Instructions

Answer **all** questions in pencil on the answer sheet provided for multiple-choice questions.
Choose the response that is **correct** for the question.
A correct answer scores 1; an incorrect answer scores 0.
Marks will **not** be deducted for incorrect answers.
No marks will be given if more than one answer is completed for any question.
Unless otherwise indicated, the diagrams in this book are **not** drawn to scale.

Data analysis

Use the following information to answer Questions 1 and 2.

The *blood pressure* (low, normal, high) and the *age* (under 50 years, 50 years or over) of 110 adults were recorded. The results are displayed in the two-way frequency table below.

Blood pressure	**Age**	
	Under 50 years	**50 years or over**
low	15	5
normal	32	24
high	11	23
Total	58	52

Question 1

The variables *blood pressure* (low, normal, high) and *age* (under 50 years, 50 years or over) are

A. both categorical variables.
B. both numerical variables.
C. a numerical variable and a categorical variable respectively.
D. a categorical variable and a numerical variable respectively.
E. a discrete variable and a continuous variable respectively.

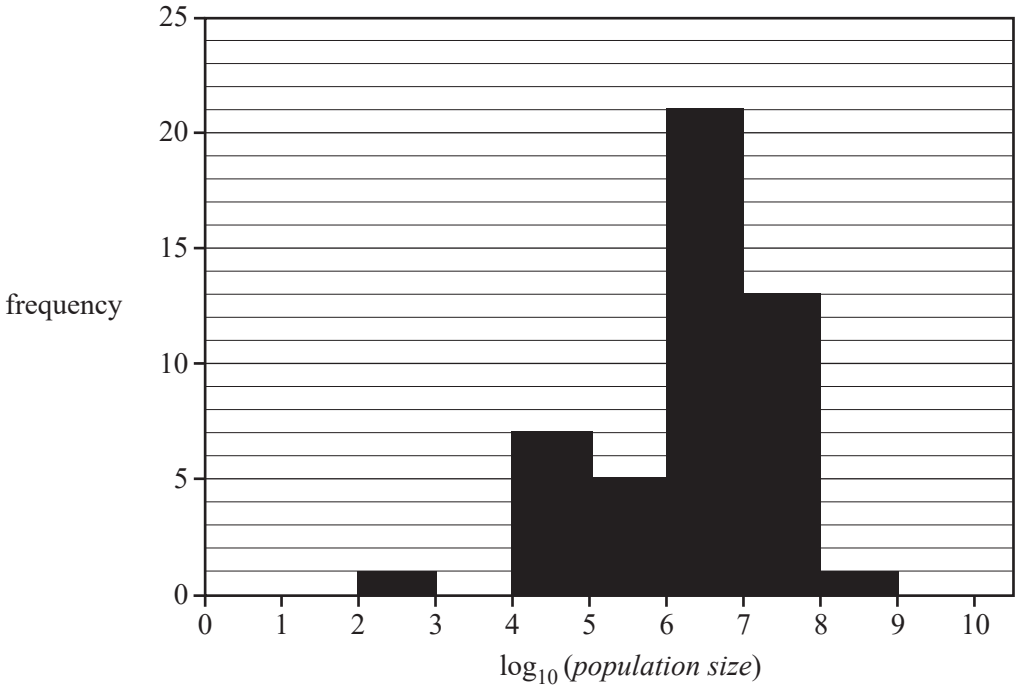
Question 2

The percentage of adults under 50 years of age who have high blood pressure is closest to

A. 11%
B. 19%
C. 26%
D. 44%
E. 58%

Question 3

The histogram below shows the *population size* for 48 countries plotted on a logarithmic (base 10) scale.



Data: Worldometers, <www.worldometers.info/>

Based on this histogram, the number of countries with a *population size* that is less than 100 000 people is

A. 1

B. 7

C. 8

D. 13

E. 48

Use the following information to answer Questions 4–6.

The pulse rates of a population of Year 12 students are approximately normally distributed with a mean of 69 beats per minute and a standard deviation of 4 beats per minute.

Question 4

A student selected at random from this population has a standardised pulse rate of $z = -2.5$

This student's actual pulse rate is

A. 59 beats per minute.

B. 63 beats per minute.

C. 65 beats per minute.

D. 73 beats per minute.

E. 79 beats per minute.

Question 5

Another student selected at random from this population has a standardised pulse rate of $z = -1$.

The percentage of students in this population with a pulse rate greater than this student is closest to

A. 2.5%

B. 5%

C. 16%

D. 68%

E. 84%

Question 6

A sample of 200 students was selected at random from this population.

The number of these students with a pulse rate of less than 61 beats per minute or greater than 73 beats per minute is closest to

A. 19

B. 37

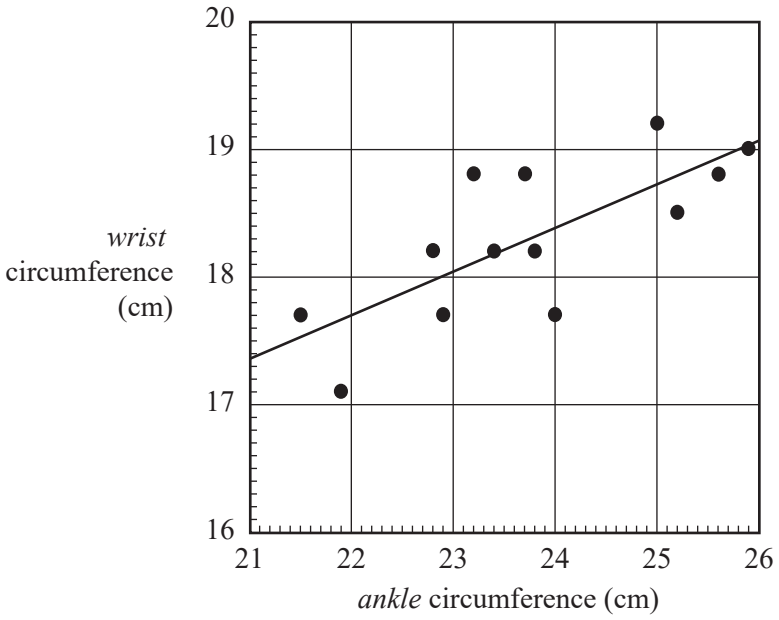
C. 64

D. 95

E. 190

Use the following information to answer Questions 7–9.

The scatterplot below shows the *wrist* circumference and *ankle* circumference, both in centimetres, of 13 people. A least squares line has been fitted to the scatterplot with *ankle* circumference as the explanatory variable.



Question 7

The equation of the least squares line is closest to

A. *ankle* = 10.2 + 0.342 × *wrist*

B. *wrist* = 10.2 + 0.342 × *ankle*

C. *ankle* = 17.4 + 0.342 × *wrist*

D. *wrist* = 17.4 + 0.342 × *ankle*

E. *wrist* = 17.4 + 0.731 × *ankle*

Question 8

When the least squares line on the scatterplot is used to predict the wrist circumference of the person with an ankle circumference of 24 cm, the residual will be closest to

A. –0.7

B. –0.4

C. –0.1

D. 0.4

E. 0.7

Question 9

The residuals for the least squares line have a mean of 0.02 cm and a standard deviation of 0.4 cm.

The value of the residual for one of the data points is found to be –0.3 cm.

The standardised value of this residual is

A. –0.8

B. –0.7

C. –0.3

D. 0.7

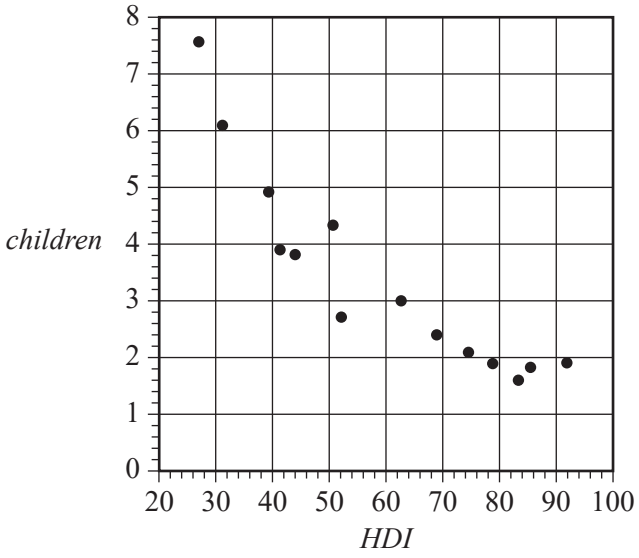
E. 0.8

Question 10

The table below gives the Human Development Index (*HDI*) and the mean number of children per woman (*children*) for 14 countries in 2007.

A scatterplot of the data is also shown.

HDI	***Children***
27.3	7.6
31.3	6.1
39.5	4.9
41.6	3.9
44.0	3.8
50.8	4.3
52.3	2.7
62.5	3.0
69.1	2.4
74.6	2.1
78.9	1.9
85.6	1.8
92.0	1.9
83.4	1.6



Data: Gapminder

The scatterplot is non-linear.

