Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel Level 1/Level 2 GCSE (9–1)

Monday 17 June 2024

Afternoon (Time: 1 hour 30 minutes)

Paper reference **1ST0/2H**

Statistics

PAPER 2
Higher Tier

You must have:
Ruler graduated in centimetres and millimetres, protractor, pair of compasses, pen, HB pencil, eraser, scientific calculator.

Total Marks

Instructions

- Use **black** ink or ball-point pen.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions.
- Answer the questions in the spaces provided
  – *there may be more space than you need.*
- Scientific calculators may be used.
- You must **show all your working out** with **your answer clearly identified** at the **end of your solution**.

Information

- The total mark for this paper is 80.
- The marks for **each** question are shown in brackets
  – *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over ▶

P75439A

F:1/1/1/1/1/1/

P 7 5 4 3 9 A 0 1 2 4

Pearson

Higher Tier Formulae

You must not write on this page.

Anything you write on this page will gain NO credit.

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

$$\text{Standard deviation} = \sqrt{\frac{1}{n}\sum (x - \bar{x})^2}$$

An alternative formula for standard deviation is

$$\text{standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Spearman's rank correlation coefficient

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

Rates of change (e.g. Crude birth rate = $\frac{\text{number of births} \times 1000}{\text{total population}}$)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 Matthew is investigating average household income for different states in the USA.

(a) Give a reason why it is appropriate to use secondary data for this.

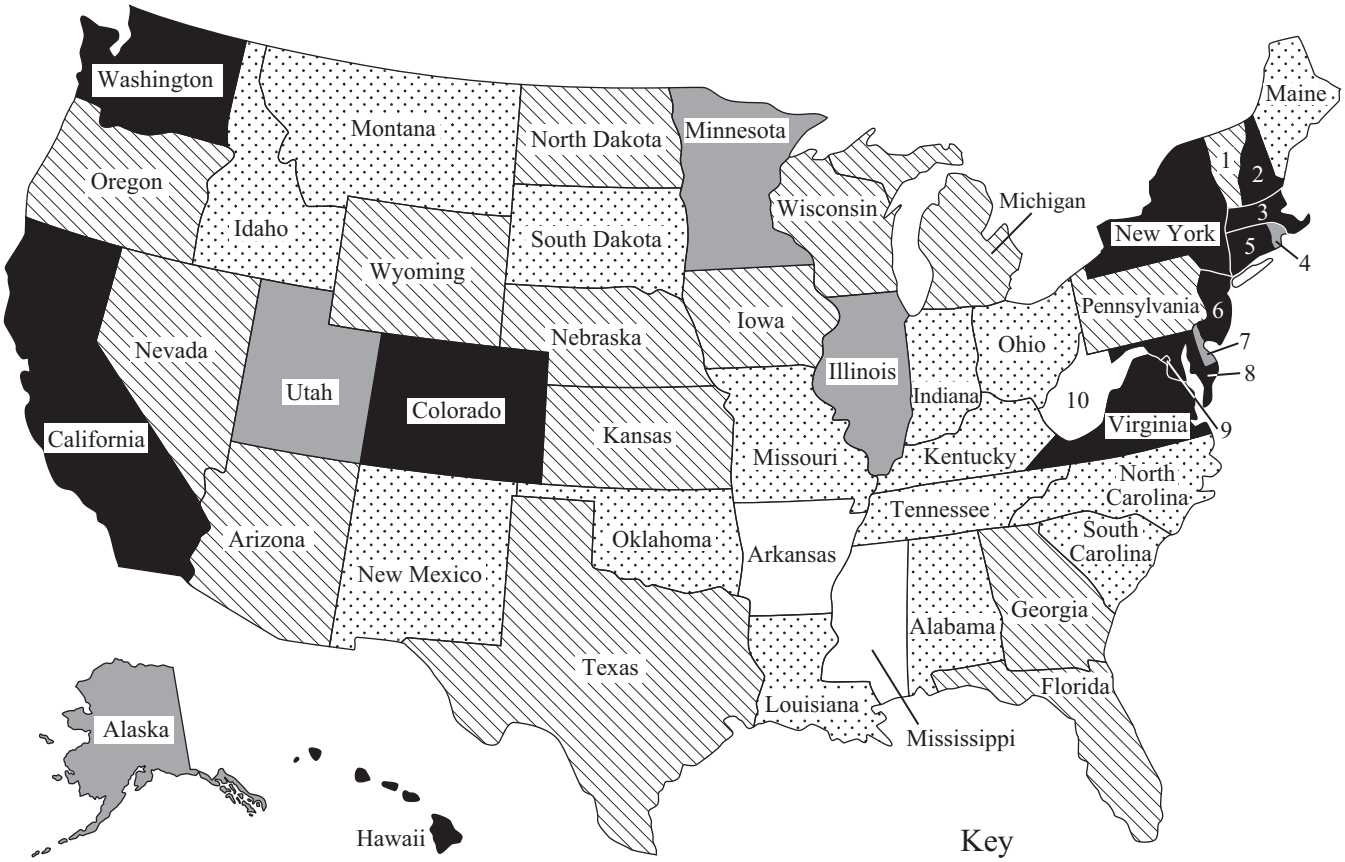
.....

