The big Stats worksheet

1. Every evening, 5 men and 5 women are chosen to take part in a phone-in competition. Of these 10 people, exactly 3 will win a prize. These 3 prize-winners are chosen at random.

i. Find the probability that, on a particular evening, 2 of the prize-winners are women and the other is a man. Give your answer as a fraction in its simplest form.

[4]

ii. Four evenings are selected at random. Find the probability that, on at least three of the four evenings, 2 of the prize-winners are women and the other is a man.

[4]

2. The weights of bags of a particular brand of flour are quoted as 1.5 kg. In fact, on average 10% of bags are underweight.

i. Find the probability that, in a random sample of 50 bags, there are exactly 5 bags which are underweight.

[3]

ii. Bags are randomly chosen and packed into boxes of 20. Find the probability that there is at least one underweight bag in a box.

[2]

iii. A crate contains 48 boxes. Find the expected number of boxes in the crate which contain at least one underweight bag.

[2]

3. A researcher is investigating whether people can identify whether a glass of water they are given is bottled water or tap water. She suspects that people do no better than they would by guessing. Twenty people are selected at random; thirteen make a correct identification. She carries out a hypothesis test.

i. Explain why the null hypothesis should be $p = 0.5$, where p represents the probability that a randomly selected person makes a correct identification.

[2]

ii. Briefly explain why she uses an alternative hypothesis of $p > 0.5$.

[1]

iii. Complete the test at the 5% significance level.

[5]

4. The birth weights in kilograms of 25 female babies are shown below, in ascending order.

1.39	2.50	2.68	2.76	2.82	2.82	2.84	3.03	3.06	3.16	3.16	3.24	3.32
3.36	3.40	3.54	3.56	3.56	3.70	3.72	3.72	3.84	4.02	4.24	4.34	

i. Find the median and interquartile range of these data.

[3]

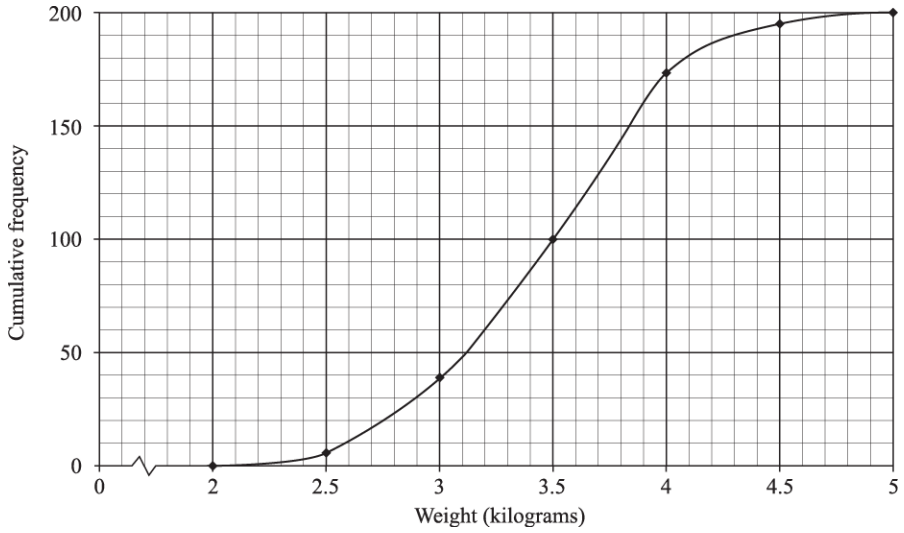
ii. Draw a box and whisker plot to illustrate the data.

[3]

iii. Show that there is exactly one outlier. Discuss whether this outlier should be removed from the data.

[4]

The cumulative frequency curve below illustrates the birth weights of 200 male babies.



iv. Find the median and interquartile range of the birth weights of the male babies.

[3]

v. Compare the weights of the female and male babies.

[2]

vi. Two of these male babies are chosen at random. Calculate an estimate of the probability that both of these babies weigh more than any of the female babies.

