

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)

Time 1 hour 30 minutes

Paper
reference

1ST0/1H

Statistics

PAPER 1

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 ►

P66367A

©2022 Pearson Education Ltd.

Q:1/1/1/1/



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

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

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

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

- 1 Bill is investigating how being grown in the shade and being grown in sunlight affects the heights of tree seedlings.

The following stem and leaf diagrams give information about the heights, in centimetres, of 17 tree seedlings grown in the shade and 17 tree seedlings grown in sunlight.

The seedlings were all planted at the same time.

Shade

1	7 8 9
2	1 2 3 4 5 7 8 9
3	1 1 1 2 4 5
4	
5	

Key:

2 | 1 represents 21 cm

Sunlight

1	
2	8 9
3	3 5 5 6 7 9
4	1 2 3 5 6 7 8 9
5	1

Key:

3 | 3 represents 33 cm

Compare the average height of the tree seedlings grown in the shade with the average height of the tree seedlings grown in sunlight.

State clearly the value of each average you use in order to make your comparison.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(Total for Question 1 is 3 marks)

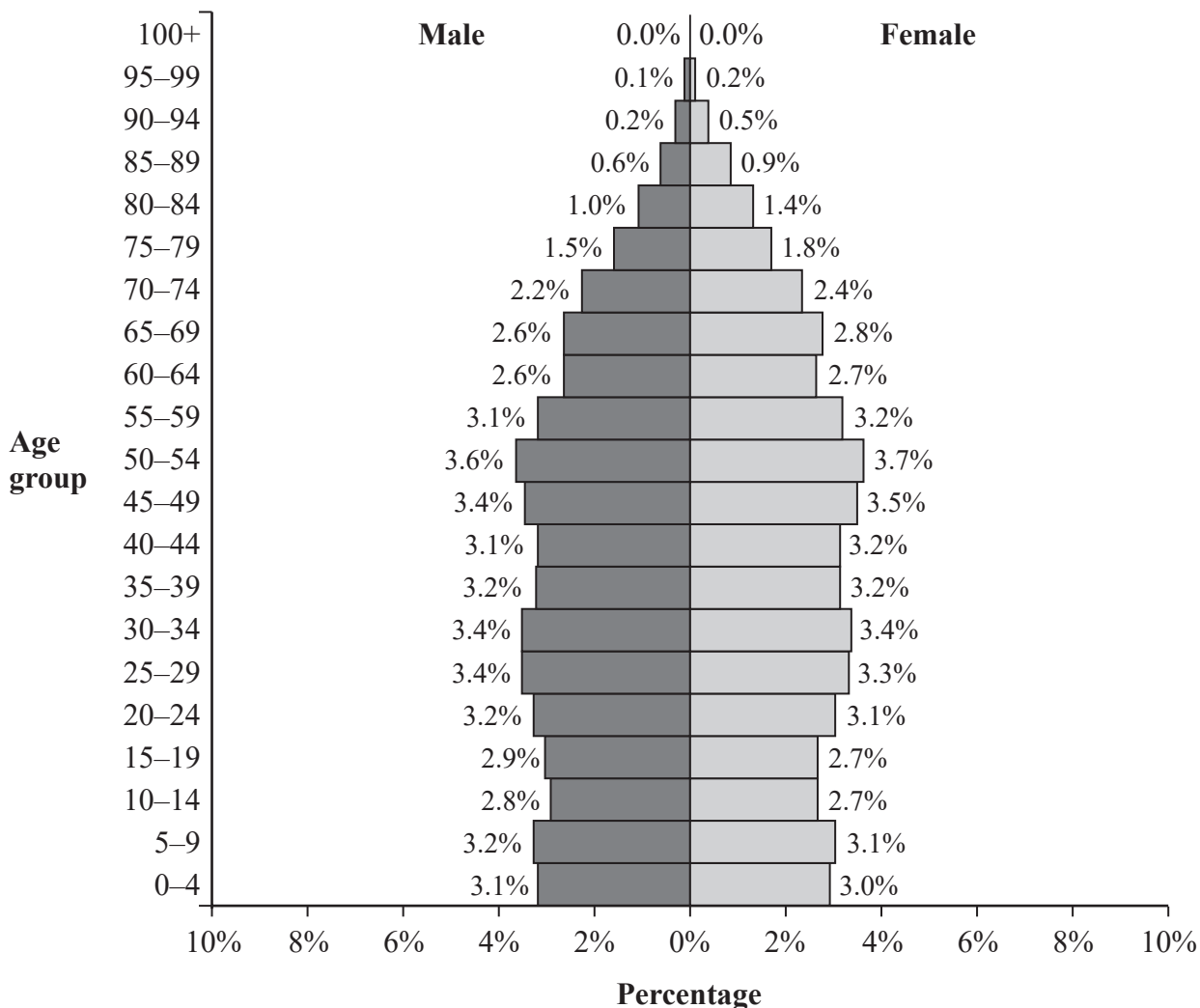


2 The population pyramid gives information about the percentages of the population of the United Kingdom who are male and who are female in each age group for 2017

Each percentage is given correct to one decimal place.

United Kingdom 2017

Population: 65 511 097



(Source: Populationpyramid.net)

(a) Write down the percentage of the population who are female in the age group 50-54

..... %
(1)

