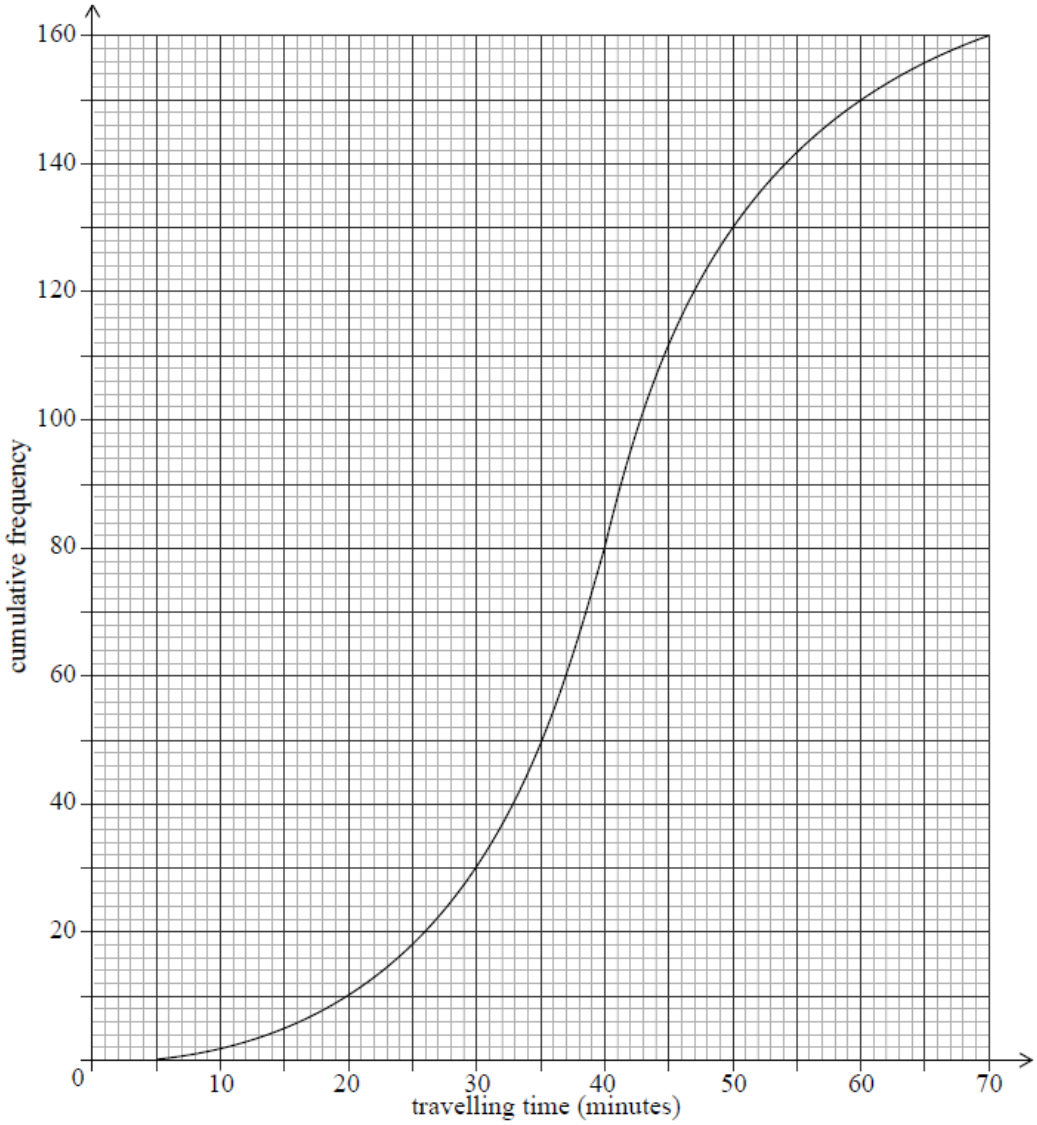
Statistics - Revision [62 marks]

1. [Maximum mark: 15] SPM.1.SL.TZ0.7

A large company surveyed 160 of its employees to find out how much time they spend traveling to work on a given day. The results of the survey are shown in the following cumulative frequency diagram.



