

Please check the examination details below before entering your candidate information

Candidate surname		Other names	
Centre Number	Candidate Number		
<input type="text"/>	<input type="text"/>		

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

**Monday 12 June 2023**

Afternoon (Time: 1 hour 30 minutes)	Paper reference	<b>1ST0/1H</b>
-------------------------------------	-----------------	----------------

**Statistics**

**Paper 1**

**Higher Tier**

<p><b>You must have:</b></p> <p>Ruler graduated in centimetres and millimetres, protractor, pair of compasses, pen, HB pencil, eraser, scientific calculator.</p>	<p>Total Marks</p>
---	--------------------

### 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 ►

R72894RA

©2023 Pearson Education Ltd.  
N:1/1/1/1/

### 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

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



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Blank page area for writing.

BLANK PAGE



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

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

- 1 The scores for Vitality Roses netball team for the 2017 season are listed below.

55 61 72 62 52 74 56 49 75 66  
64 85 55 52 55 94 61 46 74 54

(Source: [www.englandnetball.co.uk](http://www.englandnetball.co.uk))

- (a) Use the data to complete the back-to-back stem and leaf diagram.

2017 season		2018 season
	1	9
	2	5 6 9
	3	5 7 9
	4	3 3 5 5 5 7
	5	0 2 2 4 6 8 8
	6	4 5 6
	7	0 2
	8	
	9	0

Key:

(3)

- (b) Work out the interquartile range of the scores for the 2018 season.

(2)



The table gives information about the scores for the 2017 and 2018 seasons.

	Median	Interquartile range
2017 season	61	18.5
2018 season	48.5	

Naomi thinks that the results show that Vitality Roses had a higher and more consistent set of scores in the 2018 season than in the 2017 season.

- (c) (i) Do the statistics support this conclusion?  
You must give reasons for your answer.

(3)

- (ii) Give a limitation of using the data provided to comment on whether or not Vitality Roses has improved as a team between the 2017 and 2018 seasons.

(1)

Naomi plans to process the data further.

She suggests checking for outliers, by performing calculations, before doing any more processing.

- (d) Comment on the appropriateness of this suggestion.

(1)

(Total for Question 1 is 10 marks)

2 Kyle is investigating the heights and the weights of professional basketball players.

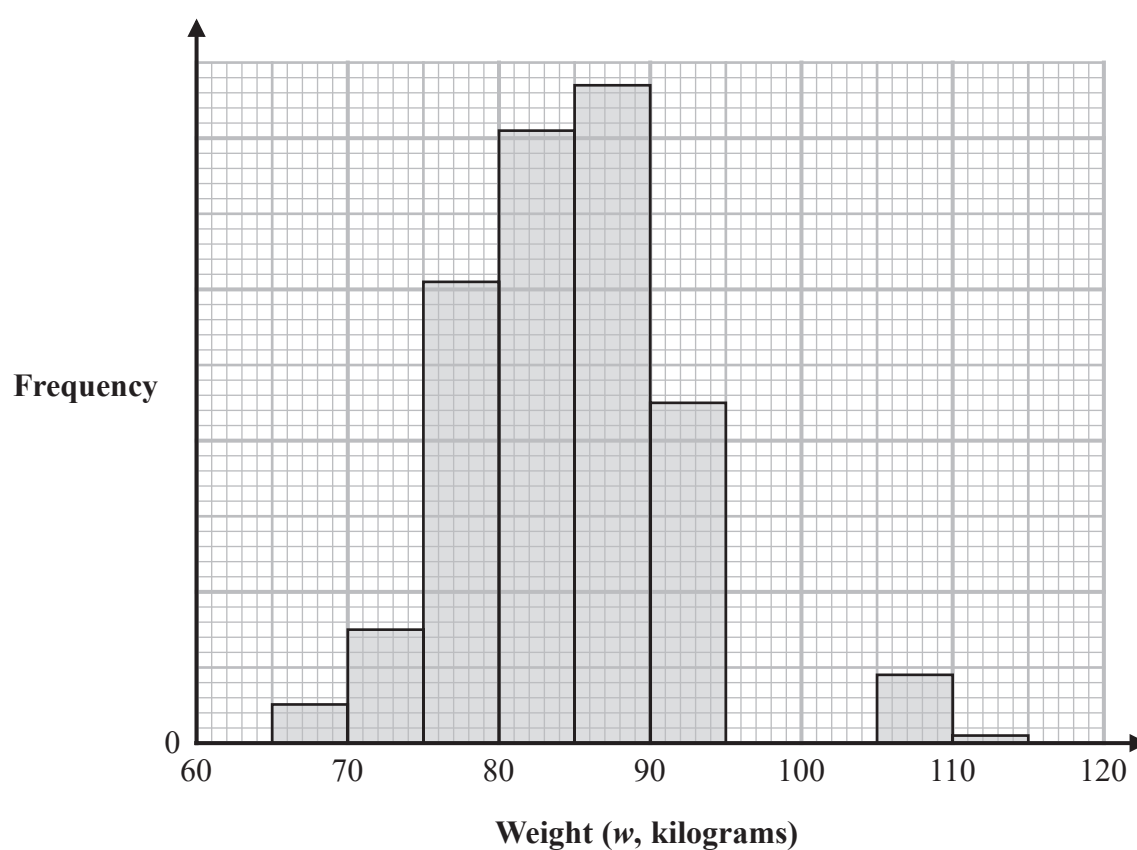
He found the weight, in kilograms, of some professional basketball players from 1950 to 1959