.....

(1)

Matthew creates a choropleth map giving information about the mean household income by state for the USA in 2023

Mean annual household income in $ thousands.



1. Vermont
2. New Hampshire
3. Massachusetts
4. Rhode Island
5. Connecticut
6. New Jersey
7. Delaware
8. Maryland
9. District of Columbia
10. West Virginia

(Source: *worldpopulationreview.com*)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(b) Which **three** states have the lowest mean household income?

.....

(1)

Matthew concludes that the mean household incomes are highest on the West coast and the East coast.

(c) Does the choropleth map support this conclusion?
Give a reason for your answer.

.....

.....

.....

(2)

(Total for Question 1 is 4 marks)

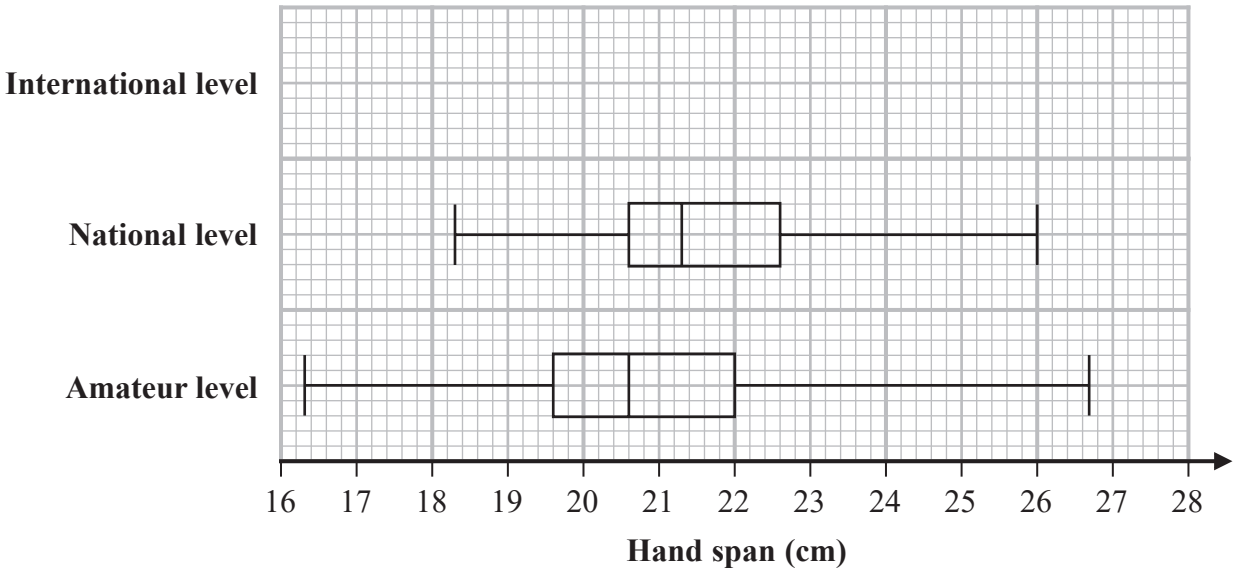
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

2 Some researchers investigated the hand span, in centimetres, of adult pianists by their level – international, national and amateur.