A log transformation applied to the variable *children* can be used to linearise the scatterplot.

With *HDI* as the explanatory variable, the equation of the least squares line fitted to the linearised data is closest to

A. $\log_{10}(children) = 1.1 - 0.0095 \times HDI$

B. $children = 1.1 - 0.0095 \times \log_{10}(HDI)$

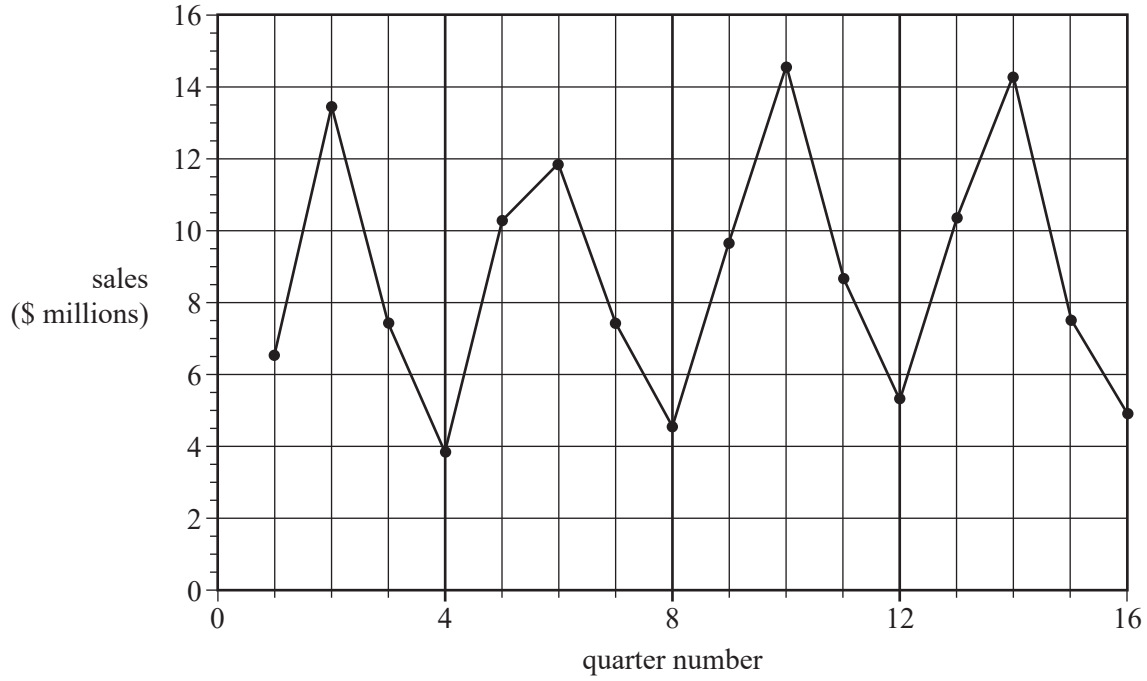
C. $\log_{10}(children) = 8.0 - 0.77 \times HDI$

D. $children = 8.0 - 0.77 \times \log_{10}(HDI)$

E. $\log_{10}(children) = 21 - 10 \times HDI$

Use the following information to answer Questions 11 and 12.

The time series plot below charts the quarterly sales figures, in millions of dollars, of a small manufacturing business over a period of four years.



Question 11

The time series plot is best described as having

A. seasonality only.

B. irregular fluctuations only.

C. seasonality with irregular fluctuations.

D. a decreasing trend with seasonality and irregular fluctuations.

E. an increasing trend with seasonality and irregular fluctuations.

Question 12

The sales figures used to generate this time series plot are displayed in the table below.

Year	Quarter 1	Quarter 2	Quarter 3	Quarter 4
2018	6.5	13.4	7.4	3.8
2019	10.2	11.8	7.4	4.5
2020	9.6	14.5	8.6	5.3
2021	10.3	14.2	7.5	4.9

The four-mean smoothed sales with centring for Quarter 3 in 2020, in millions of dollars, was closest to

A. 8.6

B. 9.3

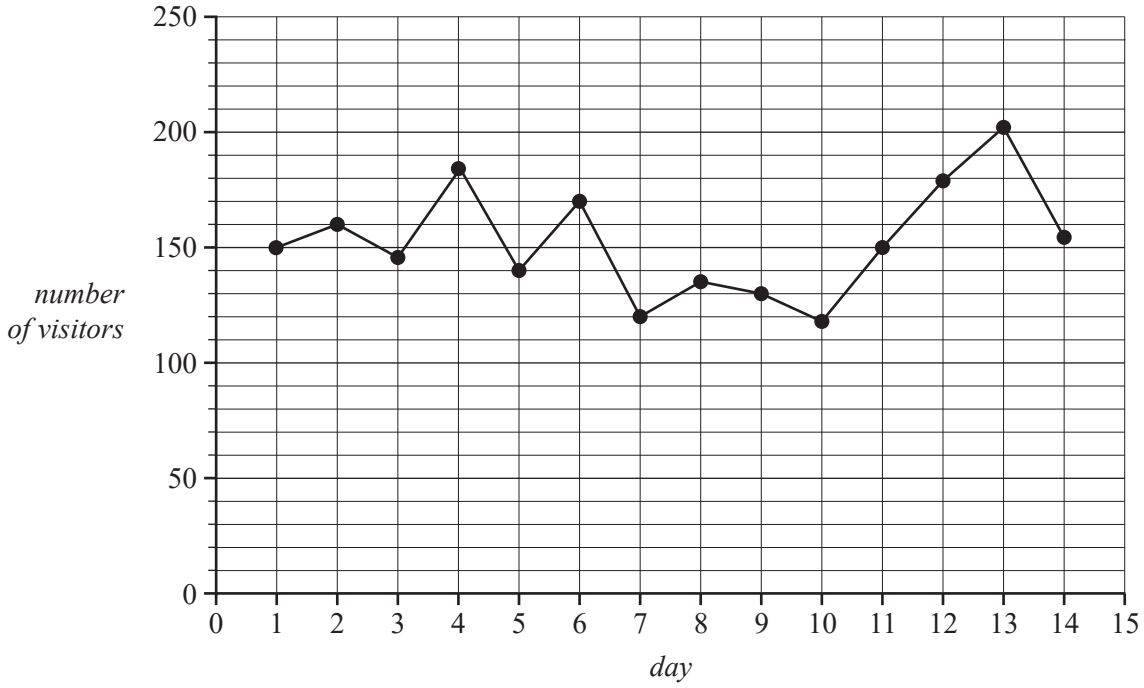
C. 9.5

D. 9.6

E. 9.7

Question 13

The time series plot below shows the daily number of visitors to a historical site over a two-week period.



This time series plot is to be smoothed using seven-median smoothing.

The seven-median smoothed number of visitors on day 4 is closest to

A. 120

B. 140

C. 145

D. 150

E. 160

Use the following information to answer Questions 14–16.

The table below shows the long-term average of the number of meals served each day at a restaurant. Also shown is the daily seasonal index for Monday through to Friday.

	Day of the week						
	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**
Long-term average	89	93	110	132	145	190	160
Seasonal index	0.68	0.71	0.84	1.01	1.10		

Question 14

The seasonal index for Wednesday is 0.84

This tells us that, on average, the number of meals served on a Wednesday is

A. 16% less than the daily average.

B. 84% less than the daily average.

C. the same as the daily average.

D. 16% more than the daily average.

E. 84% more than the daily average.

Question 15

Last Tuesday, 108 meals were served in the restaurant.

The deseasonalised number of meals served last Tuesday was closest to

A. 77

B. 93

C. 110

D. 131

E. 152

Question 16

The seasonal index for Saturday is closest to

A. 1.22

B. 1.31

C. 1.38

D. 1.45

E. 1.49

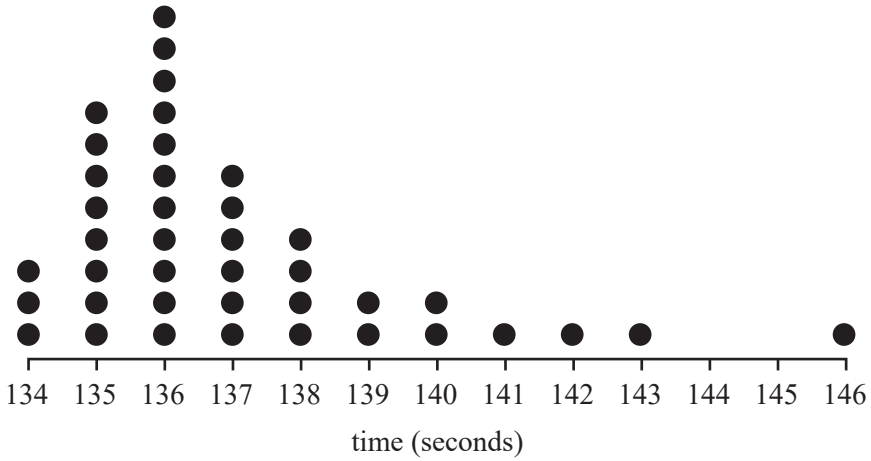
Instructions

Answer **all** questions in pencil on the answer sheet provided for multiple-choice questions.
Choose the response that is **correct** for the question.
A correct answer scores 1; an incorrect answer scores 0.
Marks will **not** be deducted for incorrect answers.
No marks will be given if more than one answer is completed for any question.
Unless otherwise indicated, the diagrams in this book are **not** drawn to scale.