[3]

5. Jenny has six darts. She throws darts, one at a time, aiming each at the bull's-eye. The probability that she hits the bull's-eye with her first dart is 0.1. For any subsequent throw, the probability of hitting the bull's-eye is 0.2 if the previous dart hit the bull's-eye and 0.05 otherwise.

i. Illustrate the possible outcomes for her first, second and third darts on a probability tree diagram.

[4]

ii. Find the probability that

A. she hits the bull's-eye with at least one of her first three darts,

[3]

B. she hits the bull's-eye with exactly one of her first three darts.

[4]

iii. Given that she hits the bull's-eye with at least one of her first three darts, find the probability that she hits the bull's-eye with exactly one of them.

[3]

Jenny decides that, if she hits the bull's-eye with any of her first three darts, she will stop after throwing three darts. Otherwise she will throw all six darts.

iv. Find the probability that she hits the bull's-eye three times in total.

[4]

6. The scores, X, in Paper 1 of an English examination have an underlying Normal distribution with mean 76 and standard deviation 12. The scores are reported as integer marks. So, for example, a score for which $75.5 \leqslant X < 76.5$ is reported as 76 marks.

i. Find the probability that a candidate's reported mark is 76.

[4]

ii. Find the probability that a candidate's reported mark is at least 80.

[3]

iii. Three candidates are chosen at random. Find the probability that exactly one of these three candidates' reported marks is at least 80.

[2]

The proportion of candidates who receive an A* grade (the highest grade) must not exceed 10% but should be as close as possible to 10%.

iv. Find the lowest reported mark that should be awarded an A* grade.

[5]

The scores in Paper 2 of the examination have an underlying Normal distribution with mean μ and standard deviation 12.

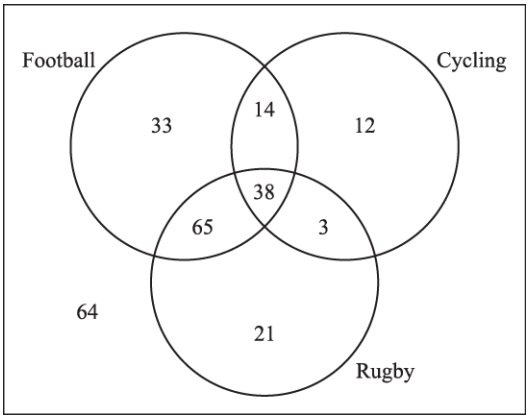
v. Given that 20% of candidates receive a reported mark of 50 or less, find the value of μ.

[4]

7. A survey is being carried out into the sports viewing habits of people in a particular area. As part of the survey, 250 people are asked which of the following sports they have watched on television in the past month.

- Football
- Cycling
- Rugby

The numbers of people who have watched these sports are shown in the Venn diagram.



One of the people is selected at random.

i. Find the probability that this person has in the past month

A. watched cycling but not football,

[1]

B. watched either one or two of the three sports.

[2]

ii. Given that this person has watched cycling, find the probability that this person has not watched football.

[2]

8. At a tourist information office the numbers of people seeking information each hour over the course of a 12-hour day are shown below.

6	25	38	39	31	18	35	31	33	15	21	28

i. Construct a sorted stem and leaf diagram to represent these data.

[3]

ii. State the type of skewness suggested by your stem and leaf diagram.

[1]

iii. For these data find the median, the mean and the mode. Comment on the usefulness of the mode in this case.

[4]

9. A rugby team of 15 people is to be selected from a squad of 25 players.

i. How many different teams are possible?

[2]

ii. In fact the team has to consist of 8 forwards and 7 backs. If 13 of the squad are forwards and the other 12 are backs, how many different teams are now possible?

[2]

iii. Find the probability that, if the team is selected at random from the squad of 25 players, it contains the correct numbers of forwards and backs.

[2]

10. A drug for treating a particular minor illness cures, on average, 78% of patients. Twenty people with this minor illness are selected at random and treated with the drug.

i.

