MATHEMATICS

Stage 5

Statistics and Probability
Data Analysis
Bivariate Data
Box Plots and 5-Number Summary
Standard Deviation

Teacher: Haeun Jang
Exam Equivalent Time: 174 minutes *(based on allocation ~1.25 min/mark)*

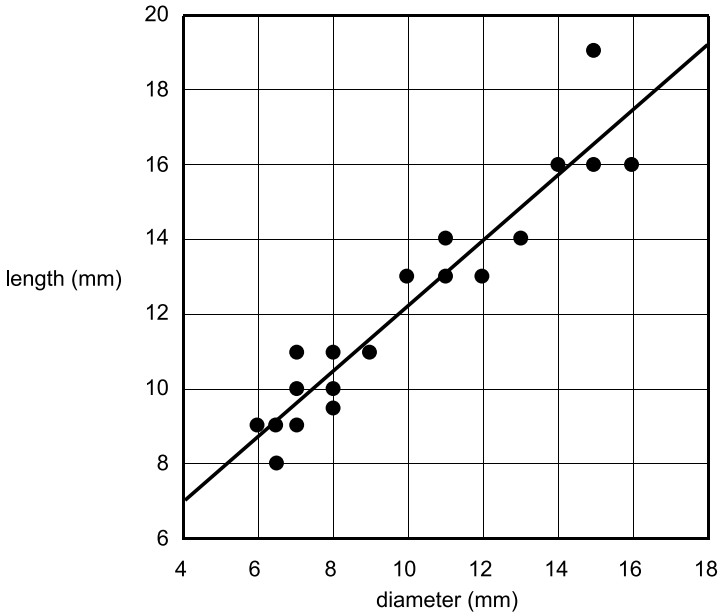
norWEST CHRISTIAN COLLEGE
STRENGTH OF PURPOSE

Questions

1. Bivariate Data (262876)

The lengths and diameters, in millimetres, of a sample of jellyfish selected were recorded and displayed in the scatterplot below.

A line of best fit for this data is shown.



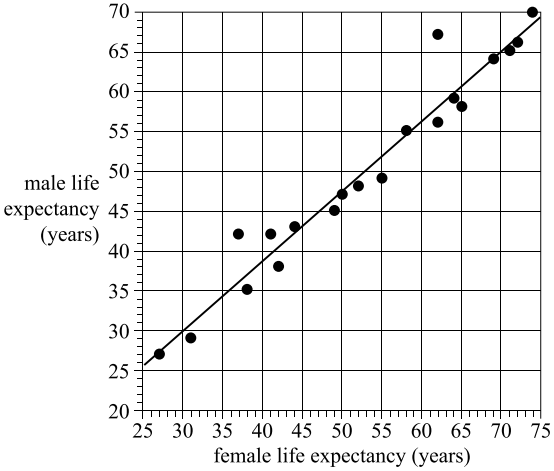
i. Determine the expected length of a jellyfish with a diameter of 12 millimetres. *(1 mark)*

ii. Determine the expected diameter of a jellyfish with a length of 16 millimetres. *(1 mark)*

2. Bivariate Data (263112)

The scatterplot below plots male life expectancy (***male***) against female life expectancy (***female***) in 1950 for a number of countries.

A line of best fit has been fitted to the scatterplot as shown.



i. State the dependent variable in the graph. *(1 mark)*

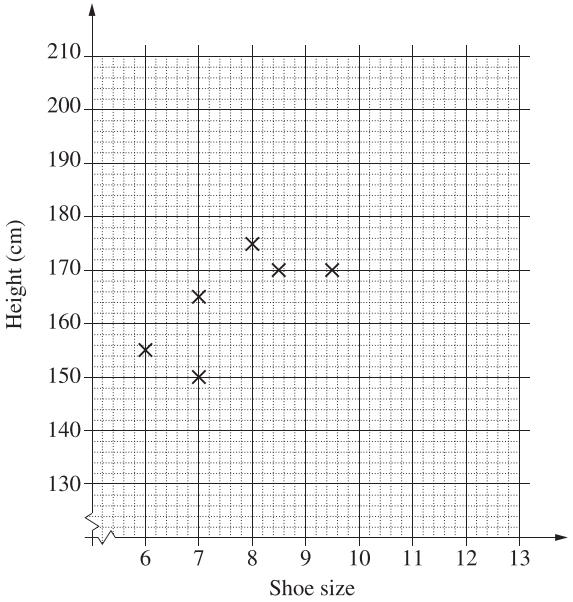
ii. Determine the age at which males and females have the same life expectancy. *(1 mark)*

3. Bivariate Data (262752)

The shoe size and height of ten students were recorded.

Shoe size	6	7	7	8	8.5	9.5	10	11	12	12
Height	155	150	165	175	170	170	190	185	200	195

i. Complete the scatter plot AND draw a line of best fit by eye. ***(2 marks)***



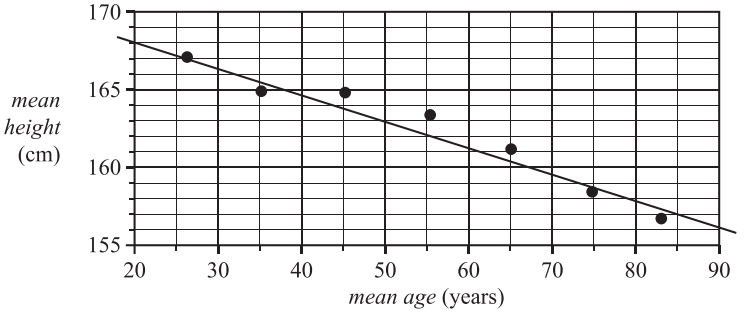
ii. Use the line of best fit to estimate the height difference between a student who wears a size 7.5 shoe and one who wears a size 9 shoe. ***(1 mark)***

---

4. Bivariate Data (263114)

The scatterplot below displays the ***mean age***, in years, and the ***mean height***, in centimetres, of 648 women from seven different age groups.