Data analysis

Use the following information to answer Questions 1 and 2.

The dot plot below shows the times, in seconds, of 40 runners in the qualifying heats of their 800 m club championship.



Question 1

The median time, in seconds, of these runners is

A. 135.5
B. 136
C. 136.5
D. 137
E. 137.5

Question 2

The shape of this distribution is best described as

A. positively skewed with one or more possible outliers.
B. positively skewed with no outliers.
C. approximately symmetric with one or more possible outliers.
D. approximately symmetric with no outliers.
E. negatively skewed with one or more possible outliers.

Question 3

Gemma's favourite online word puzzle allows her 12 attempts to guess a mystery word.

Her number of attempts for the last five days is displayed in the table below.

Day	Number of attempts
1	8
2	11
3	5
4	6
5	9

On day six, how many attempts can she make so that the mean number of attempts for these six days is exactly eight?

A. 5

B. 6

C. 7

D. 8

E. 9

Question 4

The time spent by visitors in a museum is approximately normally distributed with a mean of 82 minutes and a standard deviation of 11 minutes.

2380 visitors are expected to visit the museum today.

Using the 68–95–99.7% rule, the number of these visitors who are expected to spend between 60 and 104 minutes in the museum is

A. 1128

B. 1618

C. 2256

D. 2261

E. 2373

Question 5

The heights of a group of Year 8 students have a mean of 163.56 cm and a standard deviation of 8.14 cm.

One student's height has a standardised z-score of -0.85.

This student's height, in centimetres, is closest to

A. 155.4

B. 156.6

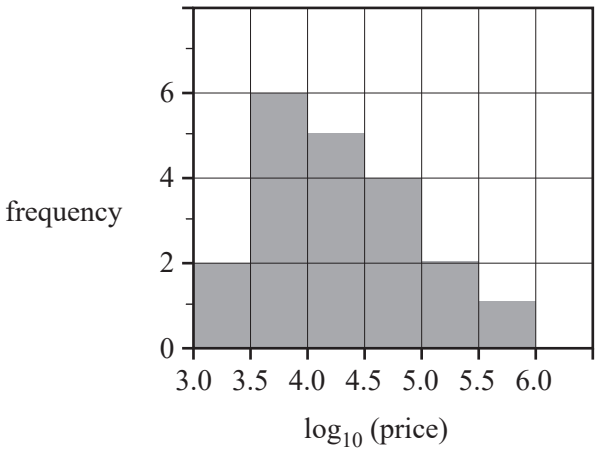
C. 162.7

D. 170.5

E. 171.7

Question 6

The histogram below displays the distribution of prices, in dollars, of the cars for sale in a used-car yard.
The histogram has a logarithm (base 10) scale.



Six of the cars in the yard have the following prices:

- \$2450
- \$3175
- \$4999
- \$8925
- \$10 250
- \$105 600

How many of the six car prices listed above are in the modal class interval?

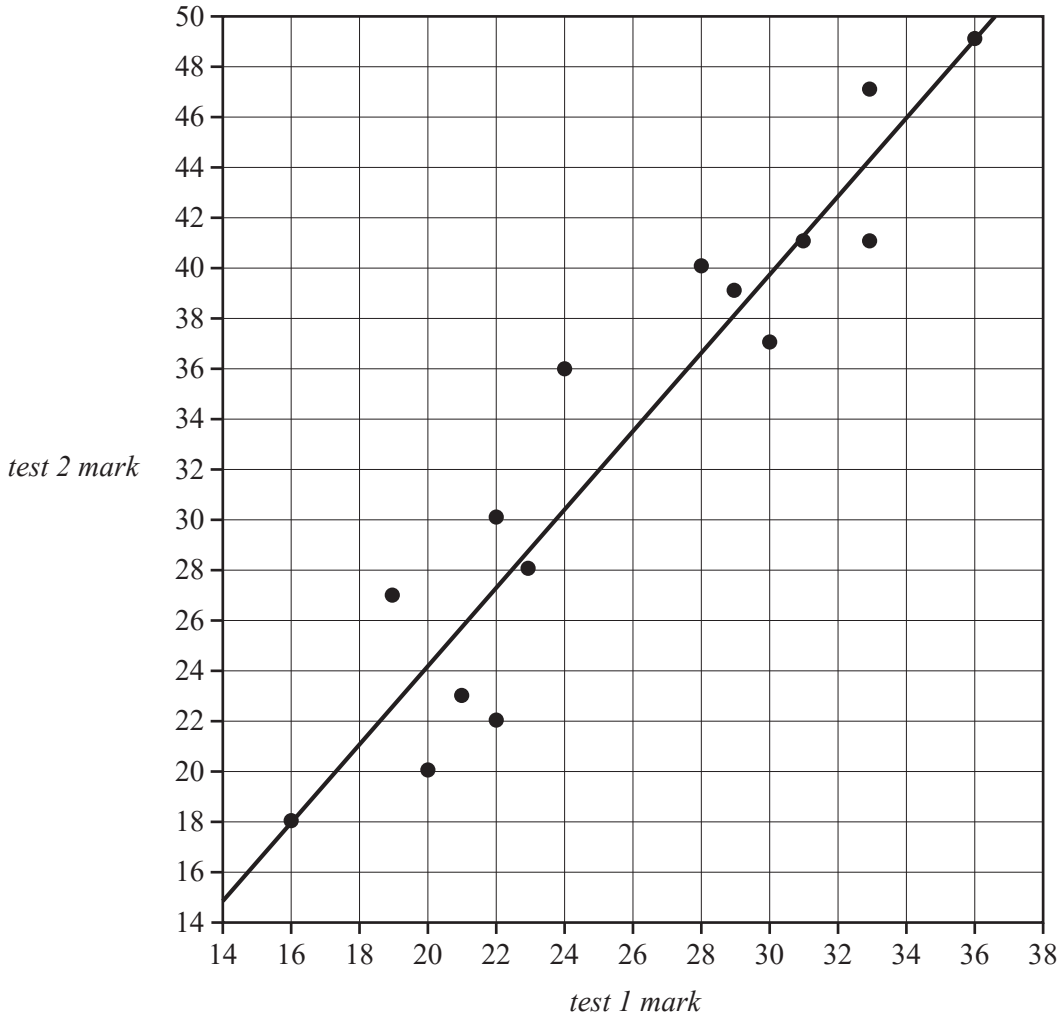
A. 1
B. 2
C. 3
D. 4
E. 6

Use the following information to answer Questions 7 and 8.

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 least squares line has been fitted to the scatterplot.



Question 7

The equation of the least squares line is closest to

A. *test 2 mark* = −6.83 + 1.55 × *test 1 mark*

B. *test 2 mark* = 15.05 + 0.645 × *test 1 mark*

C. *test 2 mark* = −6.78 + 0.645 × *test 1 mark*

D. *test 2 mark* = 1.36 + 1.55 × *test 1 mark*

E. *test 2 mark* = 6.83 + 1.55 × *test 1 mark*

Question 8

The least squares line shows the predicted *test 2 mark* for each student based on their *test 1 mark*.

The number of students whose actual *test 2 mark* was within two marks of that predicted by the line is

A. 3

B. 4

C. 5

D. 6

E. 7

Question 9

A least squares line can be used to model the *birth rate* (children per 1000 population) in a country from the average daily food *energy intake* (megajoules) in that country.

When a least squares line is fitted to data from a selection of countries it is found that:

- for a country with an average daily food *energy intake* of 8.53 megajoules, the *birth rate* will be 32.2 children per 1000 population
- for a country with an average daily food *energy intake* of 14.9 megajoules, the *birth rate* will be 9.9 children per 1000 population.