(a) Circle the word in the list below that describes weight, in kilograms, as a type of data.

discrete      continuous      ordinal      categorical

(1)

The incomplete histogram and incomplete grouped frequency table give information about the weights, in kilograms, of the professional basketball players from 1950 to 1959



Weight ( $w$ kilograms)	Frequency
$65 < w \leq 70$	5
$70 < w \leq 75$	15
$75 < w \leq 80$	61
$80 < w \leq 85$	81
$85 < w \leq 90$	
$90 < w \leq 95$	
$95 < w \leq 100$	35
$100 < w \leq 105$	14
$105 < w \leq 110$	9
$110 < w \leq 115$	1

(b) Use the information in the histogram to complete the table.

(2)

(c) Use the information in the table to complete the histogram.

(2)

Kyle also drew a histogram for the weights of professional basketball players from 2000 to 2009

This histogram was negatively skewed.

(d) Interpret the negative skew of the weights of professional basketball players from 2000 to 2009

(1)

Kyle also collected data about the heights of professional basketball players from 1950 to 1959 and the heights of professional basketball players from 2000 to 2009

The grouped frequency table below gives information about the heights of professional basketball players from 2000 to 2009

Height ( $h$ centimetres)	Frequency		
$170 < h \leq 180$	12		
$180 < h \leq 190$	146		
$190 < h \leq 200$	175		
$200 < h \leq 210$	323		
$210 < h \leq 220$	146		
$220 < h \leq 230$	8		
<b>Total</b>	<b>810</b>		

(Source: [www.kaggle.com](http://www.kaggle.com))

The estimate of the mean height for professional basketball players from 1950 to 1959 is calculated to be 190.9 cm to one decimal place.

- (e) (i) Calculate an estimate of the mean height of basketball players from 2000 to 2009

..... cm  
(3)

- (ii) Comment on how the mean height of professional basketball players has changed between the two sets of data.

(1)

(Total for Question 2 is 10 marks)



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**  
**Turn over for next question**