The box plots below give information about the hand spans for national level and amateur level pianists.



(Source: *www.appca.com.au*)

(a) Circle the word in the list below that describes hand span, in centimetres, as a type of data.

qualitative ordinal continuous bivariate

(1)

The table gives information about the hand spans of the international level pianists.

Greatest hand span	27.4 cm
Median hand span	23.9 cm
Lower quartile	23.2 cm
Range	5.1 cm
Interquartile range	1.1 cm

(b) Using the information in the table, draw on the grid above a box plot for the hand spans of the international level pianists.

(3)

(c) Compare the three distributions of hand spans.
Give **three** comparisons and interpret **two** of your comparisons.

(5)

Pavel owns a music shop.
He wants to investigate the keyboard sizes used by pianists with different hand spans.
He collects data about the hand spans of the pianists who use his shop.

The table gives information about the number of these pianists with hand spans in each of four size categories.

Hand span (cm)	A (less than 19)	B ($19 \leqslant$ span < 22)	C ($22 \leqslant$ span < 24)	D (24 or more)
Number of pianists	24	65	57	14

Pavel plans to sample 20 of these pianists stratified by hand span size.

(d) Explain how Pavel can obtain his stratified sample.
You should include details of any calculations he should use.

(3)

(Total for Question 2 is 12 marks)

3 Khatia organises two different training courses, Course A and Course B, to help people to learn to type.
She wants to compare the two different courses to see which is better.
At the end of each course the people are given a skills test.

The table shows the number of participants who passed and failed the skills test for each of the two courses.

	Passed	**Failed**	**Total**
Course A	35	15	50
Course B	48	32	80

(a) Find the relative risk of failing the skills test having taken Course A compared to Course B.

.....

(3)

(b) Give an interpretation of your answer to part (a).

.....

.....

(1)

(Total for Question 3 is 4 marks)

BLANK PAGE

Turn over for next question

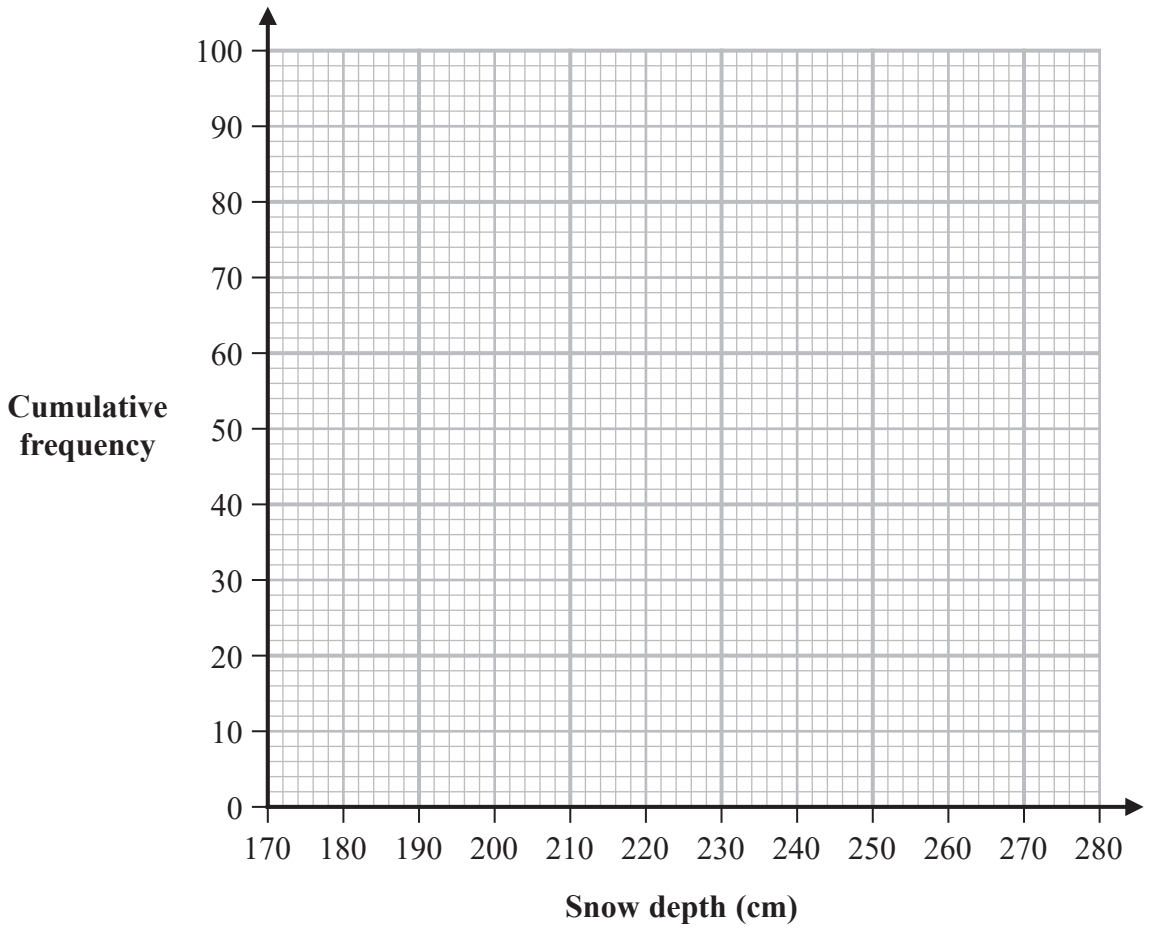
4 The snow depth, in centimetres, at Whistler Mountain was measured daily in February, in March and in April 2022