In an initial analysis of the data, a line of best fit is drawn, as shown.



i. Describe this association in terms of strength and direction. ***(2 marks)***

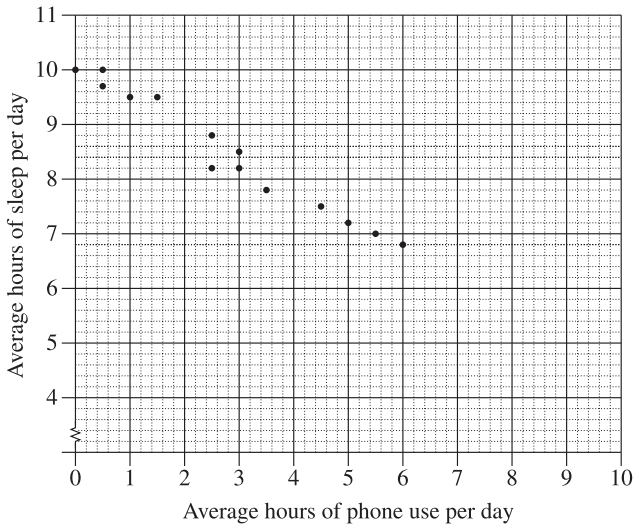
ii. Determine the mean height predicted for a group of 65 year old women. ***(1 mark)***

---

5. Bivariate Data (202645)

A teacher surveyed the students in her Year 8 class to investigate the relationship between the average number of hours of phone use per day and the average number of hours of sleep per day.

The results are shown on the scatterplot below.



a. The data for two new students, Alinta and Birrani, are shown in the table below. Plot their results on the scatterplot. ***(2 marks)***

	Average hours of phone use per day	*Average hours of sleep per day*
Alinta	4	8
Birrani	0	10.5

b. By first fitting the line of best fit by eye on the scatterplot, estimate the average number of hours of sleep per day for a student who uses the phone for an average of 2 hours per day. ***(2 marks)***

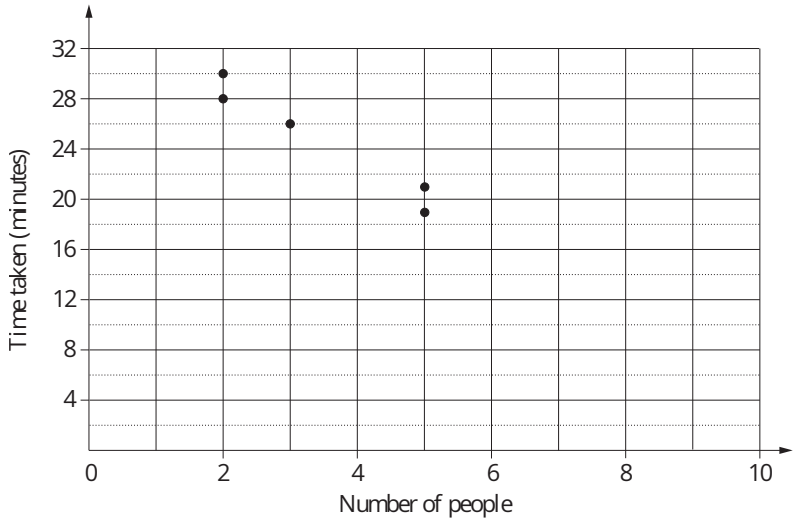
6. Bivariate Data (262725)

People are placed into groups to complete a puzzle. There are 9 different groups.

The table shows the number of people in each group and the amount of time, in minutes, for each group to complete the puzzle.

Number of people	2	2	3	5	5	7	7	7	8
Time taken (min)	28	30	26	19	21	12	13	11	8

a. Complete the scatterplot by adding the last four points from the table. ***(1 mark)***



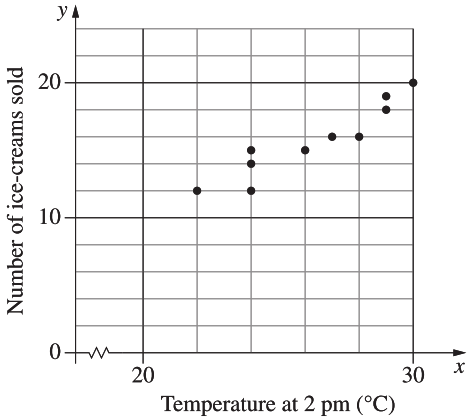
b. Add a line of best fit by eye to the graph in part (a). ***(1 mark)***

c. The graph in part (a) shows the association between the time to complete the puzzle and the number of people in the group.

Identify the form (linear or non-linear), the direction and the strength of the association. ***(2 marks)***

7. Bivariate Data (234689)

The scatterplot shows the number of ice-creams sold, y, at a shop over a ten-day period, and the temperature recorded at 2 pm on each of these days.



a. The data are modelled by the equation of the line of best fit given below.

$y = 0.936x - 8.929$, where x is the temperature.

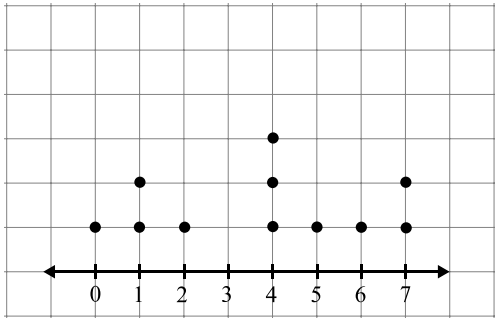
Sam used a particular temperature with this equation and predicted that 23 ice-creams would be sold.