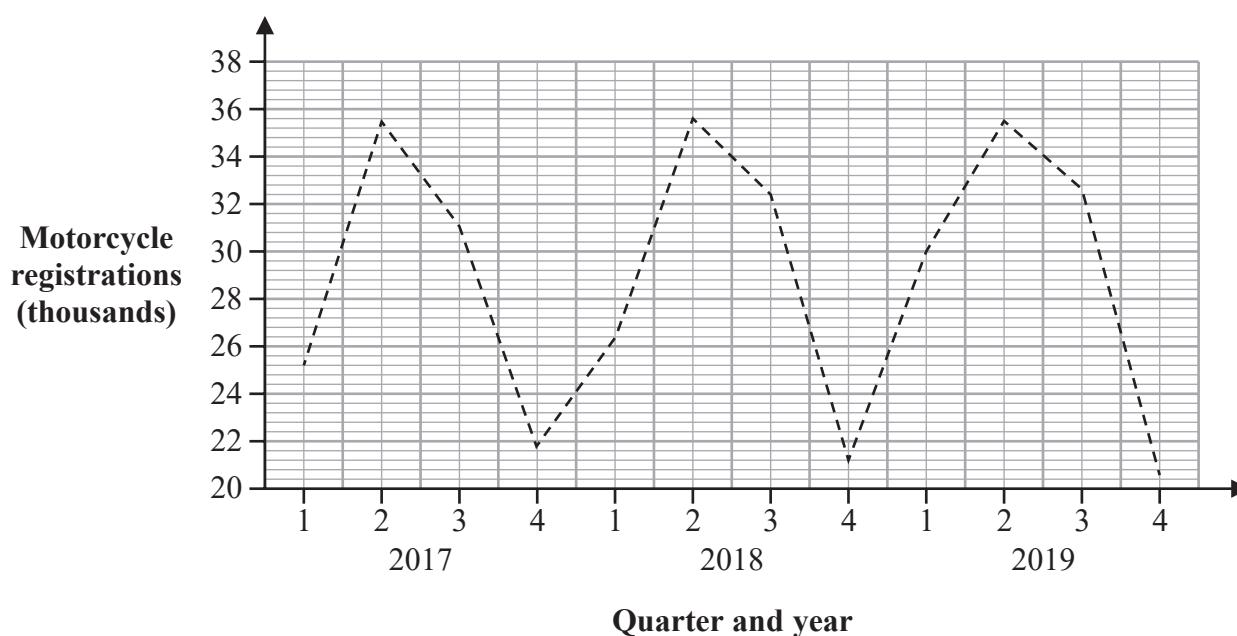


- 3 Claire is investigating sales of different types of vehicle over time. She plans to collect data on the numbers of motorcycles first registered in the UK over time.

(a) Write down a suitable hypothesis for this investigation.

(1)

The time series graph shows some information about the numbers of motorcycles first registered in the UK from 2017 to 2019



(Source: [www.gov.uk](http://www.gov.uk))

(b) Identify and interpret one example of seasonal trend shown by the time series graph.

(2)

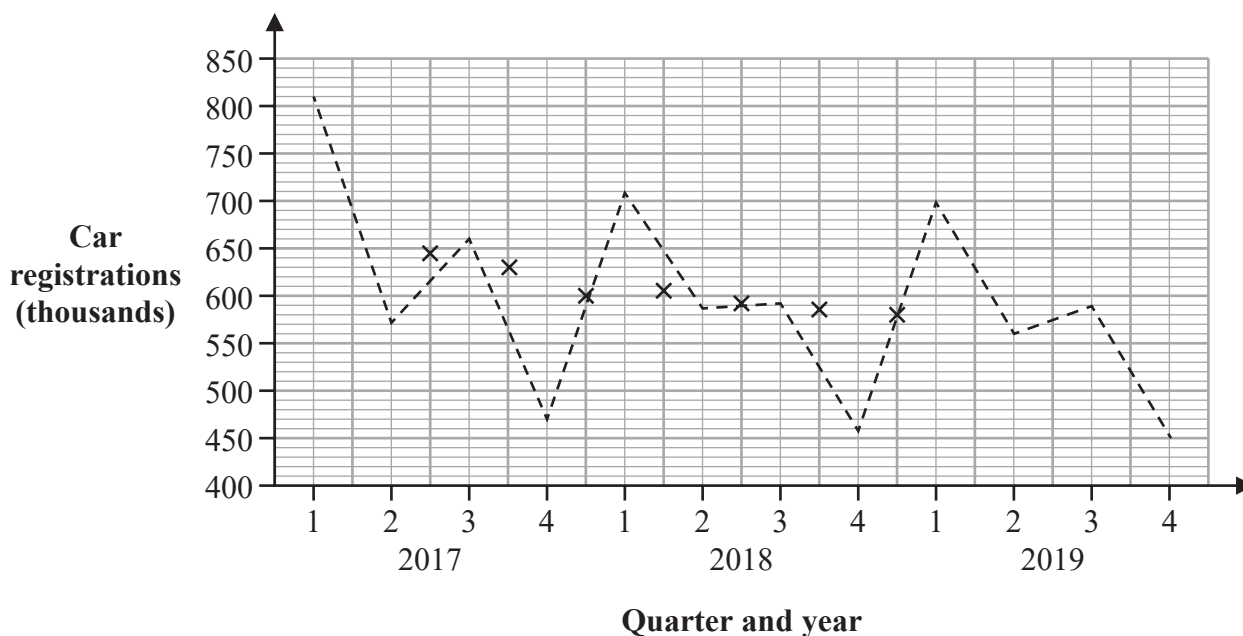
Claire calculated 4-point moving averages for the information shown in the time series graph.

(c) Explain why this is appropriate.

(1)

Claire also collected data on the numbers of cars first registered in the UK from 2017 to 2019

The time series graph shows some information about the numbers of cars first registered in the UK from 2017 to 2019 together with the first seven 4-point moving averages.