The table gives information about the snow depths for these 89 days.

Snow depth (x cm)	Cumulative frequency
$170 \leqslant x < 180$	16
$170 \leqslant x < 190$	37
$170 \leqslant x < 200$	55
$170 \leqslant x < 210$	74
$170 \leqslant x < 220$	83
$170 \leqslant x < 280$	89

(Source: *aqrt.nrs.gov.bc.ca*)

(a) Draw a cumulative frequency diagram for this information.



(2)

(b) Find an estimate for the number of these days where the snow depth was between 195 cm and 215 cm.

.....

(2)

(Total for Question 4 is 4 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

5 Zack is comparing the players from the England Rugby Union team with the players from the Welsh Rugby Union team.
The players are listed as Rugby Union Forwards or Rugby Union Backs.

Here are the weights, in kilograms, of the England Rugby Union team players who are listed as Backs.

96	112	94	87	93	88	90	92
98	96	96	82	107	111	84	88

(Source: *www.englandrugby.com/england/senior-men#squad*)

The back-to-back stem and leaf diagram gives the weights, in kilograms, of the Welsh Rugby Union Backs.

(a) Complete the back-to-back stem and leaf diagram with the weights, in kilograms, of the England Rugby Union Backs.

Welsh		**England**
8	7	
7 6 4 4 4	8	
8 6 5 5 2 1	9	
9 4 0	10	
	11	

Key:

(Source: *www.wru.wales/fixtures-and-teams/teams/wales/*)

(3)

(b) Work out the median weight for the Welsh Rugby Union Backs.

..... kg

(1)

(c) Work out the interquartile range for the Welsh Rugby Union Backs.

..... kg

(2)

The table gives information about the weights for the England Rugby Union Backs.

Median	Interquartile range
93.5 kg	9 kg

Zack thinks that the Welsh Rugby Union Backs are heavier and have less consistent weights than the England Rugby Union Backs.

(d) Do the statistics support these conclusions?
You must give reasons for your answer.

.....

.....

.....

.....

.....

(3)

Amy wants to use the median and interquartile range statistics in a news article for a sports magazine. The article will compare the players on the two teams who are Backs.

(e) Comment on the appropriateness of using the median and the interquartile range in the article.

.....

.....

(1)

(f) Give a limitation of using Zack's statistics to compare **all** the players on the two teams.

.....

.....

(1)

(Total for Question 5 is 11 marks)

6 Researchers investigated the effect of changing land usage on the amount of surface runoff water for areas of land around Brandywine Creek.

They recorded the change in forest cover area (km^2) from 1992 to 2000
They also recorded the change in the amount of surface runoff (mm water) from 1992 to 2000

The table gives information about their data.