What was the temperature used by Sam, to the nearest degree? ***(2 marks)***

b. In using the equation to make the prediction in part (a), was Sam interpolating or extrapolating? Justify your answer. ***(2 marks)***

---

8. Box Plots and 5-Number Summary (261334)



Using a 5-number summary of the data represented in this graph, or otherwise, determine

i. the range ***(1 mark)***

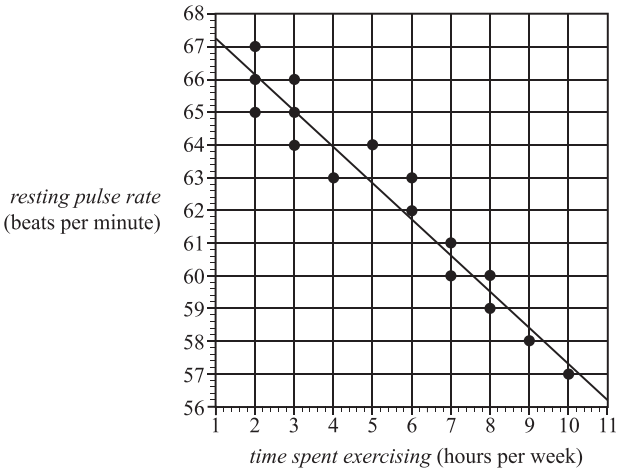
ii. the interquartile range ***(2 marks)***

---

9. Bivariate Data (263026)

The scatterplot below displays the resting pulse rate, in beats per minute, and the time spent exercising, in hours per week, of 16 students.

A line of best fit has been fitted to the data.



i. If a student spends 8 hours exercising per week, determine the resting pulse rate predicted by the line of best fit. ***(1 mark)***

ii. Provide TWO descriptions of the association between the variables ***time spent exercising*** and ***resting pulse rate***. ***(2 marks)***

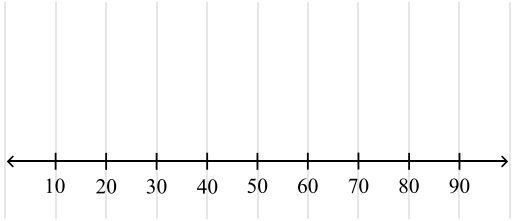
---

10. Box Plots and 5-Number Summary (261449)

Eli sat 22 Maths quizzes throughout year 8 and her results were recorded in the 5-number summary below.

Min	Q_1	Q_2	Q_3	Max
30	50	60	70	80

i. Using the number line below as reference, draw a box plot that displays Eli's results. ***(2 marks)***



ii. Describe the skew of Eli's results. ***(1 mark)***

---

11. Box Plots and 5-Number Summary (261221)

Le Bron plays basketball and records the number of free throws he shoots in the first nine games of the season.

$$12,\ 8,\ 4,\ 15,\ 13,\ 7,\ 8,\ 3,\ 10$$

Using a 5-number summary of this dataset, or otherwise, determine

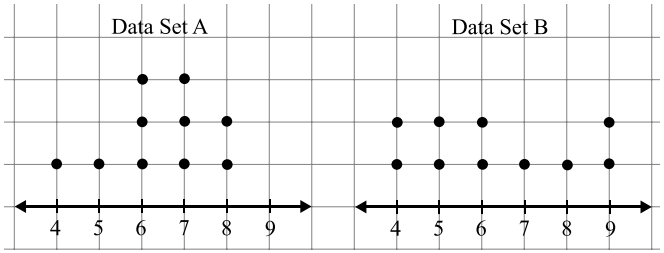
i. the median *(1 mark)*

ii. the interquartile range *(2 marks)*

---

12. Standard Deviation (262424)

Two data sets are represented in the dot plot diagrams below.



i. Calculate the standard deviation of Data Set B, giving your answer correct to one decimal place. *(1 mark)*

ii. Without using calculations, explain which data set has the highest standard deviation. *(2 marks)*

---

13. Box Plots and 5-Number Summary (261238)

The data set below consists of 8 data points.

$$7.0,\ 7.4,\ 8.1,\ 9.3,\ 9.5,\ 10.7,\ 11.1,\ 12.9$$

Complete the 5-number summary table of the data set below. *(2 marks)*

Min	Q_1	Q_2	Q_3	Max

---

14. Box Plots and 5-Number Summary (261453)

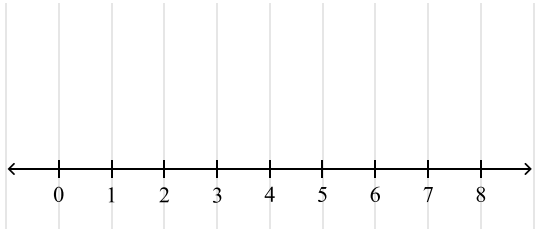
Ellyse recorded the number of wickets she took in the first seven games of the cricket season, which are as follows:

$$3,\ 5,\ 0,\ 8,\ 1,\ 2,\ 2$$

i. Complete the 5-number summary table of Ellyse's results below. *(2 marks)*

Min	Q_1	Q_2	Q_3	Max

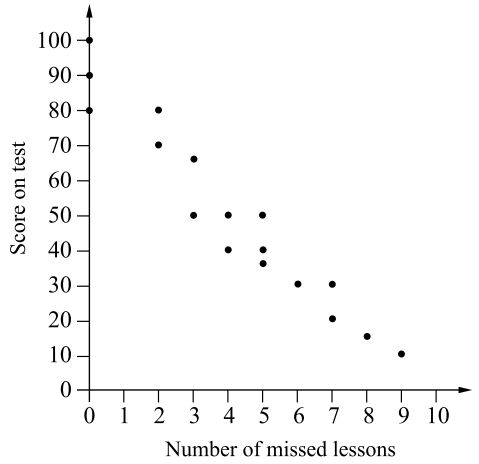
ii. Draw a box plot, using the number line below for reference, that displays Ellyse's results. *(2 marks)*



iii. Describe the skew of Ellyse's results. *(1 mark)*

15. Bivariate Data (159055)

A group of students sat a test at the end of term. The number of lessons each student missed during the term and their score on the test are shown on the scatterplot.



a. Describe the strength and direction of the linear association observed in this dataset. *(2 marks)*

b. Calculate the range of the test scores for the students who missed no lessons. *(1 mark)*

c. Draw a line of the best fit in the scatterplot above. *(1 mark)*

d. Meg did not sit the test. She missed five lessons.
Use the line of the best fit drawn in part (c) to estimate Meg's score on this test. *(1 mark)*

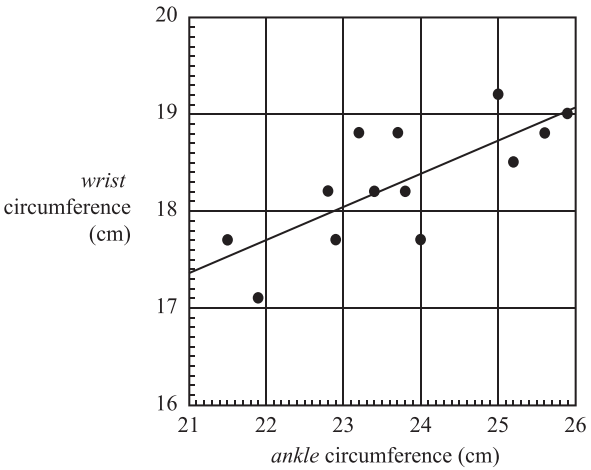
e. John also did not sit the test and he missed 16 lessons.
Is it appropriate to use the line of the best fit to estimate his score on the test? Briefly explain your answer. *(1 mark)*

---

16. Bivariate Data (263015)

The scatterplot below shows the ***wrist*** circumference and ***ankle*** circumference, both in centimetres, of 13 people.