(Source: [www.gov.uk](http://www.gov.uk))

- (d) Compare the seasonal trend shown for the numbers of motorcycles first registered in the UK with the seasonal trend for the numbers of cars first registered in the UK.

(1)

The last three 4-point moving averages (thousands) for the number of cars registered in the UK from 2017 to 2019 are

576.0    575.3    573.9

- (e) Plot these three moving averages on the time series graph and draw a trend line.

(3)

- (f) Describe and interpret the trend in the numbers of cars first registered in the UK from 2017 to 2019

(2)

(Total for Question 3 is 10 marks)

- 4 Maciek is investigating the performance of athletes in the long jump. He wants to find out how the height of an athlete affects the distance jumped.

Here is how Maciek plans to collect, process and present his data.

**Collecting data**

- Ask all 50 of the students in his school athletics club to take part in the investigation.
- Ask each student to complete a long jump and measure their distance jumped.
- Ask each student to record their gender, age, height and distance jumped on his spreadsheet.

**Processing and presenting data**

- Produce two scatter diagrams – one for male students and one for female students.
- Plot the scatter diagrams with distance jumped on the horizontal axis and height of the student on the vertical axis.
- Look for outliers on the scatter diagram.
- Insert a line of best fit through the double mean point on each scatter diagram.
- Use the line of best fit to predict the results for the athletes at the Olympic Games.

Discuss whether Maciek's plans for collecting, processing and presenting data are appropriate.

(Total for Question 4 is 5 marks)



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**  
Turn over for next question



- 5 A local arts festival has a needlecraft competition and a flower arranging competition. Mr Smith judges both competitions, giving each entry a rank.

Mr Smith thinks there is an association between the ranks given to competitors in the needlecraft competition and the ranks given to competitors in the flower arranging competition.

He plans to calculate Spearman's rank correlation coefficient for the ranks in the two competitions.

- (a) Explain why this may **not** be appropriate.

(1)

The arts festival also has a photography competition.

In the photography competition, the 8 entries were given the following ranks by two judges, Mrs John and Mr Nowak.

Photography entry	Mrs John's rank	Mr Nowak's rank		
A	3	1		
B	6	6		
C	2	5		
D	4	3		
E	7	8		
F	1	4		
G	8	7		
H	5	2		



- (b) Using suitable calculations, investigate the level of agreement between the two judges.  
You may use the blank columns in the table for your working.

(5)

(Total for Question 5 is 6 marks)

- 6 The table gives information about the monthly average price per troy ounce, in pounds, of gold over a period of six months in 2019

The table also gives some of the chain base index numbers, correct to one decimal place, for this information.

	Jan	Feb	Mar	Apr	May	Jun
Monthly average price (£)	1001.4	1014.2	988.1	986.8	1000.0	1071.9
Chain base index number		101.3	97.4	99.9	101.3	

(Source: *gold.org*)

- (a) Calculate the chain base index number for June and write it in the table.  
Give your answer correct to one decimal place.

(2)

- (b) (i) Calculate the geometric mean of the five chain base index numbers.  
You must show your working.  
Give your answer correct to one decimal place.

(2)

- (ii) Interpret your answer to part (i) in context.

(2)

(Total for Question 6 is 6 marks)



- 7 A class is learning about the Petersen capture-recapture method.

Their teacher provides them with the following data for burbot fish in Tolsona Lake, Alaska.

The data was collected by tagging all of the fish in the initial sample before releasing them, then counting the number of fish with tags in the second sample.

Initial sample size	683
Second sample size	587
Number of tagged fish in second sample	148

(Source: [www.adfg.alaska.gov](http://www.adfg.alaska.gov))

- (a) Calculate an estimate of the total number of burbot fish in Tolsona Lake.

(2)

Sophie is planning her own investigation using the Petersen capture-recapture method. Here is her plan for data collection.

Data collection to estimate the population of frogs in the science garden.

Monday 9 am

Capture 5 frogs and tag them using stick-on tags.

Release them.

Friday 9 am

Capture 5 frogs and record how many have tags on.

- (b) Discuss the appropriateness of Sophie's plan for data collection.