Area of land	Change in forest cover area (km^2)	Change in surface runoff (mm water)	Change in forest cover area rank	Change in surface area runoff rank	
1	0.28	–8.75	15		
2	–0.33	3.60	11		
3	0.07	–5.44	14		
4	–1.48	213.34	7		
5	–0.39	191.29	10		
6	–2.66	106.06	3		
7	–1.52	178.85	6		
8	–3.04	137.54	2		
9	0.00	0.00	13		
10	–3.46	168.34	1		
11	–0.58	116.85	9		
12	–1.72	204.32	5		
13	–0.32	29.94	12		
14	–1.85	228.54	4		
15	–1.08	136.31	8		

(Source: *www.researchgate.net*)

(a) Calculate Spearman's rank correlation coefficient for this data.
Interpret this correlation in the context of the question.

(6)

Elizabeth wants to investigate possible drainage solutions that reduce surface runoff.
She wants to establish whether a particular drainage solution reduces surface runoff.

Elizabeth plans to measure surface runoff in 10 areas for a year without the drainage solution.
She will then install the drainage solution in these 10 areas and measure the surface runoff for a year.

(b) Explain how Elizabeth's plan to collect data controls some extraneous variables.
You should include in your answer an example of an extraneous variable that is likely to be controlled in this investigation.

(2)

Elizabeth would like to reduce the time that she is collecting data to one year overall.

(c) Describe how she could do this using a matched pairs approach.

(1)

(Total for Question 6 is 9 marks)

7 The Consumer Price Index (CPI) is a measure of the rate of change of prices in everyday life.

The table shows the annual average CPI from 2017 to 2021 with 2015 as the base year.

Year	2017	2018	2019	2020	2021
Annual average CPI	103.6	106.0	107.8	108.9	111.6

(Source: *www.ons.gov.uk*)

(a) Give an interpretation of the number 108.9 in the table.

.....

.....

(2)

(b) Find the percentage increase in consumer prices from 2017 to 2021
You must show your working.
Give your answer correct to one decimal place.

.....

(2)

A representative sample of consumer product prices is included in the calculations for CPI.
In the non-alcoholic beverages category there are two classes of product included.

The table gives the weightings for these two classes and the price index for one of the classes for November 2022

Non-alcoholic beverages	**Weights**	**Price index**
Coffee, tea and cocoa	2	115.6
Mineral waters, soft drinks and juices	7	

The price index for non-alcoholic beverages in November 2022 was 124.2

(Source: *www.ons.gov.uk*)

(c) Calculate the price index for mineral waters, soft drinks and juices in November 2022
Give your answer correct to one decimal place.

.....

(2)

The weights in the non-alcoholic beverage category were different in 2021 and in 2022

The table gives these weights.

Non-alcoholic beverages	2021	2022
Coffee, tea and cocoa	3	2
Mineral waters, soft drinks and juices	7	7

(d) Explain what this means for the relative importance in the CPI of 'Coffee, tea and cocoa' and 'Mineral waters, soft drinks and juices' in 2021 compared with 2022

.....

.....

.....

(1)

The rate of inflation is the change in prices for goods and services over time.

David wants to compare annual inflation in Germany from 2017 to 2021 with annual inflation in the UK from 2017 to 2021

The table gives the annual inflation for the UK from 2017 to 2021

Year	2017	2018	2019	2020	2021
Annual inflation	2.56%	2.29%	1.74%	0.99%	2.52%

(Source: *worlddata.info*)

From 2017 to 2021 the average annual inflation for Germany was 1.66%.

(e) By calculating an appropriate geometric mean, compare the average annual inflation for Germany from 2017 to 2021 with the average annual inflation for the UK from 2017 to 2021
You must show your working.

.....

.....

.....

(3)

(Total for Question 7 is 10 marks)

8 Roxann wants to investigate the resting heart rates for members of her running club.

She believes that the resting heart rates will be normally distributed.

Here is the plan for her investigation.

Data collection

Take a census of the 110 members of the running club (50 female runners and 60 male runners).

Each runner will measure their resting heart rate and input the information into the spreadsheet when they come to the club.

Processing and presenting data

Construct a box plot for the resting heart rate of the male runners and a box plot for the resting heart rate of the female runners.

Work out the mean, median and mode for the resting heart rate of the male runners and for the resting heart rate of the female runners.