(b) Work out the percentage of the population who are male in the age group 10-19

..... %
(2)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

In 2017, the number of people age 100 and older (100+) in the United Kingdom was 13310

(Source: *Office for National Statistics*)

Using the information above and information from the population pyramid,

- (c) explain why the percentage of the population in the age group 100+ is given as 0.0% on the population pyramid.
You must show your working.

(3)

Jamie is carrying out research into the ages of people in the United Kingdom. He uses the information in the population pyramid to claim,

“In the United Kingdom in 2017 the number of males who were older than 40 was greater than the number of females who were older than 40”

- (d) Explain whether or not Jamie’s claim is correct.

(2)

(Total for Question 2 is 8 marks)



- 3 Weronika works for a road traffic organisation. One day she is investigating the speeds of cars and the speeds of motorcycles along a motorway.

Here is part of the spreadsheet that Weronika used to record her results.

Speed (s miles per hour)	Percentage of cars	Percentage of motorcycles
$30 \leq s < 40$	3	0
$40 \leq s < 50$	18	two
$50 \leq s < 60$	42	7
$60 \leq s < 70$	27	124
$70 \leq s < 80$	6	56
$80 \leq s < 90$		11
Total	100	100

- (a) Give a reason why Weronika will need to clean the data.

(1)

Weronika concludes that the value of 124 in the spreadsheet must be wrong.

- (b) Explain why.

(1)

Here is the information about motorcycles from the spreadsheet with the data cleaned.

Speed (s miles per hour)	Percentage of motorcycles
$30 \leq s < 40$	0
$40 \leq s < 50$	2
$50 \leq s < 60$	7
$60 \leq s < 70$	24
$70 \leq s < 80$	56
$80 \leq s < 90$	11
Total	100