(3)

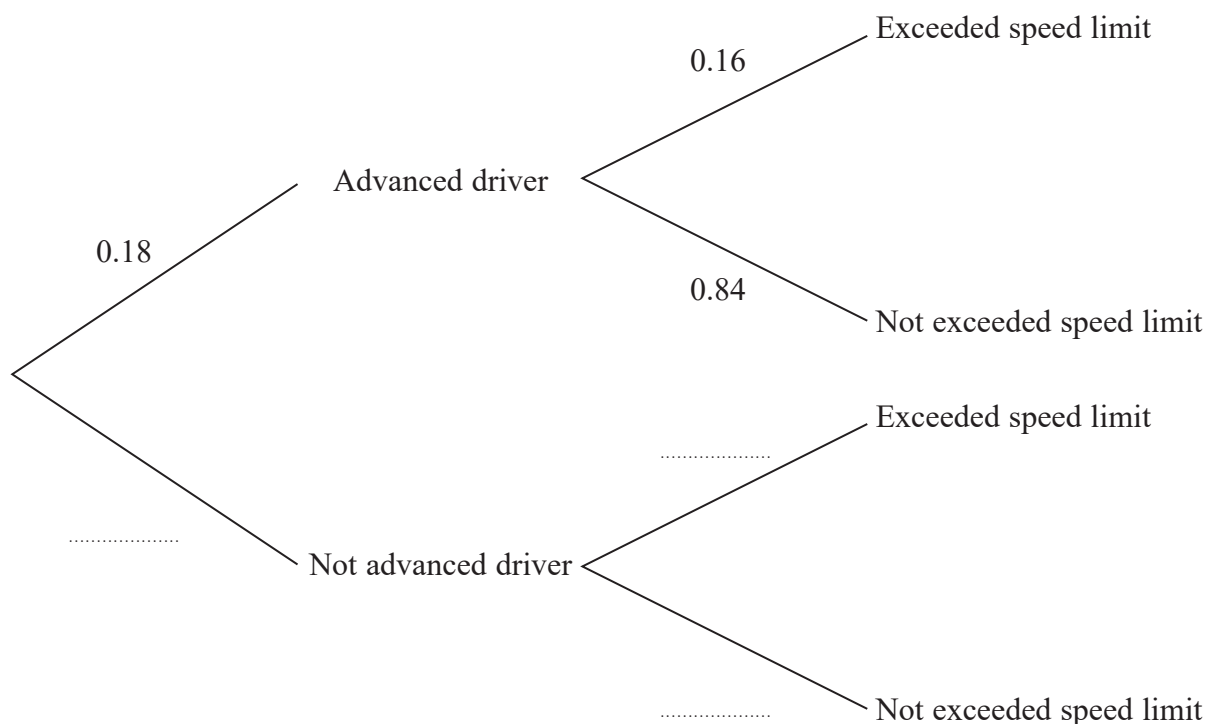
(Total for Question 7 is 5 marks)

- 8 Ian collects information about whether drivers have had advanced driver training and whether they have exceeded the speed limit in the last month.

He finds that 18% of the drivers he surveyed have had advanced driver training and that 16% of these drivers have exceeded the speed limit in the last month.

For the drivers who have not had advanced driver training 26% of them have exceeded the speed limit in the last month.

- (a) Complete the probability tree diagram for this information.



(1)

- (b) Calculate  $P(\text{Advanced driver} | \text{Not exceeded speed limit})$ .

(2)

- (c) (i) Calculate the relative risk of exceeding the speed limit for the not advanced drivers compared to the advanced drivers.

(2)

- (ii) Interpret this relative risk in context.

(1)

Ian researches the relative risk of exceeding the speed limit for newly qualified drivers compared to drivers who are not newly qualified.

He finds that in 2019 the relative risk was 0.95 and in 2020 the relative risk was 0.87

Ian concludes that the proportion of newly qualified drivers exceeding the speed limit has decreased between 2019 and 2020

