	$200 < P \leq 300$	$300 < P \leq 400$	$400 < P \leq 500$
Number of houses	12	46	29	a	b

(c) Find the value of a and of b. [2]

(d) Use mid-interval values to calculate an estimate for the mean selling price. [2]

(e) Houses which sell for more than $350000 are described as *De Luxe*.

   (i)   Use your graph to estimate the number of *De Luxe* houses sold. Give your answer to the nearest integer.

   (ii) Two *De Luxe* houses are selected at random. Find the probability that **both** have a selling price of more than $400000. [4]

....

....

....

....

....

....

....

....

....

....

....

59. [Maximum mark: 16]   *[with GDC]*

One thousand candidates sit an examination. The distribution of marks is shown in the following grouped frequency table.

Marks	1–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	81–90	91–100
Number of candidates	15	50	100	170	260	220	90	45	30	20

(a)   Complete the following table, which presents the above data as a cumulative frequency distribution.   [3]

Mark	≤10	≤20	≤30	≤40	≤50	≤60	≤70	≤80	≤90	≤100
Number of candidates	15	65					905			

(b)   Draw a cumulative frequency graph of the distribution, using a scale of 1 cm for 100 candidates on the vertical axis and 1 cm for 10 marks on the horizontal axis.   [5]

(c)   Use your graph to answer parts (i)–(iii) below,

   (i)   Find an estimate for the median score.   [2]

   (ii)   Candidates who scored less than 35 were required to retake the examination. How many candidates had to retake?   [3]

   (iii)   The highest-scoring 15% of candidates were awarded a distinction. Find the mark above which a distinction was awarded.   [3]

....

....

....

....

....

....

....

....

....

....

....

....

....

....

60*. [Maximum mark: 15]   *[with GDC]*

A supermarket records the amount of money d spent by customers in their store during a busy period. The results are as follows:

Money in $ (d)	0–20	20–40	40–60	60–80	80–100	100–120	120–140
Number of customers (n)	24	16	22	40	18	10	4

(a) Find an estimate for the mean amount of money spent by the customers, giving your answer to the nearest dollar ($). [2]

(b) Complete the following cumulative frequency table and use it to draw a cumulative frequency graph. Use a scale of 2 cm to represent $20 on the horizontal axis, and 2 cm to represent 20 customers on the vertical axis. [5]

Money in $ (d)	<20	<40	<60	<80	< 100	< 120	< 140
Number of customers (n)	24	40					

(c) The time t (minutes), spent by customers in the store may be represented by the equation $t = 2d^{\frac{2}{3}} + 3$.

   (i) Use this equation and your answer to part (a) to estimate the mean time in minutes spent by customers in the store. [3]

   (ii) Use the equation and the cumulative frequency graph to estimate the number of customers who spent more than 37 minutes in the store. [5]

....

....

....

....

....

....

....

....

....

....

....