A line of best fit been drawn with ankle circumference as the independent variable.



i. If a person has a wrist circumference of 18.5 centimetres, estimate the ankle circumference that is predicted by the line of best fit. *(1 mark)*

ii. Explain why the y-intercept of this graph has no meaning in this context. *(1 mark)*

---

17. Standard Deviation (262373)

Ali's class sits a Geography test and the results are recorded below.

$$58,\ 74,\ 65,\ 66,\ 73,\ 71,\ 72,\ 74,\ 62,\ 70$$

The mean for the test was 68.5.

i. Calculate the standard deviation for the test. Give your answer correct to one decimal place. *(2 marks)*

ii. Explain whether the mean of the test is greater than or less than the median of the test. *(1 mark)*

---

18. Box Plots and 5-Number Summary (261239)

Matt records the number of times he takes his dog Max for a walk each month. The data set below records the months January to August.

$$21,\ 24,\ 28,\ 22,\ 15,\ 12,\ 13,\ 17$$

Using a 5-number summary of this dataset, or otherwise, determine

i. the median *(1 mark)*

ii. the interquartile range *(2 marks)*

---

19. Standard Deviation (262545)

Albert teaches Physics and sets his class a mid-term exam.

The results are summarised in the Stem and Leaf plot drawn below.

Stem	Leaf
6	2 5
7	8 9
8	3 9
9	4 6 7

6 | 2 means 62

i. Determine the median test score of the data set. ***(1 mark)***

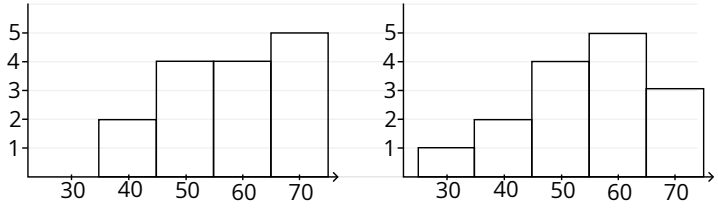
ii. Calculate the standard deviation of this data set, giving your answer to two decimal places. ***(1 mark)***

---

20. Standard Deviation (262487)

Mr Ralph has a History class that has completed two separate exams during the term.

The results of each exam are summarised in the histograms below.



Without using calculations, explain which test will have the highest standard deviation. ***(2 marks)***

---

21. Standard Deviation (096660)

Jamal surveyed eight households in his street. He asked them how many kilolitres (kL) of water they used in the last year. Here are the results.

$$220, 105, 101, 450, 37, 338, 151, 205$$

a. Calculate the mean of this set of data. ***(1 mark)***

b. What is the standard deviation of this set of data, correct to one decimal place? ***(1 mark)***

---

22. Standard Deviation (262439)

Seven players in two basketball teams, the Swifties and the Chiefs, recorded how many 3-point baskets they had shot in the last season.

The results are recorded in the two Stem and Leaf plots below.

Swifties

Stem	Leaf
0	1 2
1	7
2	2 6
3	3 9

Chiefs

Stem	Leaf
0	8
1	0 3 5
2	1 4 6
3	

2 | 2 means 22

i. Determine the standard deviation of the Swifties' results, giving your answer correct to one decimal place. ***(2 marks)***

ii. Without using calculations, explain which data set will have the smallest standard deviation. ***(2 marks)***

---

23. Standard Deviation (159260)

a. The ages in years, of ten people at the local cinema last Saturday afternoon are shown.

$$38 \;\; 25 \;\; 38 \;\; 46 \;\; 55 \;\; 68 \;\; 72 \;\; 55 \;\; 36 \;\; 38$$

The mean of this dataset is 47.1 years.

How many of the ten people were aged between the mean age and the median age? ***(2 marks)***

b. On Wednesday, ten people all aged 70 went to this same cinema.

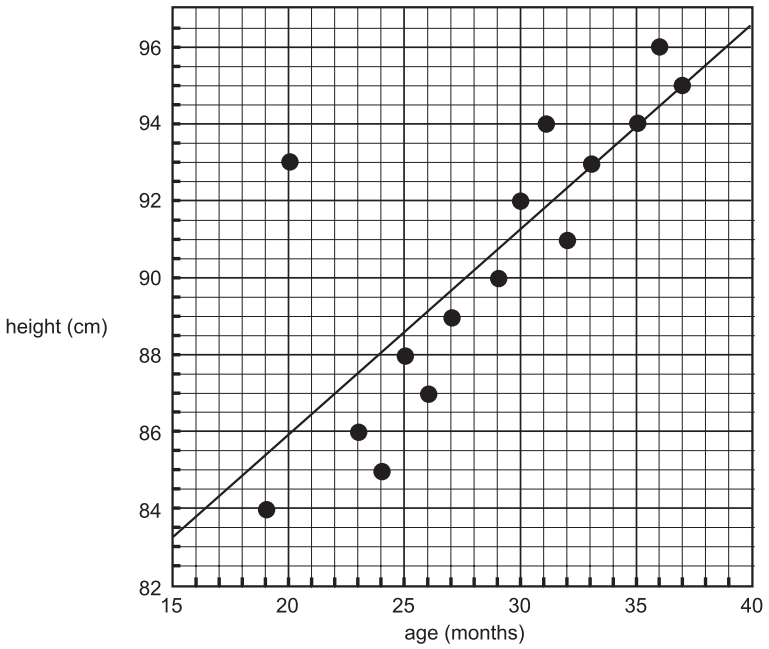
Would the standard deviation of the age dataset from Wednesday be larger than, smaller than or equal to the standard deviation of the age dataset given in part (a)? Briefly explain your answer without performing any calculations. ***(2 marks)***

---

24. Bivariate Data (263150)

The heights (in cm) and ages (in months) of a random sample of 15 boys have been plotted in the scatterplot below.