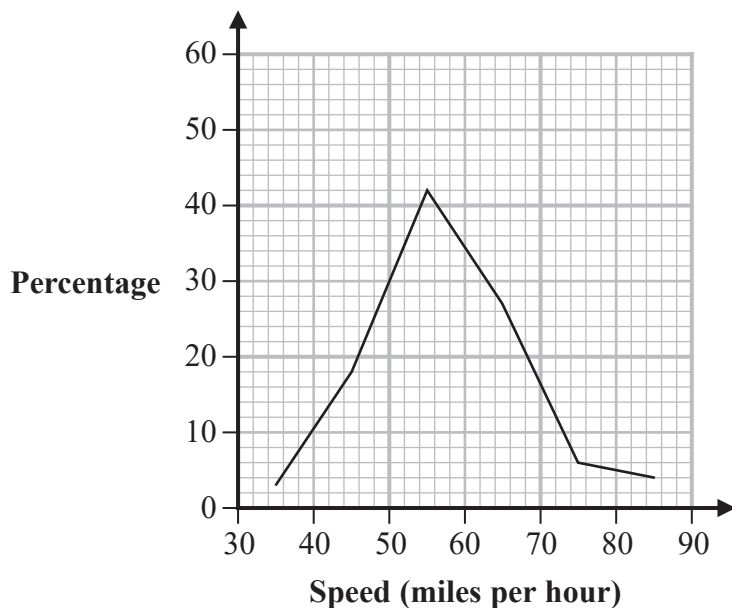


(c) Use linear interpolation to work out an estimate of the median speed of the motorcycles.

..... miles per hour
(3)

The frequency polygon for the speeds of cars has been drawn on the grid.

(d) On the same grid, draw the frequency polygon for the speeds of the motorcycles.
(2)



(e) Using the two frequency polygons, compare the skew of the distribution of the speeds of the cars with the skew of the distribution of the speeds of the motorcycles. Explain what your comparison means in context.

.....

.....

.....

.....

.....

(2)

(Total for Question 3 is 9 marks)



4 A supermarket chain has 93 stores in the United Kingdom with a total of 4502 employees. Each store is either a large store or a small store.

The directors of the supermarket chain are proposing to make changes to the employees' working hours. The directors want to survey a sample of 450 employees from the total workforce in order to ask them for their views on the proposed changes.

Three possible sampling methods in order to select the employees to be in the survey are suggested.

Method A The directors will choose two employees from each small store and six employees from each large store.

Method B All 4502 employees are allocated a number from 0001 to 4502. Start with number 0010 and take every 10th employee so that the sample consists of employees with the numbers 0010, 0020, 0030, ... and 4500.

Method C Each employee in each store will be given a raffle ticket and 10% of the number of employees in each store will be chosen at random by taking raffle tickets out of boxes, one box for each store.

For each of the three methods, identify the sampling method and discuss whether the sampling method is an appropriate way to select the employees to be in the survey.

As part of your discussion you should also state, with reasons, which of the three sampling methods is the most appropriate method for the directors to use.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 4 is 6 marks)