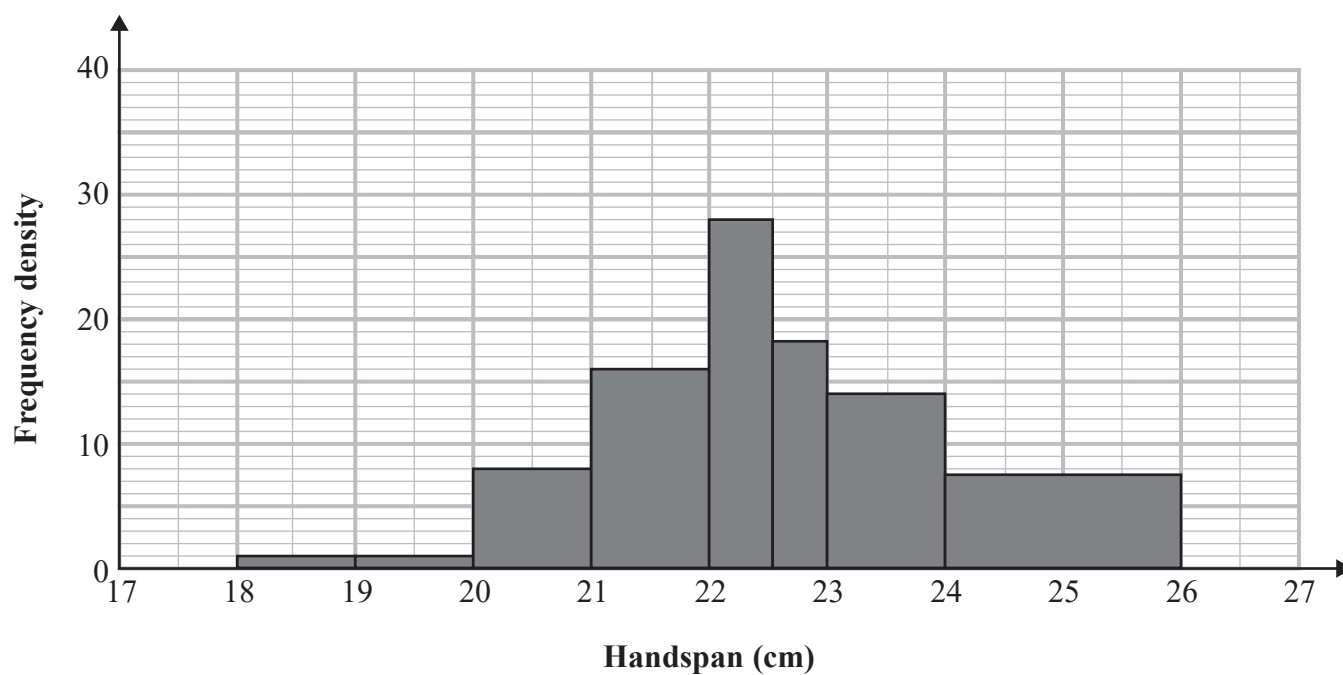
- (d) Do the statistics support Ian's conclusion?  
You must give a reason for your answer.

(2)

(Total for Question 8 is 8 marks)

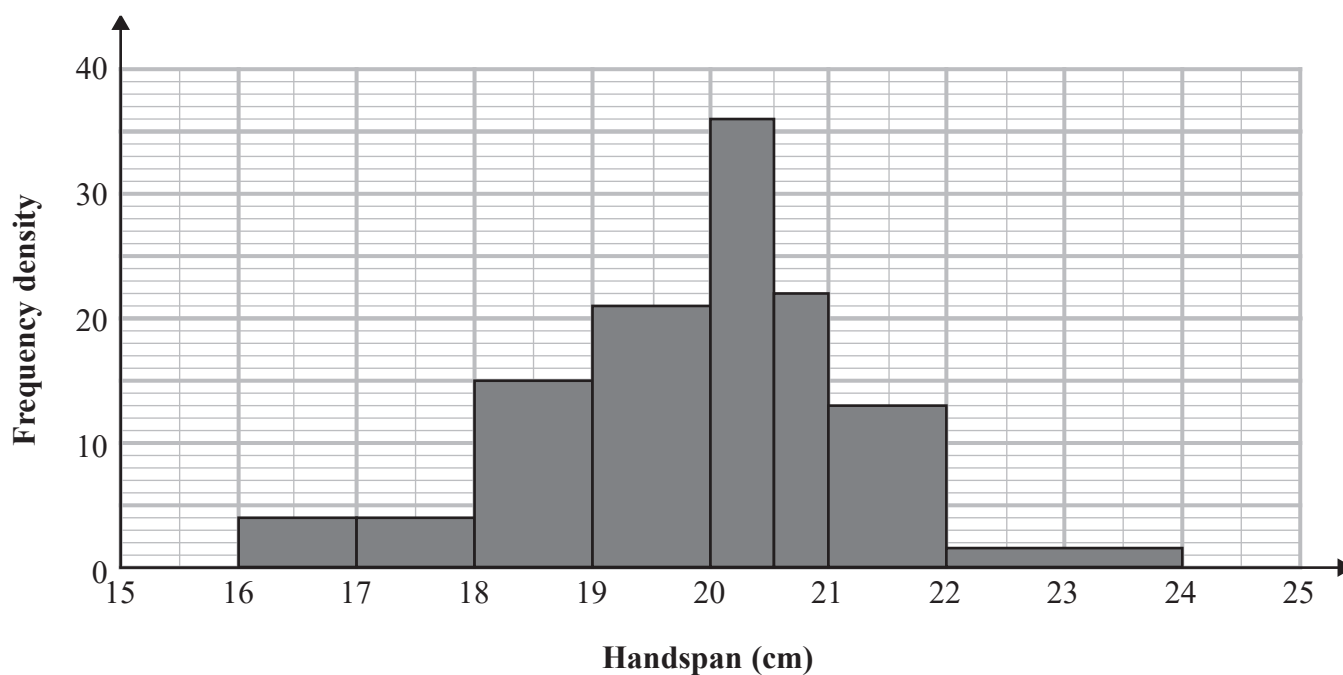
- 9 Ligita collected data on the handspans, in cm, of men and the handspans, in cm, of women. She processed her data using statistical software, here are her results.