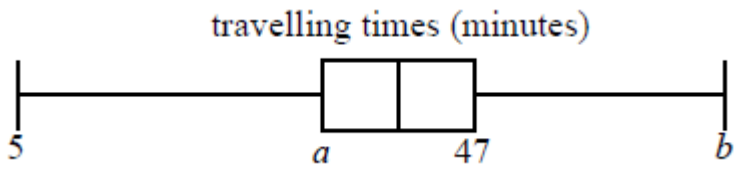
(a) Find the median number of minutes spent traveling to work. [2]

(b) Find the number of employees whose travelling time is within 15 minutes of the median. [3]

Only 10% of the employees spent more than k minutes traveling to work.

(c) Find the value of k. [3]

The results of the survey can also be displayed on the following box-and-whisker diagram.



(d) Write down the value of b. [1]

(e.i) Find the value of a. [2]

(e.ii) Hence, find the interquartile range. [2]

(f) Travelling times of less than p minutes are considered outliers.

Find the value of p. [2]

2. [Maximum mark: 8] EXN.2.SL.TZ0.4

The following table shows the systolic blood pressures, p mmHg, and the ages, t years, of 6 male patients at a medical clinic.

Patient	P1	P2	P3	P4	P5	P6
t (years)	40	72	35	47	21	61
p (mmHg)	105	145	100	130	95	132

(a.i) Determine the value of Pearson's product-moment correlation coefficient, r, for these data. [2]

(a.ii) Interpret, in context, the value of r found in part (a) (i). [1]

The relationship between t and p can be modelled by the regression line of p on t with equation $p = at + b$.

(b) Find the equation of the regression line of p on t. [2]

A 50-year-old male patient enters the medical clinic for his appointment.

(c) Use the regression equation from part (b) to predict this patient's systolic blood pressure. [2]

(d) A 16-year-old male patient enters the medical clinic for his appointment.

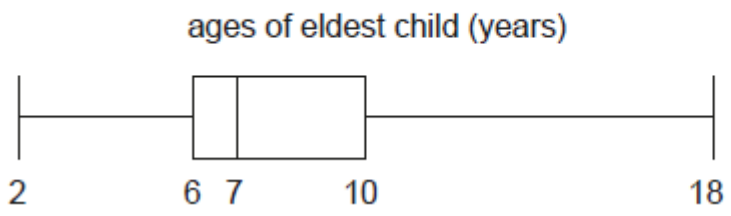
Explain why the regression equation from part (b) should not be used to predict this patient's systolic blood pressure. [1]

3. [Maximum mark: 7] 22M.1.SL.TZ1.3

A survey at a swimming pool is given to one adult in each family. The age of the adult, a years old, and of their eldest child, c years old, are recorded.

The ages of the eldest child are summarized in the following box and whisker diagram.

diagram not to scale



(a) Find the largest value of c that would not be considered an outlier. [3]

The regression line of a on c is $a = \frac{7}{4}c + 20$. The regression line of c on a is $c = \frac{1}{2}a - 9$.

(b.i) One of the adults surveyed is 42 years old. Estimate the age of their eldest child. [2]

(b.ii) Find the mean age of all the adults surveyed. [2]

4. [Maximum mark: 4] 22M.2.SL.TZ1.2

The number of hours spent exercising each week by a group of students is shown in the following table.

Exercising time (in hours)	Number of students
2	5
3	1
4	4
5	3
6	x

The median is 4.5 hours.

(a) Find the value of x. [2]

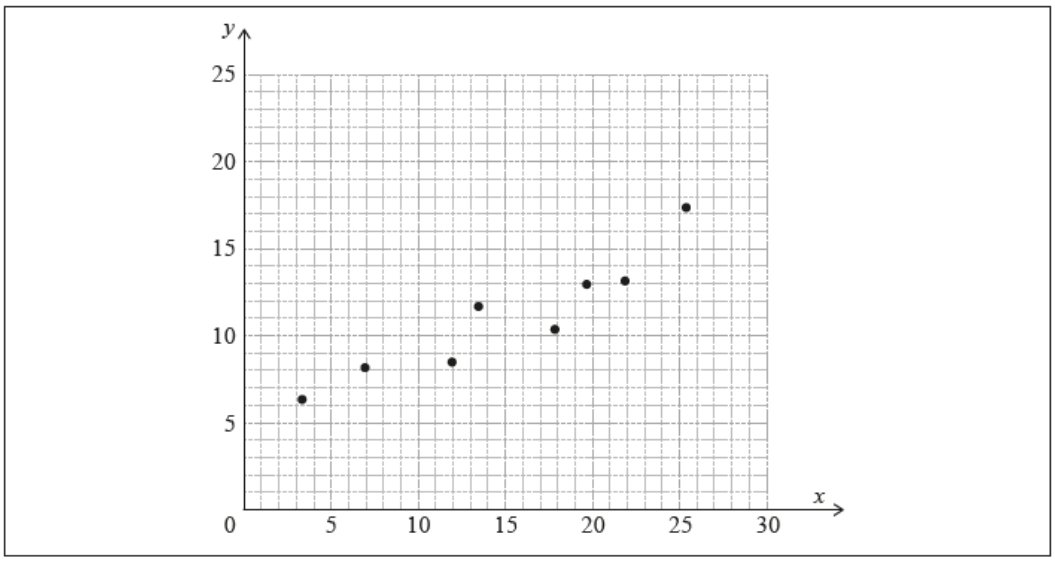
(b) Find the standard deviation. [2]