The slope of this least squares line is closest to

A. −4.7

B. −3.5

C. −0.29

D. 2.7

E. 25

Question 10

A study of Year 10 students shows that there is a negative association between the scores of topic tests and the time spent on social media. The coefficient of determination is 0.72

From this information it can be concluded that

A. a decreased time spent on social media is associated with an increased topic test score.

B. less time spent on social media causes an increase in topic test performance.

C. an increased time spent on social media is associated with an increased topic test score.

D. too much time spent on social media causes a reduction in topic test performance.

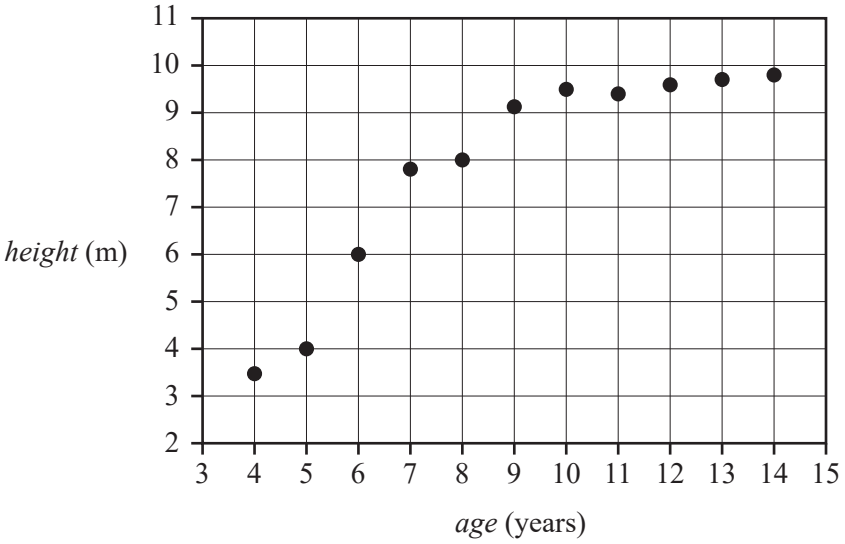
E. a decreased time spent on social media is associated with a decreased topic test score.

Use the following information to answer Questions 11 and 12.

The table below shows the *height*, in metres, and the *age*, in years, for 11 plantation trees.

A scatterplot displaying this data is also shown.

age **(years)**	***height*** **(m)**
10	9.5
8	8.0
13	9.7
9	9.1
11	9.4
14	9.8
6	6.0
4	3.5
12	9.6
7	7.8
5	4.0



Question 11

A reciprocal transformation applied to the variable *age* can be used to linearise the scatterplot.

With $\frac{1}{age}$ as the explanatory variable, the equation of the least squares line fitted to the linearised data is closest to

A. $height = -13.04 + 40.22 \times \frac{1}{age}$

B. $height = -10.74 + 8.30 \times \frac{1}{age}$

C. $height = 2.14 + 0.63 \times \frac{1}{age}$

D. $height = 13.04 - 40.22 \times \frac{1}{age}$

E. $height = 16.56 - 22.47 \times \frac{1}{age}$

Question 12

The scatterplot can also be linearised using a logarithm (base 10) transformation applied to the variable *age*. The equation of the least squares line is

$$height = -3.8 + 12.6 \times \log_{10}(age)$$

Using this equation, the *age*, in years, of a tree with a *height* of 8.52 m is closest to

A. 7.9

B. 8.9

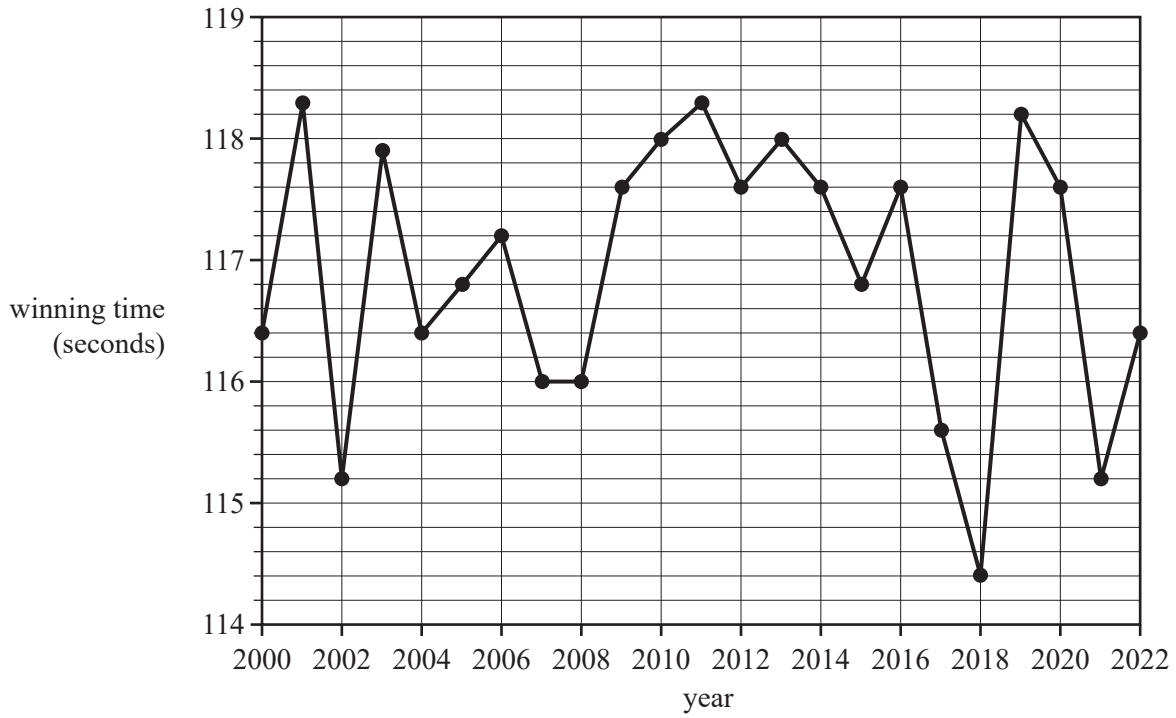
C. 9.1

D. 9.5

E. 9.9

Use the following information to answer Questions 13 and 14.

The following graph shows a selection of winning times, in seconds, for the women's 800 m track event from various athletic events worldwide. The graph shows one winning time for each calendar year from 2000 to 2022.



Data: https://www.worldathletics.org/records

Question 13

The time series is smoothed using seven-median smoothing.

The smoothed value for the winning time in 2006, in seconds, is closest to

A. 116.0

B. 116.4

C. 116.8

D. 117.2

E. 117.6

Question 14

The median winning time, in seconds, for all the calendar years from 2000 to 2022 is closest to

A. 116.8

B. 117.2

C. 117.6

D. 118.0

E. 118.3

Question 15

The number of visitors to a public library each day for 10 consecutive days was recorded.

These results are shown in the table below.

Day number	1	2	3	4	5	6	7	8	9	10
Number of visitors	337	317	313	335	322	335	322	338	302	349

The eight-mean smoothed number of visitors with centring for day number 6 is

A. 323

B. 324

C. 325

D. 326

E. 327

Question 16

The number of visitors each month to a zoo is seasonal.

To correct the number of visitors in January for seasonality, the actual number of visitors, to the nearest percent, is increased by 35%**.**

The seasonal index for that month is closest to

A. 0.61

B. 0.65

C. 0.69

D. 0.74

E. 0.77

Instructions

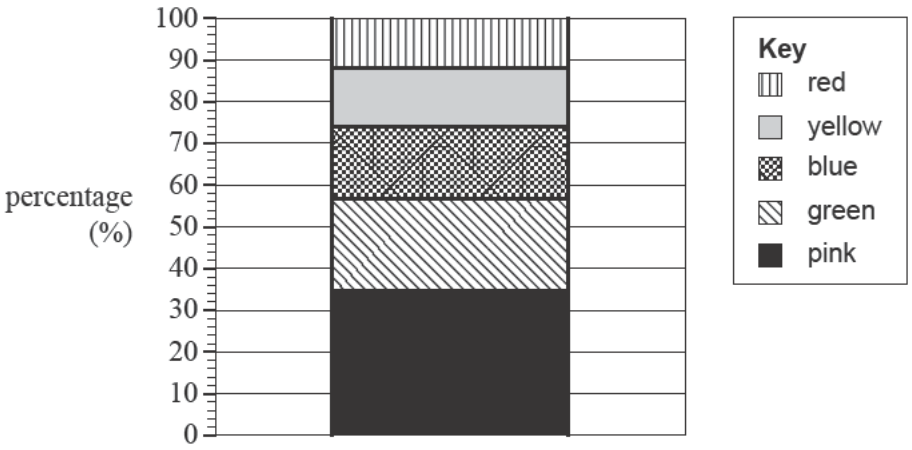
- Answer **all** questions in pencil on your Multiple-Choice Answer Sheet.
- Choose the response that is **correct** for the question.
- A correct answer scores 1; an incorrect answer scores 0.
- Marks will **not** be deducted for incorrect answers.
- No marks will be given if more than one answer is completed for any question.
- Unless otherwise indicated, the diagrams in this book are **not** drawn to scale.

Data analysis

Question 1

A group of students were asked to name their favourite colour.

The results are displayed in the percentage segmented bar chart below.



The percentage of students who named blue as their favourite colour is closest to

A. 14%

B. 18%

C. 26%

D. 74%

Question 2

Freddie organised a function at work. He surveyed the staff about their preferences.