5 David is investigating the prices of houses in Streetly, which is in the West Midlands.

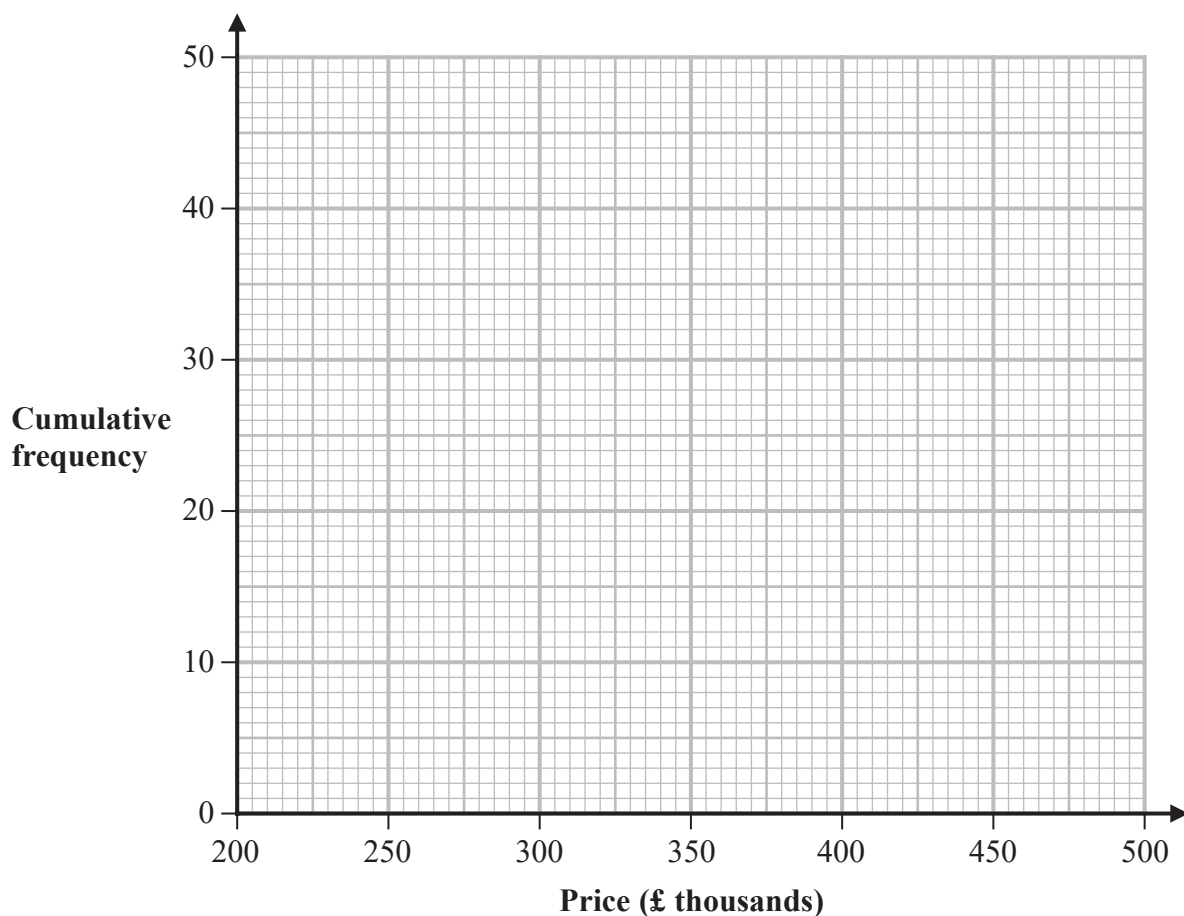
David took a sample of 41 houses in Streetly from a property website and recorded the price of each of these houses.

The cumulative frequency table shows information about the prices of these houses.

Price (£ P thousand)	Cumulative frequency
$240 < P \leq 280$	7
$240 < P \leq 320$	15
$240 < P \leq 360$	25
$240 < P \leq 400$	34
$240 < P \leq 440$	38
$240 < P \leq 480$	41

(Source: Adapted from *rightmove.co.uk*)

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



(2)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

In David's sample there were N houses with a price in the interval $300 < P \leq 350$

(b) Find an estimate for the value of N .

.....
(2)

David wants to use his investigation to predict the number of houses in Central London that have a price between £300 000 and £350 000

(c) Assess whether or not it is sensible for David to make this prediction using the results of his investigation.

.....
.....
.....
.....
(2)

(Total for Question 5 is 6 marks)



- 6 Gaby is researching the final position of a football team in the English Premier League and the mean value of all the players in the team.
- (a) Suggest a diagram that Gaby could draw to see if there is an association between the final position and the mean value.

(1)

The table gives information about the data for the 2018–2019 season that Gaby used.

Team	Mean value of all players in the team (£ million)		Final position of the team	d	d^2
	Value	Rank			
Arsenal FC	14.13	6	5	1	1
AFC Bournemouth	4.67	15	14	1	1
Brighton	4.21	17	17	0	0
Burnley FC	5.63	13	15	-2	4
Cardiff City	2.17	20	18	2	4
Chelsea FC	19.38	5	3	2	4
Crystal Palace	7.29	10	12	-2	4
Everton FC	10.74	7	8	-1	1
Fulham FC	6.70	12	19	-7	49
Huddersfield Town	3.43	19	20	-1	1
Leicester City	8.66	8	9	-1	1
Liverpool FC	23.08	3	2	1	1
Manchester City	22.92	4	1	3	9
Manchester United	23.80	2	6	-4	16
Newcastle United	5.41	14	13		
Southampton FC	6.76	11	16		
Tottenham Hotspur	24.11	1	4		
Watford FC	4.37	16	11		
West Ham United	8.37	9	10		
Wolverhampton	4.01	18	7		

(Source: www.transfermarket.com)



(b) Calculate Spearman's rank correlation coefficient for the information in the table.
You may complete the d column and d^2 column of the table to help you.