5. [Maximum mark: 7] 21M.2.SL.TZ1.2

The following table shows the data collected from an experiment.

x	3.3	6.9	11.9	13.4	17.8	19.6	21.8	25.3
y	6.3	8.1	8.4	11.6	10.3	12.9	13.1	17.3

The data is also represented on the following scatter diagram.



The relationship between x and y can be modelled by the regression line of y on x with equation $y = ax + b$, where $a,\ b \in \mathbb{R}$.

(a) Write down the value of a and the value of b. [2]

(b) Use this model to predict the value of y when $x = 18$. [2]

(c) Write down the value of $\overline{x}$ and the value of $\overline{y}$. [1]

(d) Draw the line of best fit on the scatter diagram. [2]

6. [Maximum mark: 6] 20N.1.SL.TZ0.T_3

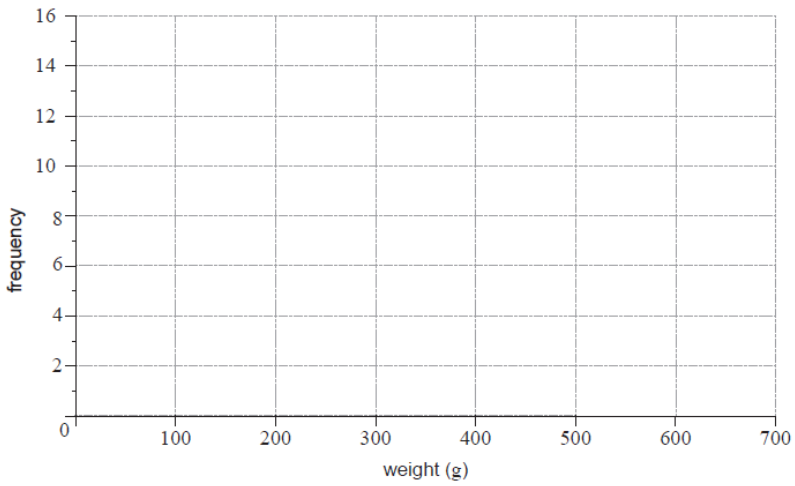
Hafizah harvested 49 mangoes from her farm. The weights of the mangoes, w, in grams, are shown in the following grouped frequency table.

Weight (g)	$100 \le w < 200$	$200 \le w < 300$	$300 \le w < 400$	$400 \le w < 500$	$500 \le w < 600$
Frequency	4	7	14	16	8

(a) Write down the modal group for these data. [1]

(b) Use your graphic display calculator to find an estimate of the standard deviation of the weights of mangoes from this harvest. [2]

(c) On the grid below, draw a histogram for the data in the table.



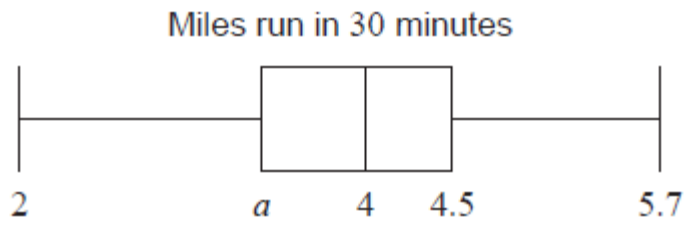
[3]

7. [Maximum mark: 15] 20N.1.SL.TZ0.S_8

Each athlete on a running team recorded the distance (M miles) they ran in 30 minutes.

The median distance is 4 miles and the interquartile range is 1.1 miles.

This information is shown in the following box-and-whisker plot.



(a) Find the value of a. [2]

The distance in miles, M, can be converted to the distance in kilometres, K, using the formula $K = \frac{8}{5}M$.