A line of best fit has been fitted to the data.



i. State the independent variable in the graph. ***(1 mark)***

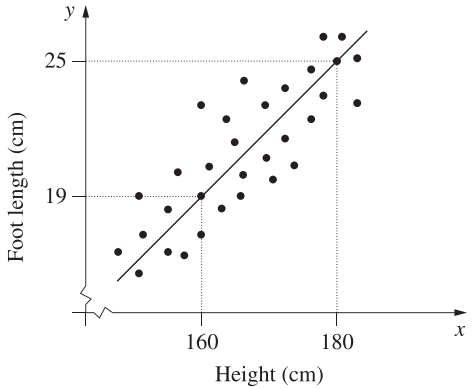
ii. Describe this association in terms of strength and direction. ***(2 marks)***

iii. Determine the gradient of the line of best fit, giving your answer correct to one decimal place. ***(2 marks)***

---

25. Bivariate Data (262839)

Each member of a group of males had his height and foot length measured and recorded. The results were graphed and a line of fit drawn.



NOT TO SCALE

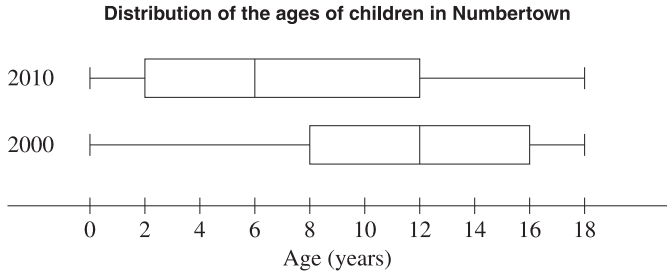
i. Why does the value of the y-intercept have no meaning in this situation? ***(1 mark)***

ii. George is 10 cm taller than his brother Harry. Use the line of fit to estimate the difference in their foot lengths. ***(1 mark)***

---

26. Box Plots and 5-Number Summary (005376)

The graphs show the distribution of the ages of children in Numbertown in 2000 and 2010.



a. In 2000 there were 1750 children aged 0–18 years.
How many children were aged 12–18 years in 2000? ***(1 mark)***

b. The number of children aged 12–18 years is the same in both 2000 and 2010.
How many children aged 0–18 years are there in 2010? ***(1 mark)***

c. Identify TWO changes in the distribution of ages between 2000 and 2010. In your answer, refer to measures of location or spread or the shape of the distributions. ***(2 marks)***

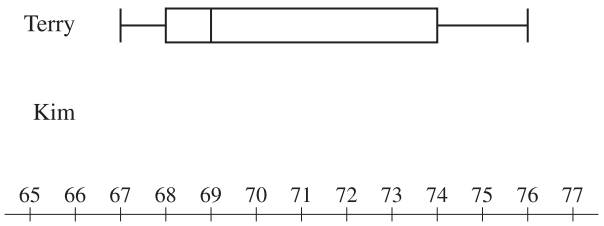
d. What would be ONE possible implication for government planning, as a consequence of this change in the distribution of ages? ***(1 mark)***

---

27. Box Plots and 5-Number Summary (261395)

Terry and Kim each sat twenty class tests. Terry's results on the tests are displayed in the box plot shown in part (i).

i. Kim's 5-number summary for the tests is 67, 69, 71, 73, 75.
Draw a box plot to display Kim's results below that of Terry's results. ***(1 mark)***

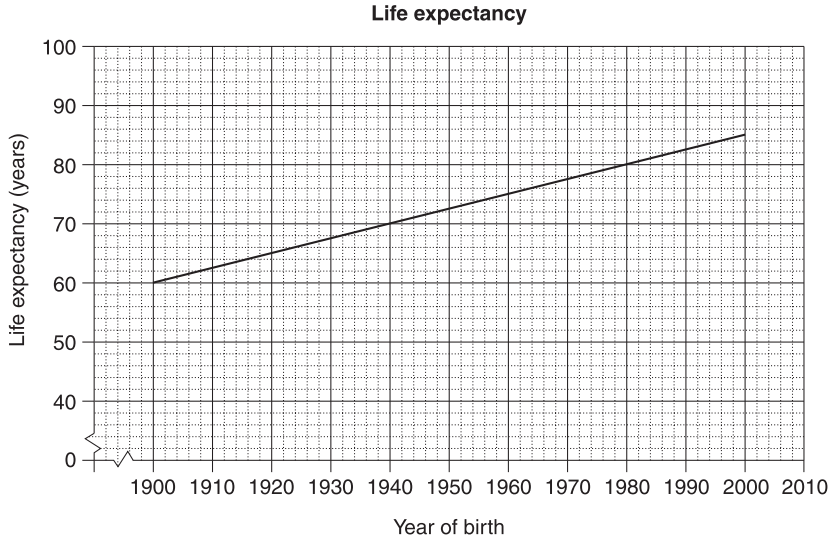


ii. What percentage of Terry's results were below 69? ***(1 mark)***

iii. Describe the skew of Kim's results. ***(1 mark)***

28. Bivariate Data (262697)

The graph shows a line of best fit describing the life expectancy of people born between 1900 and 2000.



i. According to the graph, what is the life expectancy of a person born in 1932? ***(1 mark)***

ii. Determine the value of the gradient of the line of best fit. ***(2 marks)***

29. Standard Deviation (262338)

The ages of nine students were recorded in the table below.