He asked them about their *payment preference* (cash or electronic payment) and their *budget preference* (less than \$50 or more than \$50).

The variables in this survey, *payment preference* and *budget preference*, are

A. both categorical variables.

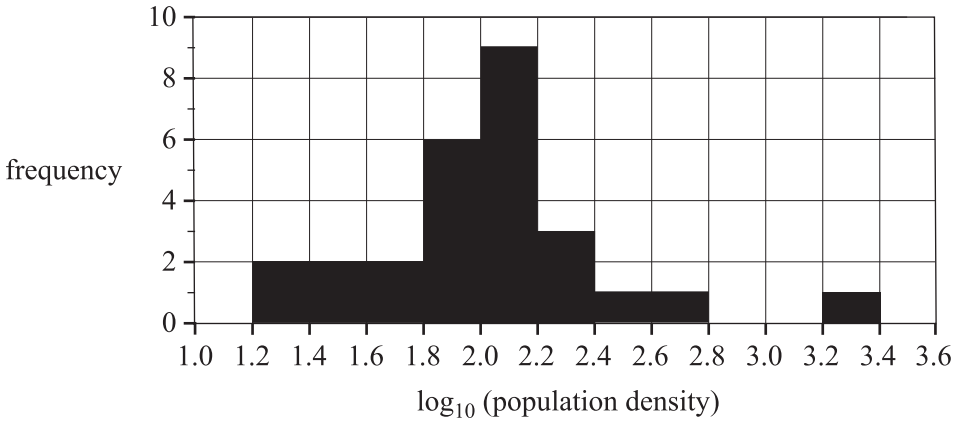
B. both numerical variables.

C. categorical and numerical variables, respectively.

D. numerical and categorical variables, respectively.

Use the following information to answer Questions 3 and 4.

The histogram below displays the population density, in people per km^2, of the 27 countries in the European Union in 2021. The histogram has a logarithmic (base 10) scale.



Data: Adapted from <https://ec.europa.eu/eurostat>

Question 3

The median value occurs in a column with a frequency of

A. 2

B. 3

C. 6

D. 9

Question 4

There is one outlier at the upper end of the histogram.
This value could be

A. 330

B. 1330

C. 2030

D. 2730

Question 5

The number of siblings of each member of a class of 24 students was recorded.

The results are displayed in the table below.

2	1	3	2	1	1	1	4	1	1	1	1
1	2	1	2	2	1	3	4	2	2	3	1

A boxplot was constructed to display the spread of the data.

Which one of the following statements about this boxplot is correct?

A. There are no outliers.

B. The value of the interquartile range (IQR) is 1.5

C. The value of the median is 1.5

D. All of the five-number summary values are whole numbers.

Question 6

More than 11 000 athletes from more than 200 countries competed in the Tokyo Summer Olympic Games.

An analysis of the number of athletes per country produced the following five-number summary.

Minimum	First quartile	Median	Third quartile	Maximum
2	5	11	48	613

Data: <https://olympics.com>

The smallest number of athletes per country that would display as an outlier on a boxplot of this data is

A. 49

B. 112

C. 113

D. 613

Question 7

Fiona plays nine holes of golf each week, and records her *score*.

Her mean *score* for all rounds in 2024 is 55.7

In one round, when she recorded a *score* of 48, her standardised score was $z = -1.75$

The standard deviation for *score* in 2024 is

A. 1.1

B. 2.3

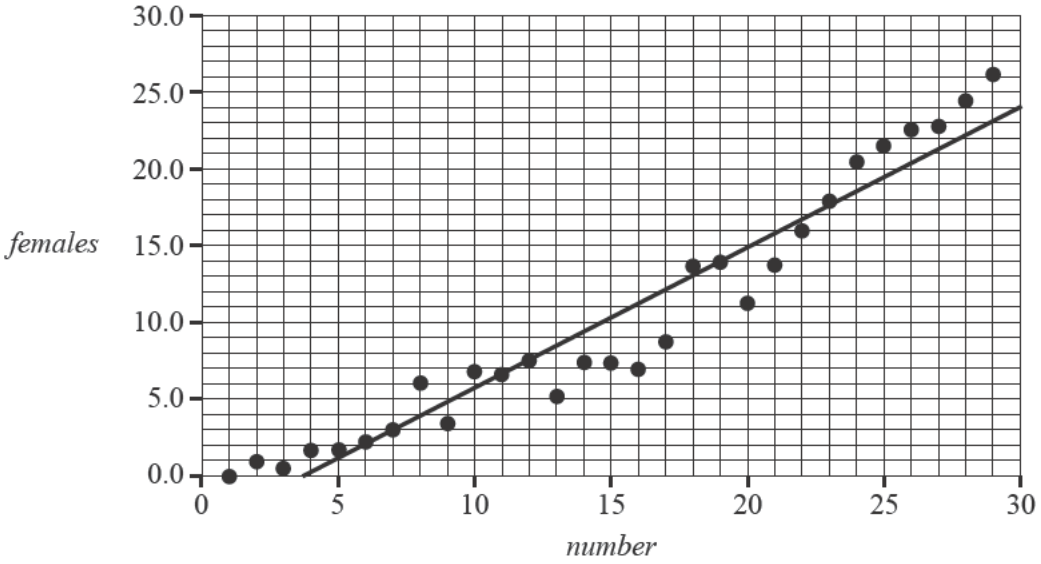
C. 4.4

D. 6.95

Question 8

The scatterplot below displays the average number of female athletes per competing nation, *females*, against the number of the Summer Olympic Games, *number*, from the first Olympic Games, in 1896, to the 29th Olympic Games, held in 2021.

A least squares line has been fitted to the scatterplot.



Data: Adapted from <www.statista.com>

The equation of the least squares line is closest to

A. $females = -4.87 + 1.02 \times number$

B. $females = -3.39 + 0.91 \times number$

C. $number = -3.39 + 0.91 \times females$

D. $number = -0.91 + 3.39 \times females$

Use the following information to answer Questions 9 and 10.

The least squares equation for the relationship between the average number of male athletes per competing nation, *males*, and the number of the Summer Olympic Games, *number*, is

$$males = 67.5 - 1.27 \times number$$

Question 9

The summary statistics for the variables *number* and *males* are shown in the table below.

	number	***males***
mean	15.0	48.4
standard deviation	8.51	19.0

The value of Pearson's correlation coefficient, r, rounded to three decimal places, is closest to

A. –0.569

B. –0.394

C. 0.394

D. 0.569

Question 10

At which Summer Olympic Games will the predicted average number of *males* be closest to 25.6?

A. 31st

B. 32nd

C. 33rd

D. 34th

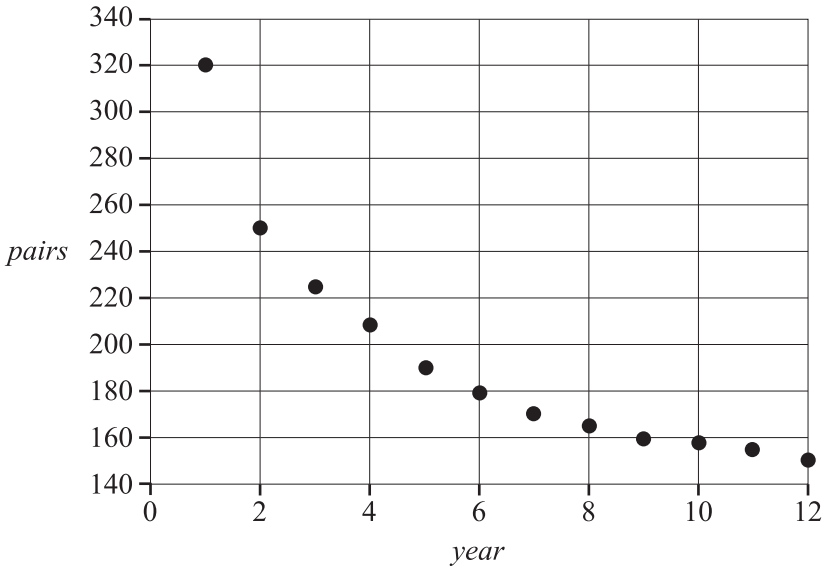
Use the following information to answer Questions 11 and 12.

The number of breeding pairs of a small parrot species has been declining over recent years.

The table below shows the number of breeding pairs counted, *pairs*, and the year number, *year*, for the last 12 years. A scatterplot of this data is also provided.

The association between *pairs* and *year* is non-linear.

year	*pairs*
1	320
2	250
3	225
4	208
5	190
6	180
7	170
8	165
9	160
10	158
11	155
12	150



Question 11

The scatterplot can be linearised using a logarithmic (base 10) transformation applied to the explanatory variable.