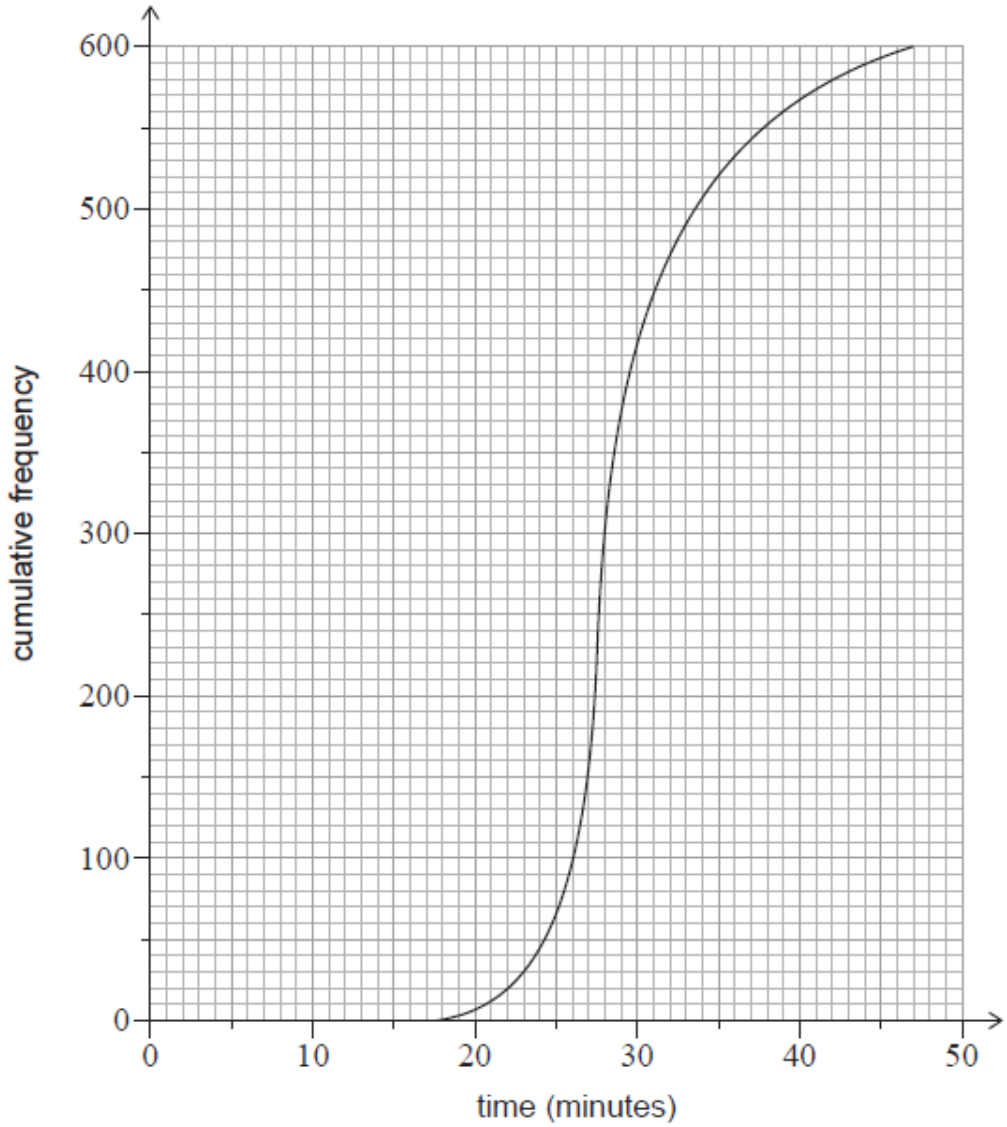
(b) Write down the value of the median distance in kilometres (km). [1]

The variance of the distances run by the athletes is $\frac{16}{9}\,\mathrm{km}^2$.

The standard deviation of the distances is b miles.

(c) Find the value of b. [4]

A total of 600 athletes from different teams compete in a $5\,\mathrm{km}$ race. The times the 600 athletes took to run the $5\,\mathrm{km}$ race are shown in the following cumulative frequency graph.



There were 400 athletes who took between 22 and m minutes to complete the $5\,\mathrm{km}$ race.

(d) Find m. [3]

(e) The first 150 athletes that completed the race won a prize.

Given that an athlete took between 22 and m minutes to complete the $5\,\mathrm{km}$ race, calculate the probability that they won a prize. [5]

© International Baccalaureate Organization, 2024