(3)

(c) Interpret your answer to part (b) in the context of Gaby's research.
You should refer to the effects of any anomalous data.

.....

.....

.....

.....

.....

(2)

Amelia suggests that Pearson's product moment correlation coefficient should be used instead of Spearman's rank correlation coefficient to measure the correlation between the data Gaby is researching.

(d) Discuss whether or not Amelia's suggestion is appropriate.

.....

.....

.....

.....

.....

.....

.....

(3)

(Total for Question 6 is 9 marks)

.....

.....

.....

.....

.....



- 7 The table gives information about the numbers, in thousands, of overseas visitors to the United Kingdom from North America.

Year	Quarter	Number of visitors (thousands)	4-point moving average (thousands)
2015	1	606	
	2	1164	
	3	1397	993.75
	4	808	1005
2016	1	651	998.25
	2	1137	1034.75
	3	1543	1070.75
	4	952	1102.25
2017	1	777	1198.25
	2	1521	1188.25
	3	1503	1186.5
	4	945	1186
2018	1	775	1146.5
	2	1363	1179.25
	3	1634	1181.75
	4	955	

(Source: Office for National Statistics)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



The table also gives the 4-point moving averages for the data in the table.

- (a) Explain why it is appropriate to calculate 4-point moving averages for the data in the table.

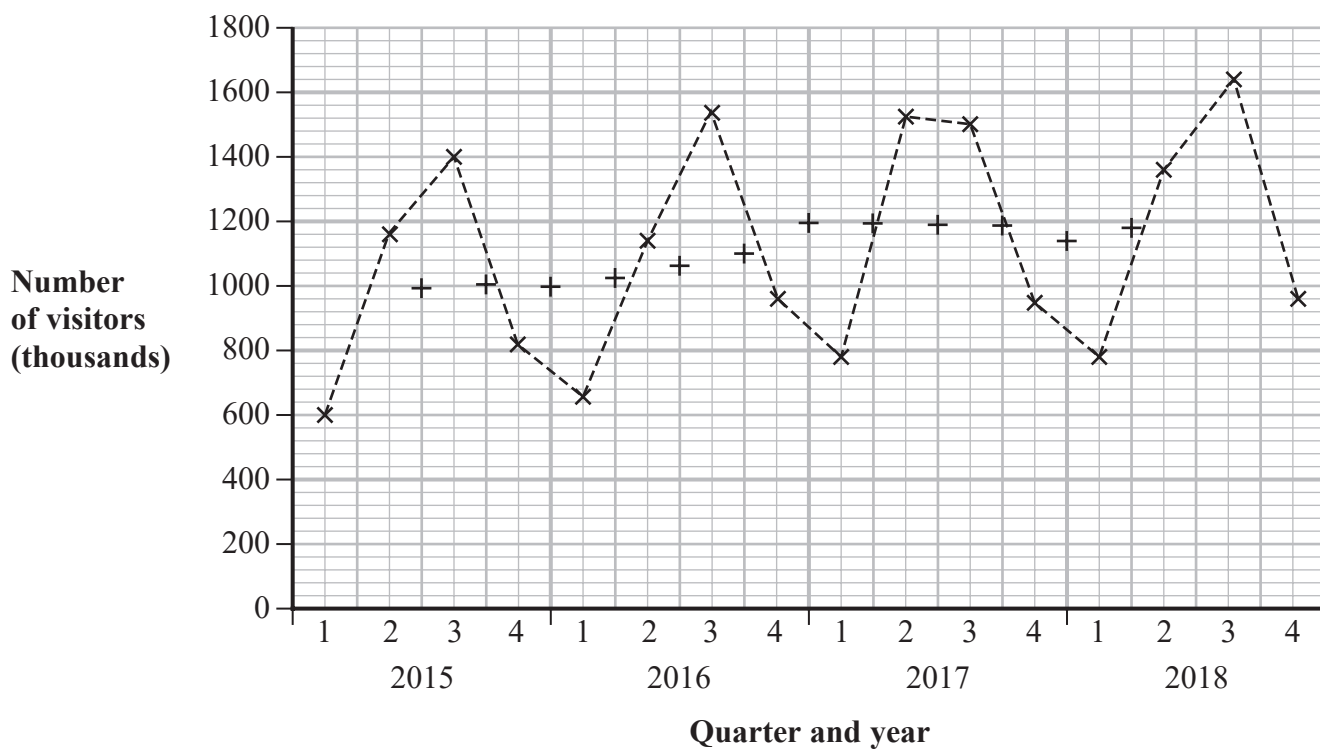
(1)