A. Find the probability that exactly 19 patients are cured.

[3]

B. Find the probability that at most 18 patients are cured.

[3]

C. Find the expected number of patients who are cured.

[1]

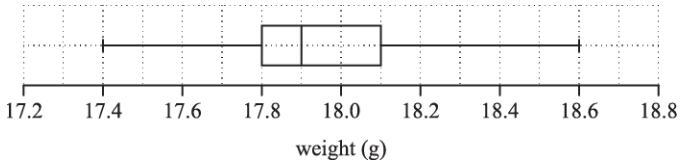
ii. A pharmaceutical company is trialling a new drug to treat this illness. Researchers at the company hope that a higher percentage of patients will be cured when given this new drug. Twenty patients are selected at random, and given the new drug. Of these, 19 are cured. Carry out a hypothesis test at the 1% significance level to investigate whether there is any evidence to suggest that the new drug is more effective than the old one.

[8]

iii. If the researchers had chosen to carry out the hypothesis test at the 5% significance level, what would the result have been? Justify your answer.

[2]

11. The box and whisker plot below summarises the weights in grams of the 20 chocolates in a box.



i. Find the interquartile range of the data and hence determine whether there are any outliers at either end of the distribution.

[5]

Ben buys a box of these chocolates each weekend. The chocolates all look the same on the outside, but 7 of them have orange centres, 6 have cherry centres, 4 have coffee centres and 3 have lemon centres.

One weekend, each of Ben's 3 children eats one of the chocolates, chosen at random.

ii. Calculate the probabilities of the following events.
   A. all 3 chocolates have orange centres
   B. all 3 chocolates have the same centres

[6]

iii. Find P($A|B$) and P($B|A$).

[3]

The following weekend, Ben buys an identical box of chocolates and again each of his 3 children eats one of the chocolates, chosen at random.

iv. Find the probability that, on both weekends, the 3 chocolates that they eat all have orange centres.

[2]

v. Ben likes all of the chocolates except those with cherry centres. On another weekend he is the first of his family to eat some of the chocolates. Find the probability that he has to select more than 2 chocolates before he finds one that he likes.

[3]

12. The ages, x years, of the senior members of a running club are summarised in the table below.

Age (x)	$20 \leqslant x < 30$	$30 \leqslant x < 40$	$40 \leqslant x < 50$	$50 \leqslant x < 60$	$60 \leqslant x < 70$	$70 \leqslant x < 80$
Frequency	10	30	42	23	9	5

i. Draw a cumulative frequency diagram to illustrate the data.

[5]

ii. Use your diagram to estimate the median and interquartile range of the data.

[3]

13. Each weekday, Marta travels to school by bus. Sometimes she arrives late.

- L is the event that Marta arrives late.
- R is the event that it is raining.

You are given that $P(L) = 0.15$, $P(R) = 0.22$ and $P(L \mid R) = 0.45$.

i. Use this information to show that the events L and R are not independent.

[1]

ii. Find $P(L \cap R)$.

[2]

iii. Draw a Venn diagram showing the events L and R, and fill in the probability corresponding to each of the four regions of your diagram.

[3]

14. The weights, w grams, of a random sample of 60 carrots of variety A are summarised in the table below.

Weight	$30 \leqslant w < 50$	$50 \leqslant w < 60$	$60 \leqslant w < 70$	$70 \leqslant w < 80$	$80 \leqslant w < 90$
Frequency	11	10	18	14	7

i. Draw a histogram to illustrate these data.

[5]

ii. Calculate estimates of the mean and standard deviation of w.

[4]

iii. Use your answers to part **(ii)** to investigate whether there are any outliers.

[3]

The weights, x grams, of a random sample of 50 carrots of variety B are summarised as follows.

$n = 50$ $\sum x = 3624.5$ $\sum x^2 = 265\ 416$

iv. Calculate the mean and standard deviation of x.

[3]

v. Compare the central tendency and variation of the weights of varieties A and B.

[2]

15. It is known that on average 85% of seeds of a particular variety of tomato will germinate. Ramesh selects 15 of these seeds at random and sows them.

i.

A. Find the probability that exactly 12 germinate.

[3]

B. Find the probability that fewer than 12 germinate.

[2]

The following year Ramesh finds that he still has many seeds left. Because the seeds are now one year old, he suspects that the germination rate will be lower. He conducts a trial by randomly selecting n of these seeds and sowing them. He then carries out a hypothesis test at the 1% significance level to investigate whether he is correct.

ii. Write down suitable null and alternative hypotheses for the test. Give a reason for your choice of alternative hypothesis.