Work out the standard deviation for the resting heart rate of the male runners and for the resting heart rate of the female runners.

Discuss whether Roxann's plans for collecting and presenting data are appropriate.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 8 is 6 marks)

9 Researchers wanted to find a method to predict the height of ancient Egyptians based upon bones from their skeletons.

They calculated the Pearson's product moment correlation coefficient between the length of particular bones and the height of the skeleton for some male skeletons and for some female skeletons.

They also found regression equations for the relationship between bone length (x centimetres) and height (y centimetres) for each of these bones in male skeletons and in female skeletons.

The table gives information about these product moment correlation coefficients and these regression equations.

	Product moment correlation coefficient	**Regression equation**
Males		
Femur	0.826	$y = 2.257x + 63.93$
Tibia	0.850	$y = 2.554x + 69.21$
Humerus	0.656	$y = 2.594x + 83.85$
Radius	0.649	$y = 2.641x + 100.91$
Females		
Femur	0.891	$y = 2.340x + 56.99$
Tibia	0.938	$y = 2.699x + 61.08$
Humerus	0.806	$y = 2.827x + 70.94$
Radius	0.580	$y = 2.509x + 96.73$

(Source: *www.semanticscholar.org*)

(a) Which bone measurement would you recommend using to estimate the height of an ancient Egyptian?
Give a reason for your answer.

.....

.....

.....

.....

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(b) Interpret in context the figure 2.594 in the regression equation for the humerus of male skeletons.

(1)

(c) Use the regression equations to compare the relationships between bone length and height for the different bones.
Include in your answer comparisons between male and female ancient Egyptians.

(3)

Dina suggests comparing the y-intercepts for the regression equations.

(d) Is Dina's suggestion appropriate?

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

A museum has some bones that were recovered from an Egyptian pyramid.

The museum wants to predict the height of the ancient Egyptians from whom the bones were recovered.

(e) Is it appropriate to use these regression equations to make this prediction?

.....

.....

.....

.....

(2)

(Total for Question 9 is 10 marks)

10 Researchers used the Petersen capture recapture method to estimate the number of humpback whales off the coasts of California, Oregon and Washington.

They used aerial photographs to identify the whales, recording the whales' unique markings in the first sampling rather than tagging the whales.

In 1995 they recorded the markings of 331 individual whales.
In 1996 they identified 264 individual whales of which 104 had been identified in the previous year.

(Source: *digitalcommons.unl.edu*)

(a) Calculate an estimate for the number of humpback whales off the coasts of California, Oregon and Washington in 1996

.....

(2)

(b) Discuss the validity and reliability of this estimate.

.....

.....

.....

.....

.....

.....

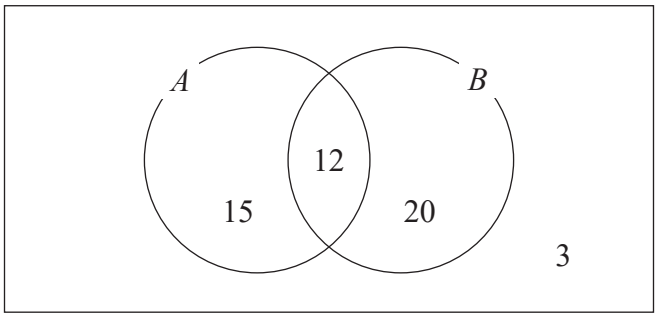
(3)

(Total for Question 10 is 5 marks)

11 The Venn diagram shows information about 50 books on a reading list.

A is the event that the book is a science fiction book.
B is the event that the book is an audio book.

The numbers in the Venn diagram show the number of books.



One of the books is chosen at random.

(a) Find P(A).

.....

(1)

(b) Find P($A \mid B$).

.....

(1)

Mike concludes that A and B are independent events.

(c) Is Mike correct?
Give a reason for your answer.

.....

.....

(1)

Two different events, C and D, are such that

$$P(C) = 0.75 \qquad P(D) = 0.4 \qquad P(C \cap D) = 0.24$$

(d) Find P($C \cup D$).

.....

(2)

(Total for Question 11 is 5 marks)

TOTAL FOR PAPER IS 80 MARKS