#### Male handspans



Minimum	18.0	Mean	22.3	Standard deviation	1.44
Maximum	25.5	Median	22.0	Skewness	0.630
Number of male handspans		78			

#### Female handspans



Minimum	16.0	Mean	19.6	Standard deviation	1.33
Maximum	23.0	Median	20.0		
Number of female handspans		89			

(Source: [mathsci.solano.edu/](http://mathsci.solano.edu/))

(a) Calculate the skew of the female handspan data.

.....  
(2)

(b) Compare in context the distribution of male handspans and the distribution of female handspans.

(5)

Ligita plans to use the mean for the males and the mean for the females to find all of the handspans.

(c) Explain why she should use a weighted mean for this calculation.

(1)

(Total for Question 9 is 8 marks)

- 10 A study took place in the southeastern United States to find if there was a relationship between the snout-vent length, in cm, and the body weight, in grams, of adult snakes.

The researchers found the equations of the regression lines for the relationship between snout-vent length ( $x$  centimetres) and the body weight ( $y$  grams) for each species of snake.

For two species of snake, the timber rattlesnake and the eastern racer, there was a difference in the relationship between the two variables for female snakes and the relationship between the two variables for male snakes.

The table below gives the equations of the regression lines for these snakes.

	Male	Female
Timber rattlesnake	$y = -1804.14 + 26.03x$	$y = -1236.54 + 19.61x$
Eastern racer	$y = -363.69 + 6.36x$	$y = -236.80 + 4.62x$

(Source: [www.jstor.org](http://www.jstor.org))

- (a) Interpret in context the figure 19.61 in the regression equation for female timber rattlesnakes.

(1)

- (b) Compare the relationships between snout-vent length and body weight in the two species of snake. Include in your comparisons reference to male snakes and female snakes.

(3)

The researchers would like to use a normal distribution as a model for the weights of male timber rattlesnakes.

They plan to draw a histogram, calculate averages and the standard deviation for these weights to check whether a normal distribution is a suitable model.

- (c) (i) Explain how they could use a histogram to check whether a normal distribution is a suitable model.

(1)

- (ii) Explain how they could use averages and the standard deviation to check whether the normal distribution is a suitable model.

(2)

(Total for Question 10 is 7 marks)

- 11 The manager of a large delivery company wants to investigate whether employees have dealt with personal errands whilst at work in the last 6 months.  
He plans to ask employees whether they have dealt with personal errands during the work day.

Libby suggests using the random response technique.

- (a) Explain why it is appropriate to use the random response technique in this case.

(1)

The manager designs the random response question below.

Think of a number between 1 and 8

If you thought of an even number, tick box A.

If you thought of an odd number, answer this question.

Have you dealt with personal errands during your work day?

If yes, tick box A. If no, tick box B.

A

☐

B

☐

- (b) Comment on the appropriateness of this random response question.  
Give reasons for your answer.

(2)



The final questionnaire will be sent to a sample of the delivery company employees.  
The delivery company has office staff, warehouse staff and delivery drivers.  
The staff work either full time or part time.

The table gives the number of each type of employee.

		Employment role		
		Office staff	Warehouse staff	Delivery drivers
Contract type	Full time	126	874	610
	Part time	93	615	208

The manager wants to take a sample stratified by employment role and contract type.  
He wants to sample at least 15 people from each strata.  
If his calculated sample for one of the strata is a decimal he will round this to the nearest integer.

(c) Work out the smallest total sample size that gives at least 15 people from each strata.