[4]

iii. In a trial with $n = 20$, Ramesh finds that 13 seeds germinate. Carry out the test.

[4]

iv. Suppose instead that Ramesh conducts the trial with $n = 50$, and finds that 33 seeds germinate. Given that the critical value for the test in this case is 35, complete the test.

[3]

v. If n is small, there is no point in carrying out the test at the 1% significance level, as the null hypothesis cannot be rejected however many seeds germinate. Find the least value of n for which the null hypothesis can be rejected, quoting appropriate probabilities to justify your answer.

[3]

16. Each weekday Alan drives to work. On his journey, he goes over a level crossing. Sometimes he has to wait at the level crossing for a train to pass.

- W is the event that Alan has to wait at the level crossing.
- L is the event that Alan is late for work.

You are given that $P(L|W) = 0.4$, $P(W) = 0.07$ and $P(L \cup W) = 0.08$.

i. Calculate $P(L \cap W)$.

[2]

ii. Draw a Venn diagram, showing the events L and W. Fill in the probability corresponding to each of the four regions of your diagram.

[3]

iii. Determine whether the events L and W are independent, explaining your method clearly.

[3]

17. The stem and leaf diagram illustrates the heights in metres of 25 young oak trees.

3	4 6 7 8 9 9
4	0 2 2 3 4 6 8 9
5	0 1 3 5 8
6	2 4 5
7	4 6
8	1

Key: 4|2 represents 4.2

i. State the type of skewness of the distribution.

[1]

ii. Use your calculator to find the mean and standard deviation of these data.

[3]

iii. Determine whether there are any outliers.

[4]

18. The heights x cm of 100 boys in Year 7 at a school are summarised in the table below.

Height	$125 \leqslant x \leqslant 140$	$140 < x \leqslant 145$	$145 < x \leqslant 150$	$150 < x \leqslant 160$	$160 < x \leqslant 170$
Frequency	25	29	24	18	4

i. Estimate the number of boys who have heights of at least 155 cm.

[2]

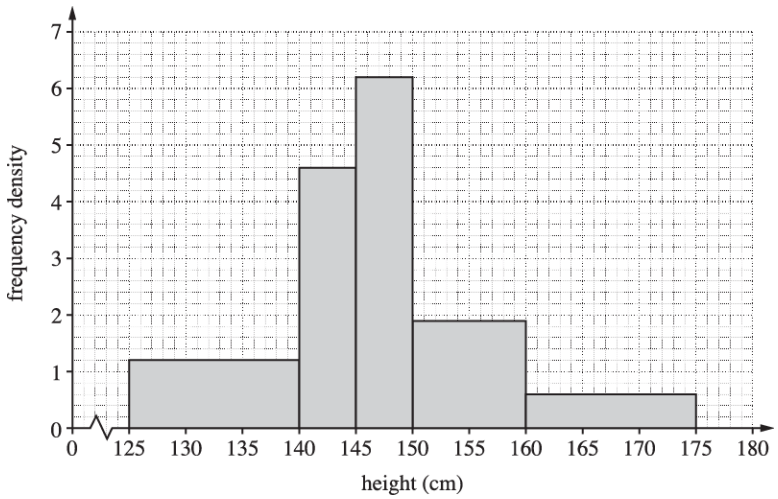
ii. Calculate an estimate of the median height of the 100 boys.

[3]

iii. Draw a histogram to illustrate the data.

[5]

The histogram below shows the heights of 100 girls in Year 7 at the same school.



iv. How many more girls than boys had heights exceeding 160 cm?

[3]

v. Calculate an estimate of the mean height of the 100 girls.

[5]

19. Malik is playing a game in which he has to throw a 6 on a fair six-sided die to start the game. Find the probability that

i. Malik throws a 6 for the first time on his third attempt,

[3]

ii. Malik needs at most ten attempts to throw a 6.

[2]

20. A coffee shop provides free internet access for its customers. It is known that the probability that a randomly selected customer is accessing the internet is 0.35, independently of all other customers.

i. 10 customers are selected at random.