The time series graph has been plotted on the grid below for the data in the table.

Twelve of the 4-point moving averages for the data in the table have also been plotted on the grid so that the final 4-point moving average is missing from the grid.

- (b) Plot this 4-point moving average on the grid.

(1)



- (c) Draw a trend line for the time series graph.

(1)

Jo wants to predict the number of visitors to the United Kingdom from North America for the first quarter of 2019

She is going to use the time series graph, the trend line and an average seasonal effect.

(d) Describe how Jo should use these to obtain her prediction.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(3)

(Total for Question 7 is 6 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



- 8 The table gives the chain base index number for the average price of houses in the United Kingdom in January for each of the years 2016 to 2019

Year	January 2015	January 2016	January 2017	January 2018	January 2019
Chain base index numbers		107.76	104.76	104.32	101.73

(Source: *landregistry.data.gov.uk*)

- (a) Describe what the chain base index numbers show about the average price of houses in January for the years 2016 to 2019

(2)

In January 2017, the average price of houses in the United Kingdom was £215 243

- (b) Calculate the average price of houses in the United Kingdom in January 2016
Give your answer correct to the nearest £

£

(2)

(Total for Question 8 is 4 marks)



9 There are 8 coloured discs in bag X and 9 coloured discs in bag Y.

In bag X, there are 3 red discs and 5 yellow discs.

In bag Y, there are 5 red discs and 4 yellow discs.

Sonia takes at random one disc from bag X and she takes at random one disc from bag Y.

Calculate the probability that the two discs taken by Sonia are the same colour.

.....
(Total for Question 9 is 3 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



- 10 Huan has applied for a job at a college. As part of the interview process all candidates must take a numeracy test and a literacy test.

The table below gives Huan's score for the numeracy test and his score for the literacy test. The table also gives the mean and the standard deviation of the scores for each of these two tests for all candidates who were interviewed for the job.

Test	Huan's score	Test mean	Test standard deviation
Numeracy	49	42	3.5
Literacy	50	40	7.5

Huan concludes that because he scored a higher mark in literacy than in numeracy he performed better in literacy than in numeracy as compared to the other candidates.

Use statistical calculations to assess Huan's conclusion.

(Total for Question 10 is 5 marks)



- 11 Viktoria collects information about the times taken, in seconds, by 50 boys to run 400 metres. She records her results in the grouped frequency table, shown below, and gives the grouped frequency table to her friend Emeka.

Time taken (t seconds)	Frequency (f)
$50 < t \leq 60$	1
$60 < t \leq 65$	5
$65 < t \leq 70$	15
$70 < t \leq 75$	17
$75 < t \leq 80$	12

Viktoria asks Emeka to use only the information in the grouped frequency table to work out an estimate of the standard deviation of the times taken.