(2)

(Total for Question 11 is 5 marks)

TOTAL FOR PAPER IS 80 MARKS



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





DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

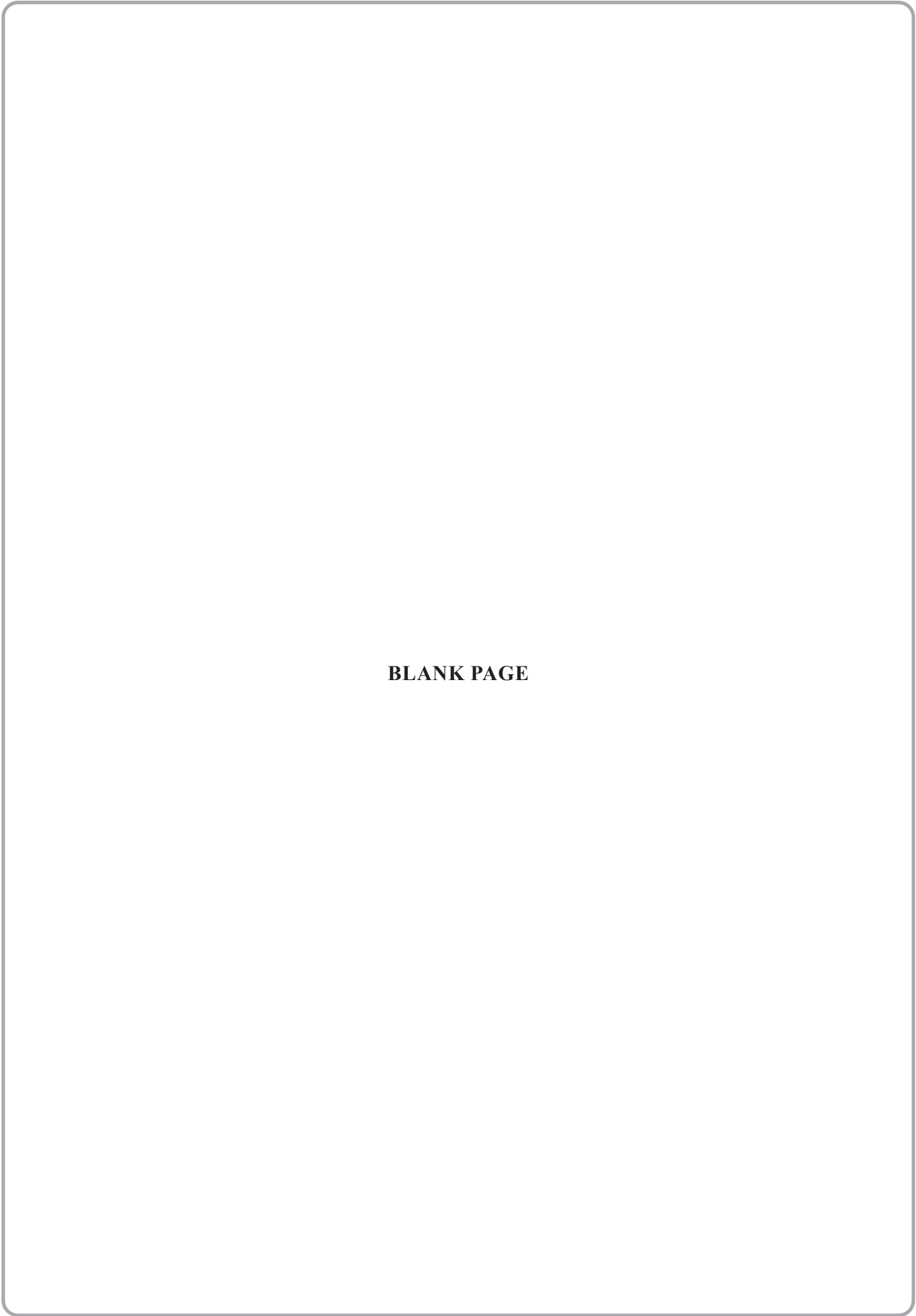
DO NOT WRITE IN THIS AREA



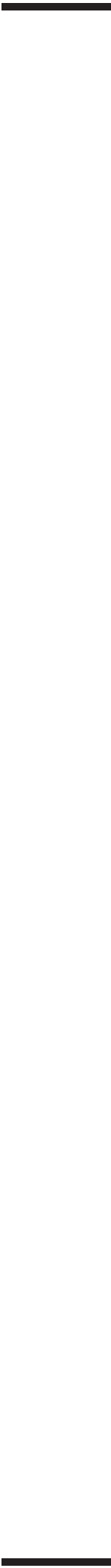
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



BLANK PAGE



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