A. Find the probability that exactly 5 of them are accessing the internet.

[3]

B. Find the probability that at least 5 of them are accessing the internet.

[2]

C. Find the expected number of these customers who are accessing the internet.

[2]

Another coffee shop also provides free internet access. It is suspected that the probability that a randomly selected customer at this coffee shop is accessing the internet may be different from 0.35. A random sample of 20 customers at this coffee shop is selected. Of these, 10 are accessing the internet.

ii. Carry out a hypothesis test at the 5% significance level to investigate whether the probability for this coffee shop is different from 0.35. Give a reason for your choice of alternative hypothesis.

[9]

iii. To get a more reliable result, a much larger random sample of 200 customers is selected over a period of time, and another hypothesis test is carried out. You are given that 90 of the 200 customers were accessing the internet. You are also given that, if X has the binomial distribution with parameters $n = 200$ and $p = 0.35$, then $P(X \geqslant 90) = 0.0022$. Using the same hypotheses and significance level which you used in part **(ii)**, complete this test.

[2]

21. The wing lengths of native English male blackbirds, measured in mm, are Normally distributed with mean 130.5 and variance 11.84.

i. Find the probability that a randomly selected native English male blackbird has a wing length greater than 135 mm.

[3]

ii. Given that 1% of native English male blackbirds have wing length more than k mm, find the value of k.

[3]

iii. Find the probability that a randomly selected native English male blackbird has a wing length which is 131 mm correct to the nearest millimetre.

[3]

It is suspected that Scandinavian male blackbirds have, on average, longer wings than native English male blackbirds. A random sample of 20 Scandinavian male blackbirds has mean wing length 132.4 mm. You may assume that wing lengths in this population are Normally distributed with variance 11.84 mm^2.

iv. Carry out an appropriate hypothesis test, at the 5% significance level.

[8]

v. Discuss briefly one advantage and one disadvantage of using a 10% significance level rather than a 5% significance level in hypothesis testing in general.

[2]

22. A geyser is a hot spring which erupts from time to time. For two geysers, the duration of each eruption, x minutes, and the waiting time until the next eruption, y minutes, are recorded.

(a) For a random sample of 50 eruptions of the first geyser, the correlation coefficient between x and y is 0.758. The critical value for a 2-tailed hypothesis test for correlation at the 5% level is 0.279. Explain whether or not there is evidence of correlation in the population of eruptions. **[2]**

The scatter diagram in Fig. 9 shows the data from a random sample of 50 eruptions of the second geyser.

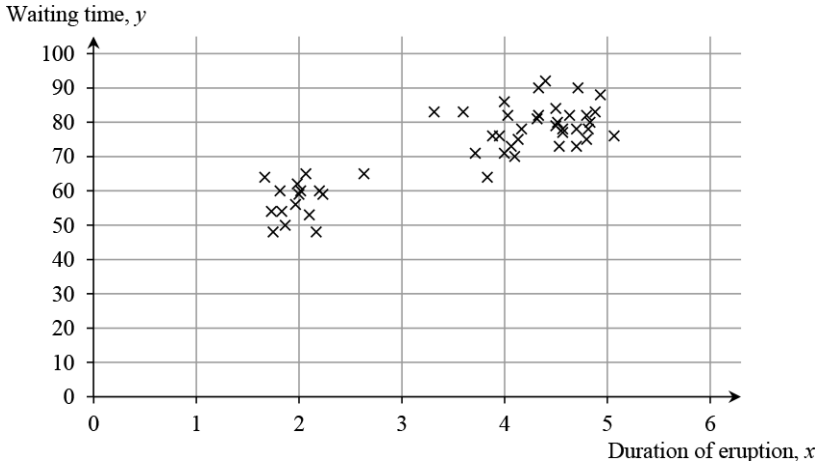


Fig.9

(b) Stella claims the scatter diagram shows evidence of correlation between duration of eruption and waiting time. Make two comments about Stella's claim. **[2]**

23. The random variable X represents the weight in kg of a randomly selected male dog of a particular breed. X is Normally distributed with mean 30.7 and standard deviation 3.5.

i. Find

A. P ($X < 30$),

[3]

B. $P\ (25 < X < 35)$.

[3]

ii. Five of these dogs are chosen at random. Find the probability that each of them weighs at least 30 kg.