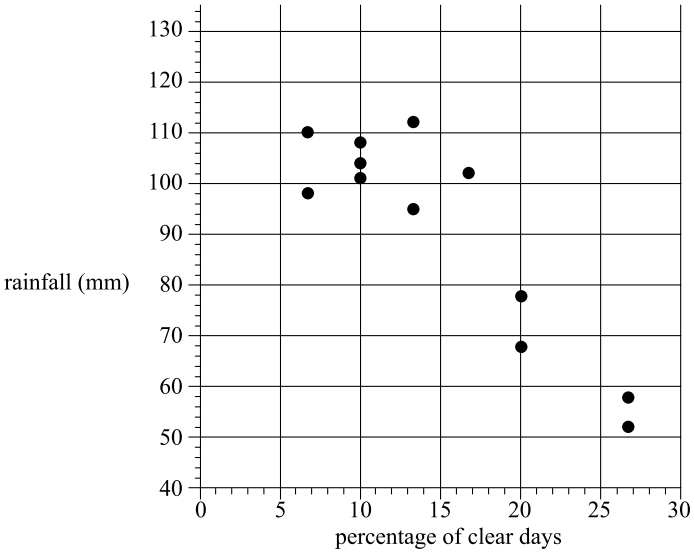
Ages		
12	11	16
14	16	15
14	15	14

i. What is the standard deviation, correct to two decimal places? ***(2 marks)***

ii. Briefly explain what is meant by the term ***standard deviation***. ***(1 mark)***

30. Bivariate Data (263158)

The scatterplot below shows the ***rainfall*** (in mm) and the ***percentage of clear days*** for each month of 2023.



An equation of the line of best fit for this data set is

$$rainfall = 131 - 2.68 \times percentage\ of\ clear\ days$$

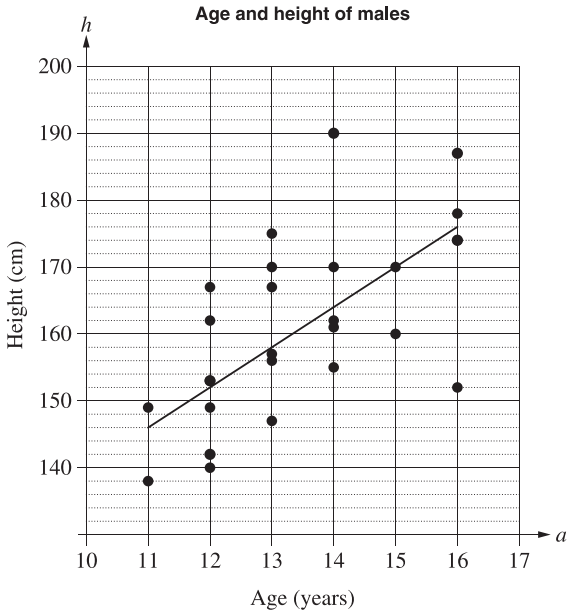
i. Using coordinates at the graph extremities or otherwise, draw this line on the scatterplot. ***(2 marks)***

ii. Describe this association in terms of strength and direction. ***(1 mark)***

31. Bivariate Data (262747)

Ahmed collected data on the age (a) and height (h) of males aged 11 to 16 years.

He created a scatterplot of the data and constructed a line of best fit to model the relationship between the age and height of males.



i. Determine the gradient of the line of best fit shown on the graph. *(1 mark)*

ii. Determine the equation of the line of best fit shown on the graph. *(2 marks)*

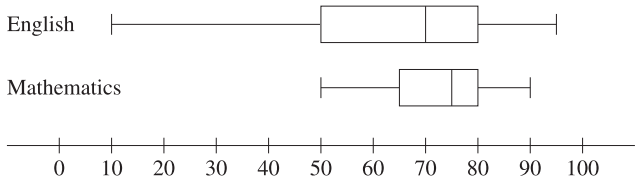
iii. Use the line of best fit to predict the height of a typical 17-year-old male. *(1 mark)*

iv. Why would this model not be useful for predicting the height of a typical 45-year-old male? *(1 mark)*

---

32. Box Plots and 5-Number Summary (261404)

The test results in English and Mathematics for a class were recorded and displayed in the box plots.



i. What is the interquartile range for English? *(1 mark)*

ii. Compare and contrast the two data sets by referring to the skewness of the distributions and the measures of their centres and spread. *(3 marks)*

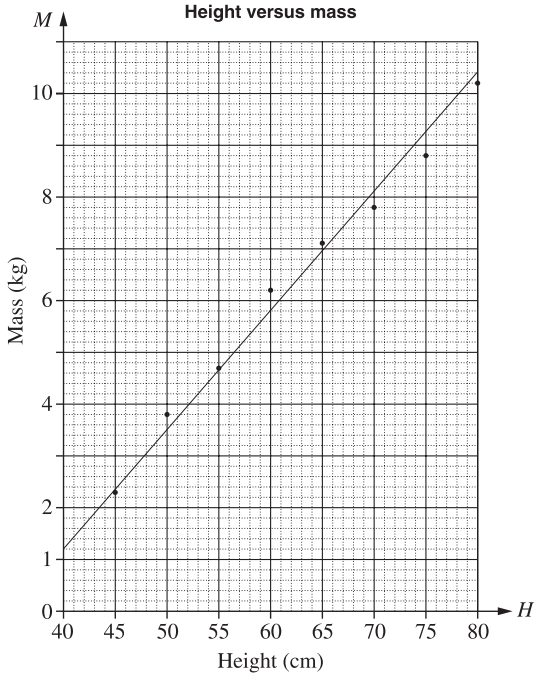
---

33. Bivariate Data (004584)

The height and mass of a child are measured and recorded over its first two years.