The least squares equation calculated from the transformed data is closest to

A. $\log_{10}(pairs) = 2.44 - 0.0257 \times year$

B. $\log_{10}(pairs) = 151 - 303 \times year$

C. $pairs = 274 - 12.3 \times \log_{10}(year)$

D. $pairs = 303 - 151 \times \log_{10}(year)$

Question 12

A reciprocal transformation applied to the variable *pairs* can also be used to linearise the scatterplot.

When a least squares line is fitted to the plot of $\frac{1}{pairs}$ versus *year*, the largest difference between the actual value and the predicted value occurs at *year*

A. 1

B. 2

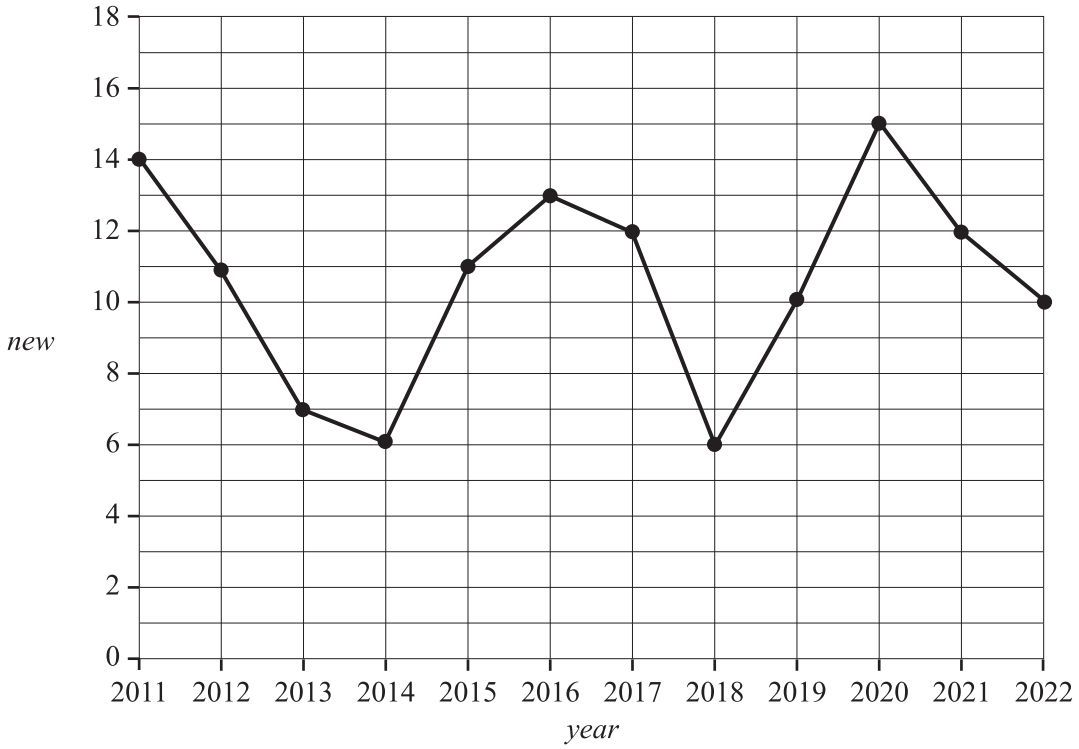
C. 11

D. 12

Use the following information to answer Questions 13 and 14.

A school runs an orientation program for new staff each January.

The time series plot below shows the number of new staff, *new*, for each year, *year*, from 2011 to 2022 (inclusive).



Question 13

The time series is smoothed using seven-median smoothing.

The smoothed value of *new* for the *year* 2016 is

A. 10

B. 11

C. 12

D. 13

Question 14

The number of new staff in 2023 is added to the total number of new staff from the previous 12 years.

For these 13 years, the mean number of new staff is 11.

The number of new staff in 2023 is

A. 11

B. 16

C. 17

D. 19

Question 15

The table below shows the total number of cans of soft drink sold each month at a suburban cafe in 2023.

Month	1	2	3	4	5	6	7	8	9	10	11	12
Cans sold	316	321	365	306	254	308	354	357	381	355	365	324

The six-mean smoothed value of the number of cans sold, with centring, for month 5 is closest to

A. 315

B. 318

C. 321

D. 324

Question 16

The table below shows the seasonal indices for the monthly takings of a bistro.

The seasonal indices for months 3 and 6 are missing.

Month	1	2	3	4	5	6	7	8	9	10	11	12
Seasonal index	1.08	1.13		0.92	0.67		1.09	1.35	0.82	0.88	1.01	0.98

The seasonal index for month 3 is twice the seasonal index for month 6.

The seasonal index for month 3 is closest to

A. 0.69

B. 1.04

C. 1.38

D. 2.07

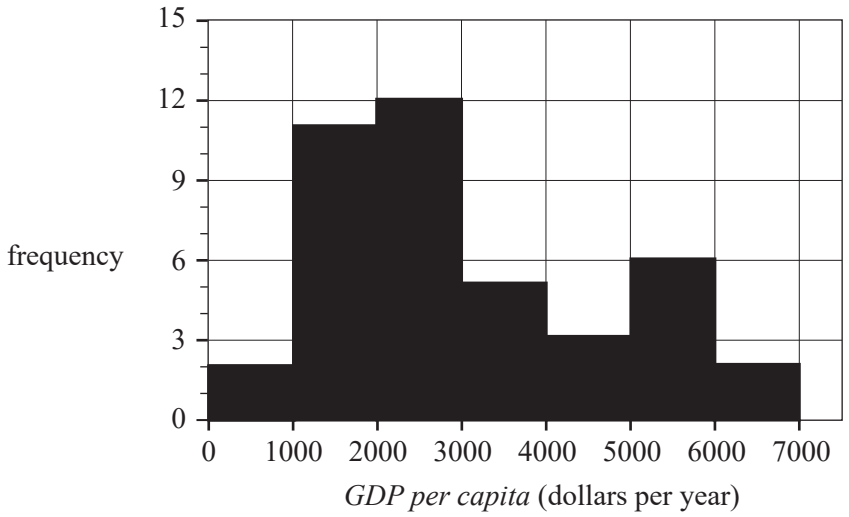
Instructions

- Answer **all** questions in pencil on the Multiple-Choice Answer Sheet.
- Choose the response that is **correct** for the question.
- A correct answer scores 1; an incorrect answer scores 0.
- Marks will **not** be deducted for incorrect answers.
- No marks will be given if more than one answer is completed for any question.
- Unless otherwise indicated, the diagrams in this book are **not** drawn to scale.

Data analysis

Question 1

The histogram below shows the distribution of the *GDP per capita* (gross domestic product per capita), in dollars per year, for a sample of 41 African countries in 2021.



Data: <data.worldbank.org>

The median *GDP per capita*, in dollars per year, for this sample will be within the range

A. greater than or equal to 0 but less than 1000

B. greater than or equal to 1000 but less than 2000

C. greater than or equal to 2000 but less than 3000

D. greater than or equal to 3000 but less than 4000

E. greater than or equal to 4000 but less than 5000

Question 2

The variables *exercise type* (aerobic, boxing, circuit) and *recovery time* (short, medium, long) are

A. a nominal variable and a numerical variable respectively.

B. a nominal variable and an ordinal variable respectively.

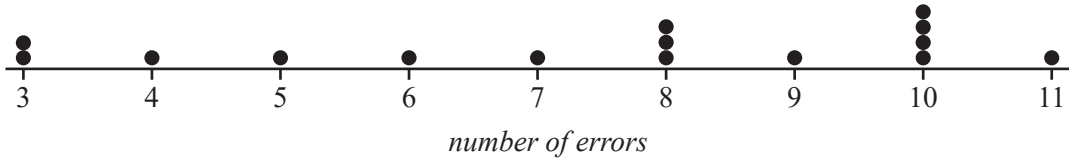
C. an ordinal variable and a nominal variable respectively.

D. both ordinal variables.

E. both nominal variables.

Question 3

The dot plot below displays the *number of errors* made in a test, for a sample of 15 students.



The mean and standard deviation for the number of errors are closest to

A. mean = 2.60 standard deviation = 7.47

B. mean = 2.70 standard deviation = 7.47

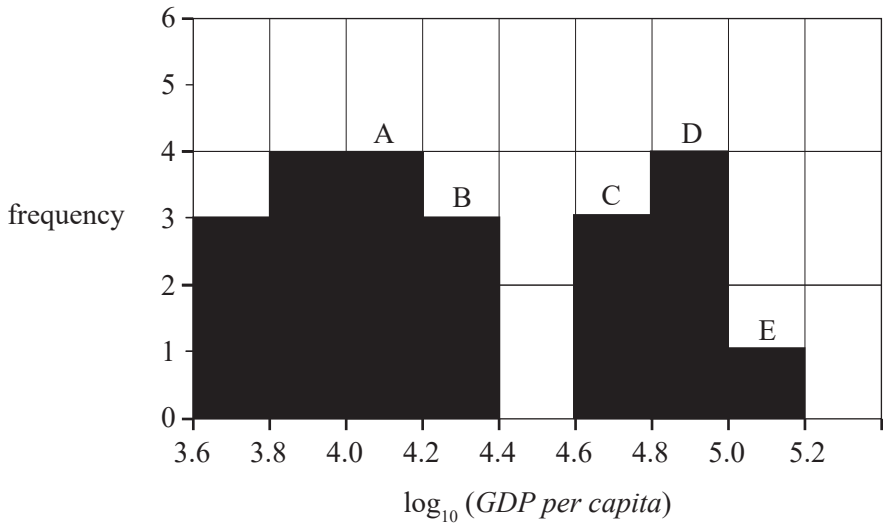
C. mean = 8.00 standard deviation = 3.00

D. mean = 7.47 standard deviation = 2.60

E. mean = 7.47 standard deviation = 2.70

Question 4

The histogram below shows the distribution of the *GDP per capita*, in dollars per year, for a sample of 22 countries in 2021 plotted on a logarithmic (base 10) scale.