- (a) Explain why Emeka is only able to work out an **estimate** of the standard deviation.

(1)



Emeka uses the information in the table to work out the following estimates of the summary statistics

$$\begin{array}{lcl} \sum ft = 3542.5 & \text{and} & \sum ft^2 = 252\,331.25 \\ \text{Mean} = 70.85 & \text{and} & \text{Median} = 71.18 \end{array}$$

- (b) Using Emeka's estimates of the summary statistics, calculate an estimate of the skew of the distribution of the times taken by the 50 boys.
Give your answer to 2 decimal places.

.....
(3)

- (c) Interpret your answer to part (b) in the context of the data in the table.

.....
.....
.....
(1)

(Total for Question 11 is 5 marks)



12 A game is played with 4 ordinary 6-faced dice.
Each dice is to be rolled once and the number of dice that land on a six is recorded.

- (a) Write down **two** conditions needed so that a binomial distribution is a suitable model for the number of sixes recorded.

.....
.....
.....
.....
(2)

- (b) Calculate the probability that all of the 4 dice land on a six.
Give your answer as a fraction.

.....
(2)

- (c) Calculate the probability that at least 2 of the 4 dice land on a six.
Give your answer as a fraction.

.....
(3)

(Total for Question 12 is 7 marks)



- 13 Suzie is investigating the profits made by two different shops, shop **A** and shop **B**. Suzie has obtained the annual percentage profits made by shop **A** for the years 2015 to 2019 and the annual percentage profits made by shop **B** for the years 2016 to 2019

The table below gives this information.

Year	Percentage profit (%)	
	Shop A	Shop B
2015	1.7	
2016	1.8	0.9
2017	2.3	1.4
2018	2.9	1.1
2019	3.5	5.7

(Source: adapted from www.tescopl.com/investors/reports)

Suzie concludes that the average annual percentage profit made by shop **B** over the 4 years is greater than the average annual percentage profit made by shop **A** over the 5 years.

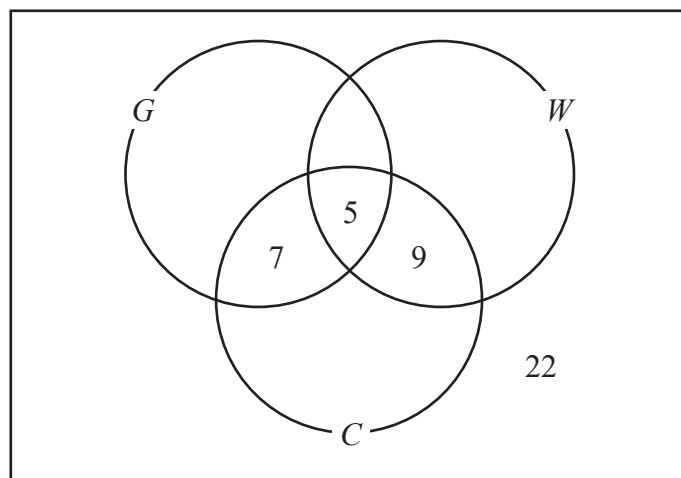
By using appropriate geometric means, assess Suzie's conclusion.
You must show your working.

(Total for Question 13 is 5 marks)



- 14 Olga is investigating the types of exercise taken by a group of 120 people. Olga asked each of these 120 people whether or not, for exercise, they go to a gym (G) or walk (W) or cycle (C).

The incomplete Venn diagram gives some information about Olga's results.



One of the people that Olga asked is to be chosen at random.

Given that, for this person,

$$P(W | G) = \frac{9}{52} \quad P(W \text{ or } G) = \frac{41}{60}$$

complete the Venn diagram.

(Total for Question 14 is 4 marks)

TOTAL FOR PAPER IS 80 MARKS