Height (cm), H	45	50	55	60	65	70	75	80
Mass (kg), M	2.3	3.8	4.7	6.2	7.1	7.8	8.8	10.2

This information is displayed in a scatter graph.



a. Describe the correlation between the height and mass of this child, as shown in the graph. *(1 mark)*

b. A line of best fit has been drawn on the graph.
Find the equation of this line. *(2 marks)*

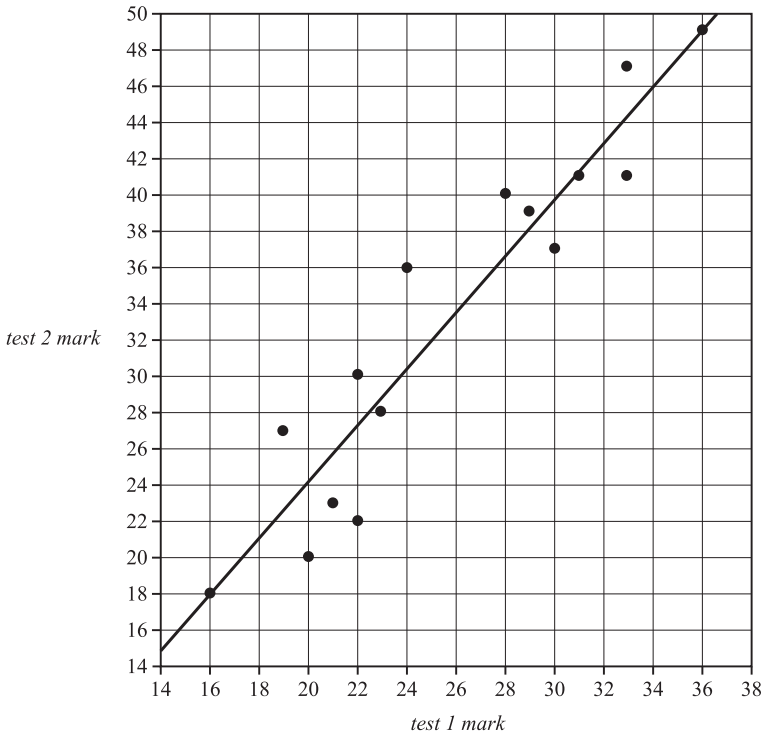
---

34. Bivariate Data (263106)

A teacher analysed the class marks of 15 students who sat two tests.

The ***test 1 mark*** and ***test 2 mark***, all whole number values, are shown in the scatterplot below.

A line of best fit has been fitted to the scatterplot.



i. If a student scored 34 in the first test, what is their expected mark in the second test. ***(1 mark)***

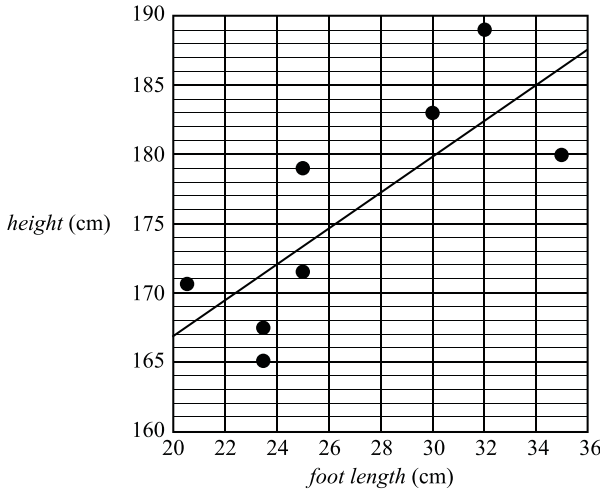
ii. The line of best fit shows the predicted ***test 2 mark*** for each student based on their ***test 1 mark***.

Determine the number of students whose actual ***test 2 mark*** was within two marks of that predicted. ***(1 mark)***

35. Bivariate Data (262982)

The ***height*** (in cm) and ***foot length*** (in cm) for each of eight Year 12 students were recorded and displayed in the scatterplot below.

A line of best fit has been fitted to the data as shown.

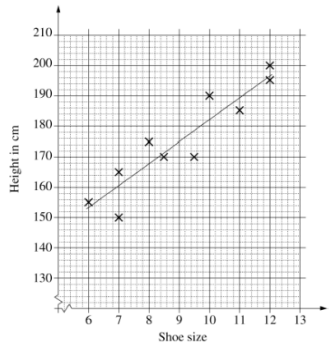


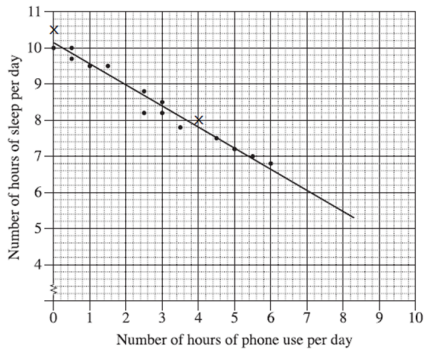
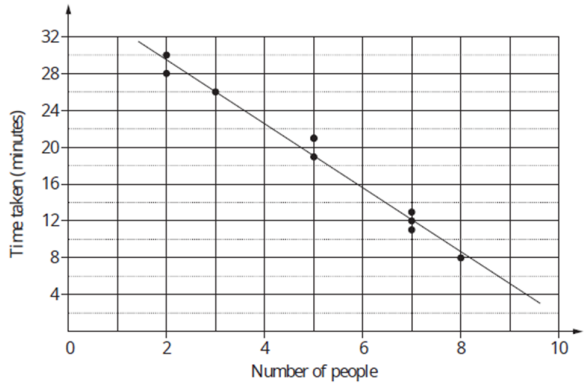
i. Determine the predicted foot size of a student who is 176 centimetres tall. ***(1 mark)***

ii. Calculate the gradient of the line of best fit, giving your answer correct to two decimal places. ***(2 marks)***

iii. What is the equation of the line of best fit? ***(2 marks)***

Single and Bivariate Data QB_Answers

Q1.	i. $14mm$ ii. $14.4mm\ (approx)$
Q2.	i. Male life expectancy ii. 30 years
Q3.	i.  ii. $13cm\ (or\ close\ given\ the\ line\ of\ best\ fit\ drawn)$
Q4.	i. strong negative ii. 160.5cm

Q5.	a)  b) 9 hours of sleep per day
Q6.	a), b)  c) Linear, strong negative
Q7.	a) $34°\ (nearest\ degree)$ b) extrapolating as $34℃$ is outside the range of data points shown on the graph.
Q8.	i. Range = 7 ii. IQR = 5

Single and Bivariate Data QB_Answers

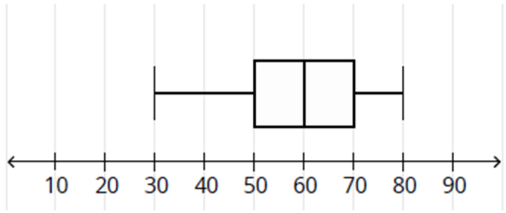
Q9.

i. 59.5

ii. Strong negative correlation

Q10.

i.



ii. Negatively skewed

Q11.

i. $Median = 8$

ii.