Data: <data.worldbank.org>

Australia is one of these 22 countries, and its *GDP per capita* value is $56 281.

Which one of the columns marked A, B, C, D or E would contain Australia's value?

A. A

B. B

C. C

D. D

E. E

Question 5

The heights of a group of Year 9 students were measured and the standard deviation was found to be 12.25 cm.

One student with a height of 174.6 cm had a standardised score of $z = 0.45$

The mean height of this group of students, in centimetres, was closest to

A. 161.9

B. 169.1

C. 180.1

D. 186.4

E. 187.3

Question 6

The weights of cans of fish on a production line are approximately normally distributed with a mean of 126.4 grams and a standard deviation of 2.4 grams.

$13\,600$ cans of fish will be produced today.

Using the 68–95–99.7% rule, the number of these cans that are expected to weigh between 121.6 and 128.8 grams is

A. 6460

B. 9248

C. $10\,812$

D. $11\,084$

E. $12\,920$

Question 7

Data was collected to investigate the association between two variables:

- *age* (in years)
- *uses public transport* (yes, no).

Which one of the following is appropriate to use in the statistical display of this data?

A. a histogram

B. a least squares line

C. parallel boxplots

D. a segmented bar chart

E. a scatterplot

Question 8

A class investigation considered 20 countries and any association between the *birth rate*, per 1000 people, and the *life expectancy*, in years.

Students were given the following table of summary statistics.

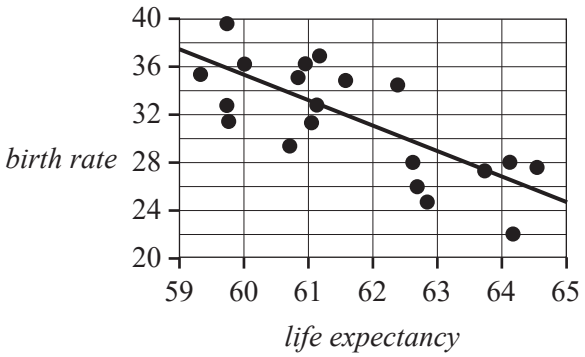
	birth rate	***life expectancy***
Mean	31.5	61.7
Standard deviation	4.70	1.64
Correlation coefficient (r)	–0.752	

Data: <data.worldbank.org> (life expectancy), CIA factbook (birth rate)

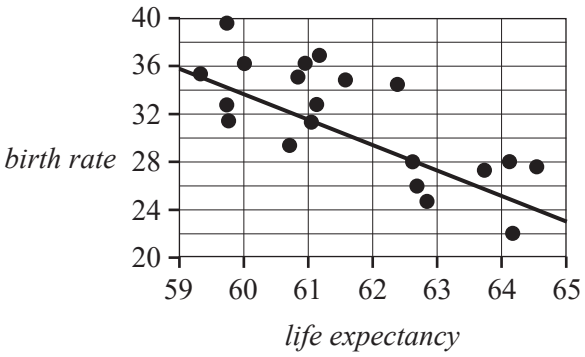
Scatterplots **A**, **B**, **C**, **D** and **E** show attempts by five students to fit the calculated least squares line to a scatterplot of the original data.

Which one of these attempts has been completed correctly?

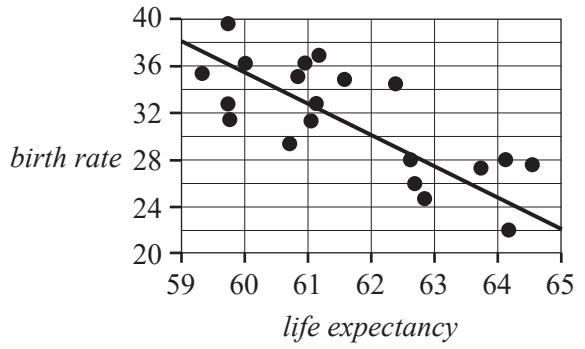
A.



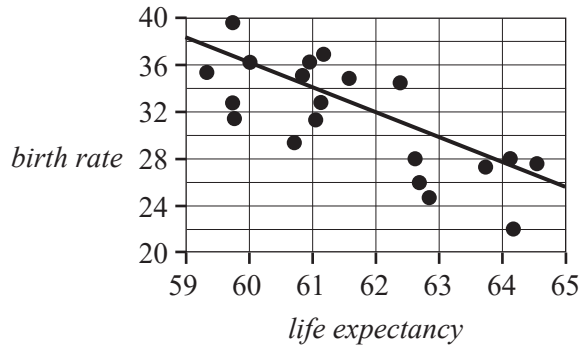
B.



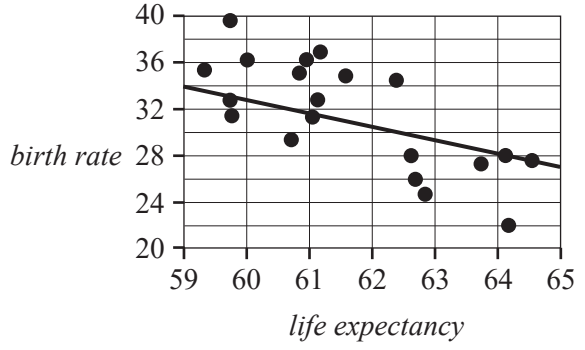
C.



D.



E.

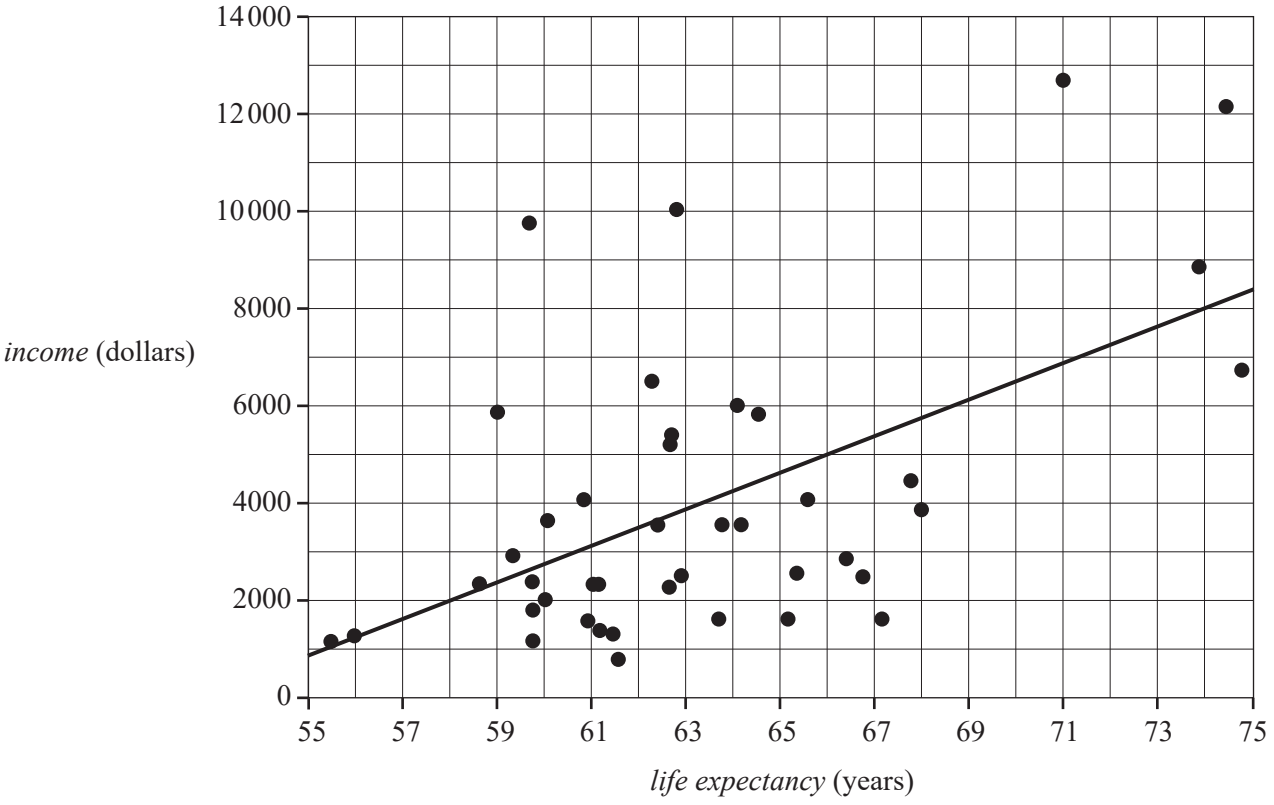


Use the following information to answer Questions 9 and 10.