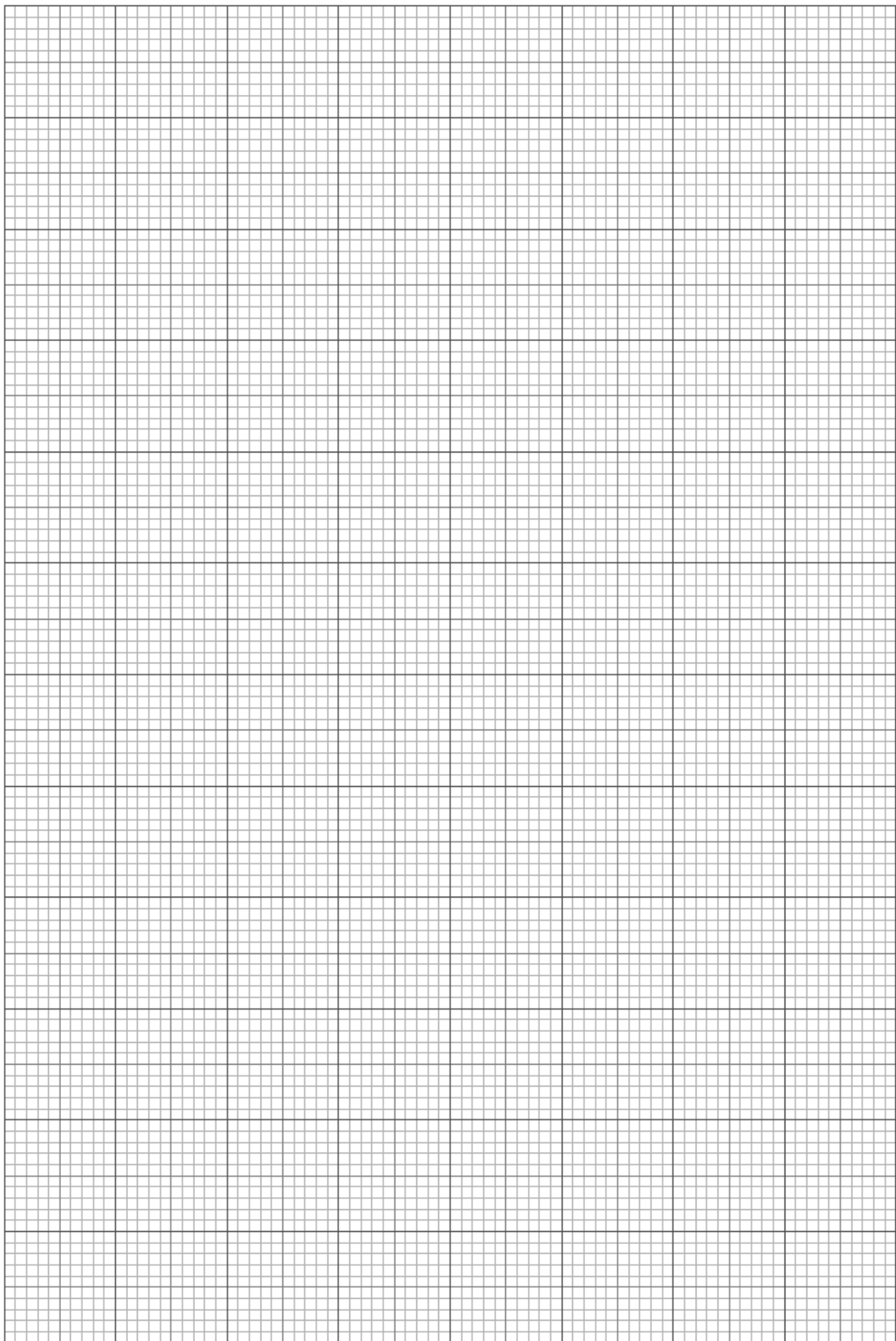...
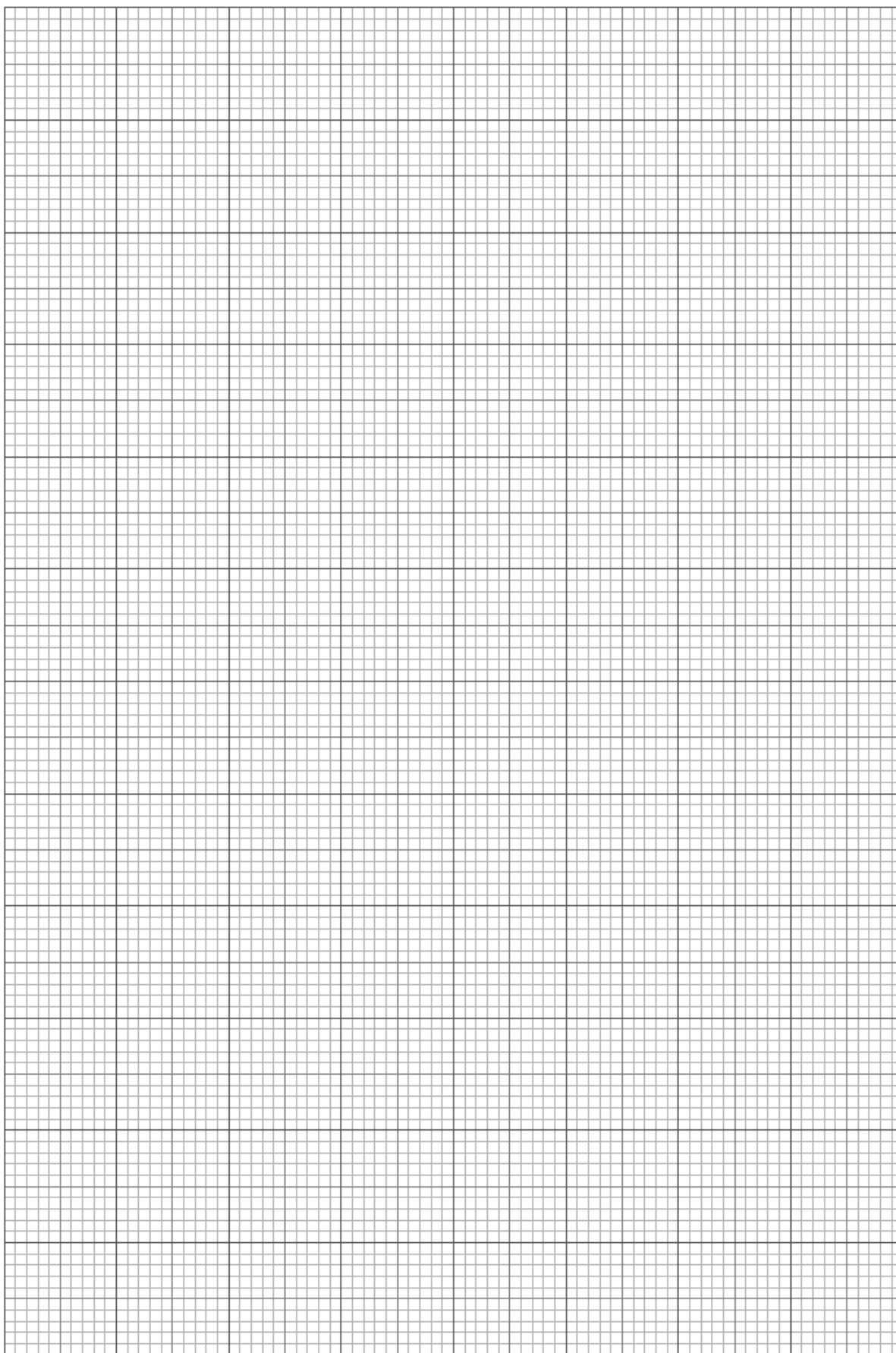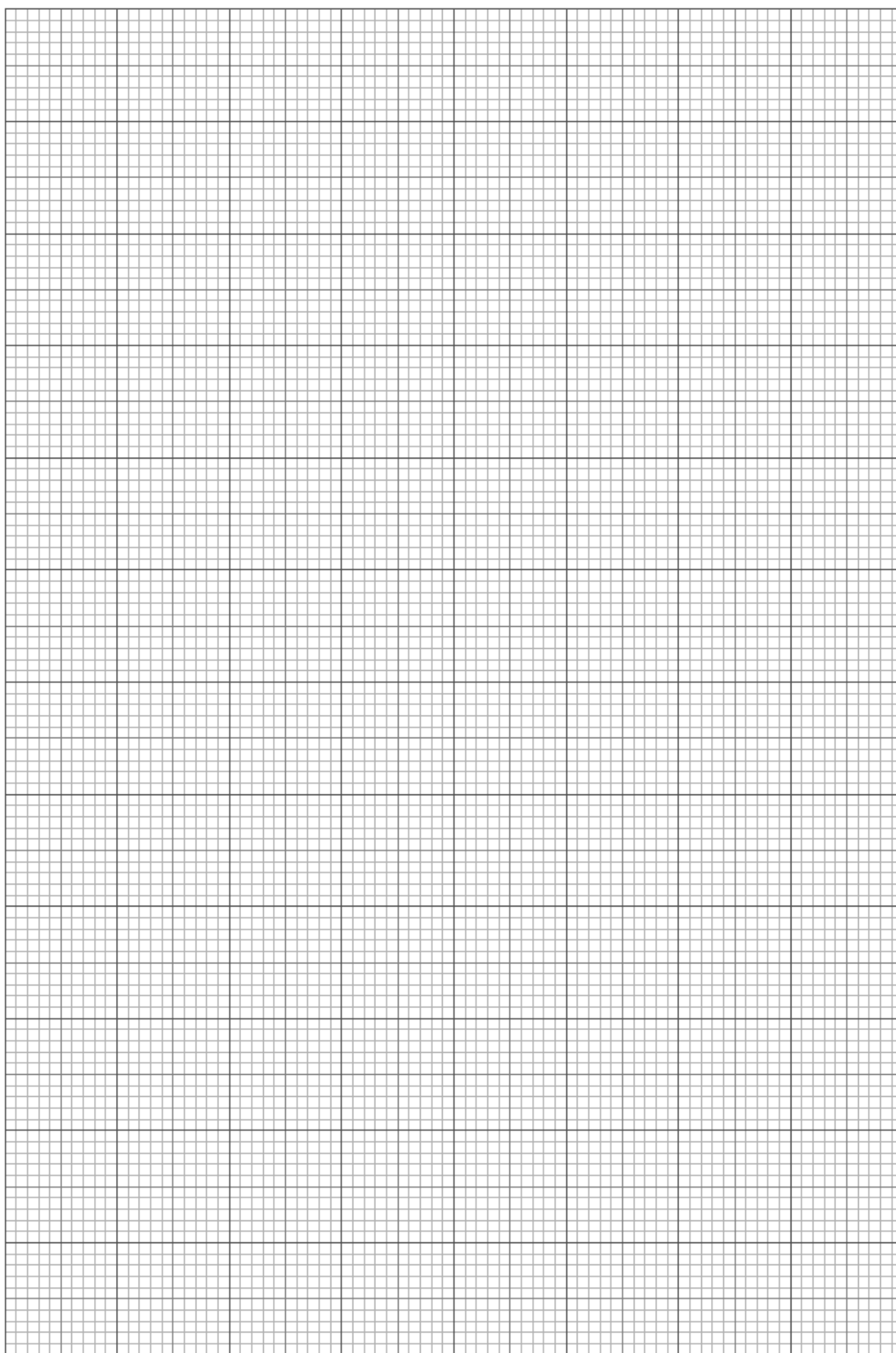
61*. [Maximum mark: 12]   *[with GDC]*

There are 50 boxes in a factory. Their weights, w kg, are divided into 5 classes, as shown in the following table.

Class	Weight (kg)	Number of boxes
A	$9.5 \leq w \leq 18.5$	7
B	$18.5 \leq w \leq 27.5$	12
C	$27.5 \leq w \leq 36.5$	13
D	$36.5 \leq w \leq 45.5$	10
E	$45.5 \leq w \leq 54.5$	8

(handwritten left of table:) 14 23 32 41 50

(a)   Show that the estimated mean weight of the boxes is 32 kg.   [3]

(b)   There are x boxes in the factory marked "Fragile". They are all in class E. The estimated mean weight of all the other boxes in the factory is 30 kg. Calculate the value of x.   [4]

(c)   An additional y boxes, all with a weight in class D, are delivered to the factory. The total estimated mean weight of **all** of the boxes in the factory is less than 33 kg. Find the largest possible value of y.   [5]

(handwritten:)

(a) $\dfrac{14(7)+23(12)+32(13)+41(10)+50(8)}{60} = \bar{x} = 32$

(b)