Min	Q_1	Q_2	Q_3	Max
3	5.5	8	12.5	15

IQR = 7

Q12.

i. $Standard\ Deviation \approx 1.8$

ii. Data Set A has a smaller range and a tighter spread against its expected mean value. Therefore Data Set B will have the highest standard deviation.

Q13.

Min	Q_1	Q_2	Q_3	Max
7.0	7.75	9.4	10.9	12.9

Q14

i. Ordering the data: 0, 1, 2, 2, 3, 5, 8

Min	Q_1	Q_2	Q_3	Max
0	1	2	5	8

ii.

iii. Positively skewed

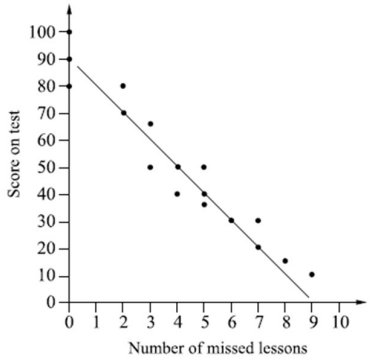
Single and Bivariate Data QB_Answers

Q15.

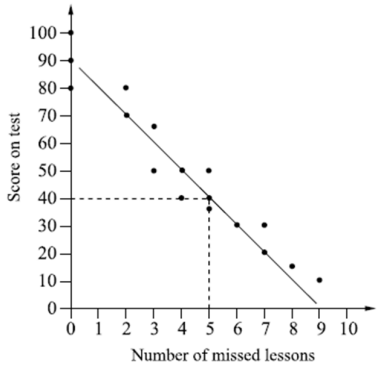
a) Strong negative

b) Range = 20

c.



d.



$\therefore$ Meg's estimated score $= 40$

e) No. John's missed days are too extreme and the model would estimate a negative score for John which is impossible.

Q16.

i. 24.5cm

ii. y-intercept occurs when ankle circumference is 0cm, which is meaningless in this context.

Q17.

i. $Standard\ Deviation \approx 5.2$

ii. By calculating both median and mean,
$Median > Mean\ (70.5 > 68.5)$

Q18.

i. Median = 19

ii. IQR = 9

Q19.

i. Median = 83

ii. $Standard\ Deviation \approx 12.14$

Q20.

Test B will have a higher standard deviation as test A results look to have a tighter spread of data points about its expected mean.

Q21.

a) Mean = 200.875

b) $Standard\ Deviation \approx 127.4$

Q22.

i. $Standard\ Deviation \approx 13.4$

ii. The Chiefs' results will have a smaller standard deviation as the Chiefs' results are a much tighter fit around the expected mean.

Single and Bivariate Data QB_Answers

Q23.

a) 1 person

b) Standard deviation is a measure of spread, so how much the ages of individuals differ from the mean age of the group. Therefore, standard deviation of Wednesday's group would be less as the mean is 70 and everyone's ages is 70.

Q24.

i. Age (months)

ii. Strong positive

iii. $m \approx 0.5\ (1.d.p)$

Q25.

i. y-intercept occurs when $x = 0cm$

ii. $Difference\ in\ their\ feet\ lengths = 3cm$

Q26.

a) 875 children

b) 3500 children

c) Changes in distribution (include 2 of the following):

- the lower quartile age is lower in 2010
- the median is lower in 2010
- the upper quartile age is lower in 2010
- the interquartile range is greater in 2010
- 2010 is positively skewed while 2000 is negatively

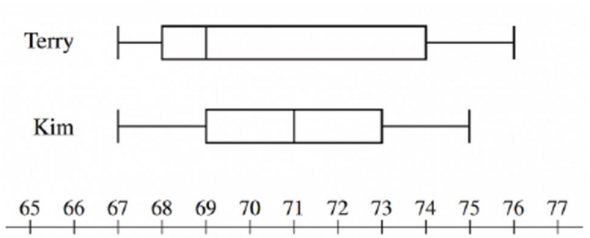
d) Implication for government planning:

Since the children are getting younger in 2010,

- Approve and build more childcare facilities
- Build more school and public playgrounds

Q27.

i.



ii. 50%

iii. Evenly skewed

Q28.

i. 68 years

ii. $m = 0.5$

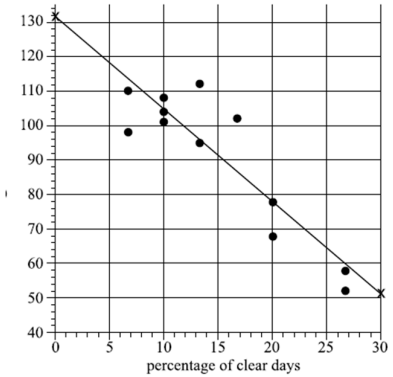
Q29.

i. $Standard\ Deviation \approx 1.59$

ii. Standard deviation is a measure of how much members of a data group differ from the mean value of the group.

Q30.

i.



ii. Moderate negative

Single and Bivariate Data QB_Answers

Q31.	i. Gradient = 6 ii. $h = 6a + 80$ iii. 182 cm iv. People slow and eventually stop growing after they become adults.
Q32.	i. IQR = 30 ii. Skewness • English has greater negative skew • Maths is more normally distributed Location and Spread • English has a range of 85, Maths has 40. • English has larger IQR than Maths (30 vs 15) • Maths' median (75) is higher than English (70) • Same upper quartile marks (80) • English has highest and lowest individual mark
Q33.	a) Strong positive correlation b) $M = 0.23H - 8$
Q34.	i. 2nd test mark = 46 ii. 5 values
Q35.	i. Foot length = 27cm ii. $Gradient \approx 1.29$ iii. $Height = 141.2 + 1.29 \times foot\ length$