The scatterplot below shows the average annual *income*, in dollars, plotted against *life expectancy*, in years, for 42 countries in 2020.

A least squares line has been fitted to the scatterplot.

The coefficient of determination is 0.306.



Data: United Nations Population Division, <data.worldbank.org>

Question 9

The equation of the least squares line is closest to

A. $income = -19\,000 + 345 \times life\ expectancy$

B. $income = -19\,250 + 355 \times life\ expectancy$

C. $income = -19\,500 + 365 \times life\ expectancy$

D. $income = -19\,750 + 375 \times life\ expectancy$

E. $income = -20\,000 + 385 \times life\ expectancy$

Question 10

Which one of the following statements is true?

A. The value of the correlation coefficient is 0.306

B. There are more data points above the least squares line than below.

C. 30.6% of the variation in annual *income* is not explained by the variation in *life expectancy*.

D. The country with the longest *life expectancy* has a positive residual associated with it.

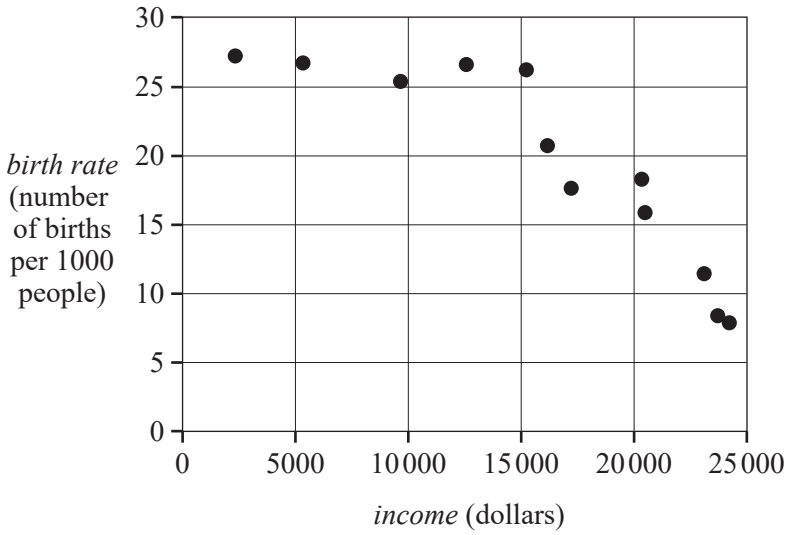
E. Using the least squares line to predict the annual *income* of a country whose citizens have a *life expectancy* of 54 years is an example of extrapolation.

Use the following information to answer Questions 11 and 12.

The table below shows the *birth rate*, in number of births per 1000 people, and the average annual *income*, in dollars per person, for a sample of 12 countries.

A scatterplot displaying the data is also shown.

birth rate	*income*
27.18	2460
26.78	5211
25.21	9846
26.44	12 706
26.14	15 176
20.60	16 304
17.55	17 402
18.24	20 463
15.69	20 615
11.35	23 285
8.15	23 926
7.72	24 364



Data: <data.worldbank.org>, <www.cia.gov/the-world-factbook/>

Question 11

A squared transformation applied to the variable *birth rate* can be used to linearise the scatterplot.

The equation of the least squares line is

$$(birth\ rate)^2 = 953 - 0.0333 \times income$$

Using this equation, the predicted *birth rate*, in number of births per 1000 people, for a country with an average annual income of $\$18\,500$ is closest to

A. 17.7

B. 18.0

C. 18.4

D. 337

E. 113 535

Question 12

Coefficients of determination were calculated for

- *birth rate* vs *income* (coefficient 1)
- $(birth\ rate)^2$ vs *income* (coefficient 2)
- *birth rate* vs $(income)^2$ (coefficient 3)

These coefficients were ranked in order from largest to smallest.

The order would be

A. coefficient 1, coefficient 2, coefficient 3

B. coefficient 2, coefficient 3, coefficient 1

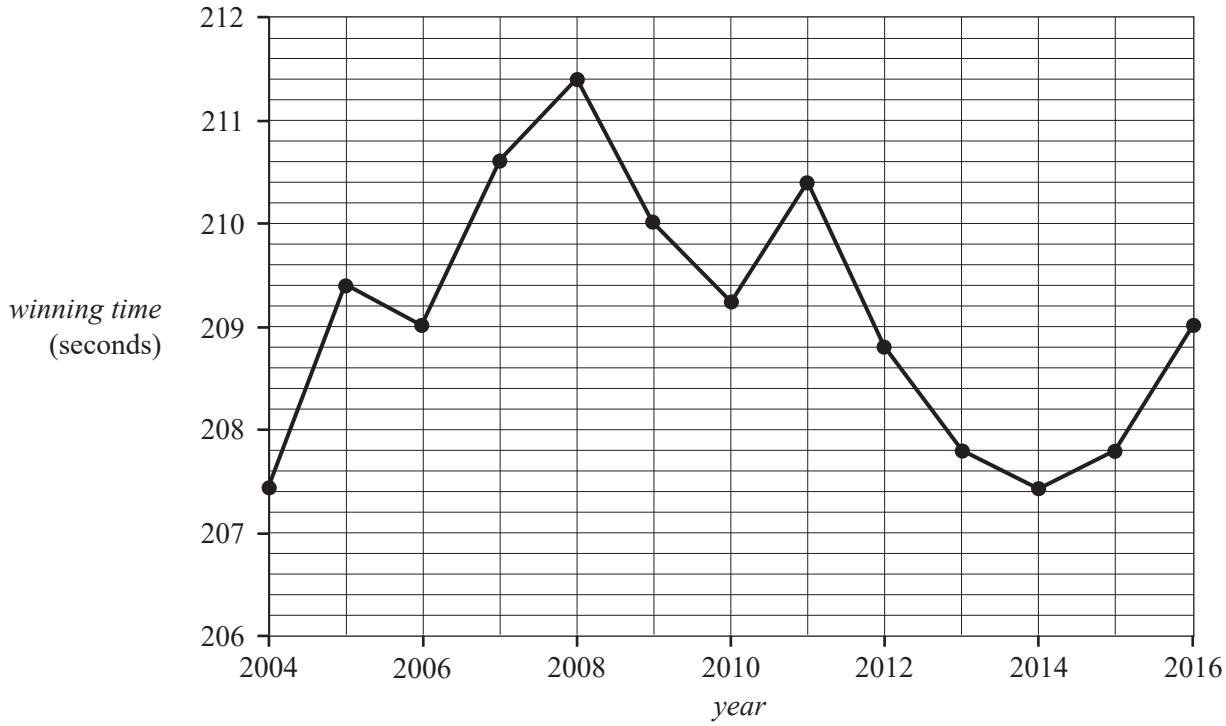
C. coefficient 2, coefficient 1, coefficient 3

D. coefficient 3, coefficient 1, coefficient 2

E. coefficient 3, coefficient 2, coefficient 1

Question 13

The following graph shows the *winning time*, in seconds, for each *year* from 2004 to 2016 for a men's 1500 m track event.



Data based on: <https://www.worldathletics.org>

The time series is smoothed using nine-median smoothing.

The smoothed value for the *winning time* in 2009, in seconds, is closest to

A. 209.0

B. 209.2

C. 209.4

D. 210.0

E. 210.4

Question 14

The number of fruit drinks sold by a fast-food shop each day for 14 consecutive days was recorded.

These results are shown in the table below.

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Number	65	84	123	154	129	187	179	71	89	131	147	141	190	185

The six-mean smoothed value with centring for day 8 is

A. 132

B. 132.5

C. 133

D. 133.5

E. 134

Question 15

The sales revenue, in dollars, from the sale of chocolate eggs is seasonal.

To correct the sales revenue in May for seasonality, the actual sales revenue, to the nearest percent, is decreased by 17%.

The seasonal index for that month is closest to

A. 0.77

B. 0.83

C. 1.17

D. 1.20

E. 1.25

Question 16

Seasonal indices for visitor numbers to a theme park in a particular year are given in the table below. The seasonal index for summer is not given.

Season	Spring	Summer	Autumn	Winter
Seasonal index	0.85		0.96	0.45

In this particular year, 33 120 visitors attended during summer.

The total annual attendance for this particular year is closest to

A. 73 960

B. 74 520

C. 75 820

D. 76 140

E. 77 380