[2]

iii. The weights of females of the same breed of dog are Normally distributed with mean 26.8 kg. Given that 5% of female dogs of this breed weigh more than 30 kg, find the standard deviation of their weights.

[4]

iv. Sketch the distributions of the weights of male and female dogs of this breed on a single diagram.

[4]

24.

i. There are 5 runners in a race. How many different finishing orders are possible? [You should assume that there are no ‘dead heats’, where two runners are given the same position.]

[1]

For the remainder of this question you should assume that all finishing orders are equally likely.

ii. The runners are denoted by A, B, C, D, E. Find the probability that they either finish in the order ABCDE or in the order EDCBA.

[2]

iii. Find the probability that the first 3 runners to finish are A, B and C, in that order.

[1]

iv. Find the probability that the first 3 runners to finish are A, B and C, in any order.

[2]

25. In a hockey league, each team plays every other team 3 times. The probabilities that Team A wins, draws and loses to Team B are given below.

- P(Wins) = 0.5
- P(Draws) = 0.3
- P(Loses) = 0.2

The outcomes of the 3 matches are independent.

i. Find the probability that Team A does not lose in any of the 3 matches.

[1]

ii. Find the probability that Team A either wins all 3 matches or draws all 3 matches or loses all 3 matches.

[2]

iii. Find the probability that, in the 3 matches, exactly two of the outcomes, 'Wins', 'Draws' and 'Loses'occur for Team A.

[4]

26. An online store has a total of 930 different types of women's running shoe on sale. The prices in pounds of the types of women's running shoe are summarised in the table below.

Price (£x)	$10 \leq x \leq 40$	$40 < x \leq 50$	$50 < x \leq 60$	$60 < x \leq 80$	$80 < x \leq 200$
Frequency	147	109	182	317	175

i. Calculate estimates of the mean and standard deviation of the shoe prices.

[4]

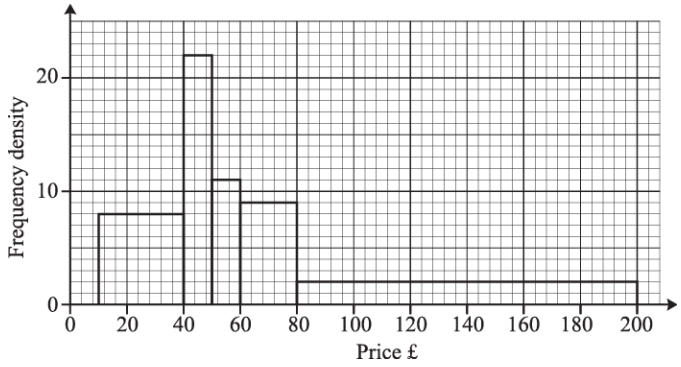
ii. Calculate an estimate of the percentage of types of shoe that cost at least £100.

[3]

iii. Draw a histogram to illustrate the data.

[5]

The corresponding histogram below shows the prices in pounds of the 990 types of men's running shoe on sale at the same online store.



iv. State the type of skewness shown by the histogram for men's running shoes.

[1]

v. Martin is investigating the percentage of types of shoe on sale at the store that cost more than £100. He believes that this percentage is greater for men's shoes than for women's shoes. Estimate the percentage for men's shoes and comment on whether you can be certain which percentage is higher.

[3]

vi. You are given that the mean and standard deviation of the prices of men's running shoes are £68.83 and £42.93 respectively. Compare the central tendency and variation of the prices of men's and women's running shoes at the store.

[2]

(b) Find the probability that each of these two people has a different blood group. **[3]**

28. There are four human blood groups; these are called O, A, B and AB. Each person has one of these blood groups. The table below shows the distribution of blood groups in a large country.

Blood group	Proportion of population
O	49%
A	38%
B	10%
AB	3%

Two people are selected at random from this country.

(a) Find the probability that at least one of these two people has blood group O. **[2]**

29. A farmer has 200 apple trees. She is investigating the masses of the crops of apples from individual trees. She decides to select a sample of these trees and find the mass of the crop for each tree.

(a) Explain how she can select a random sample of 10 different trees from the 200 trees. **[2]**

The masses of the crops from the 10 trees, measured in kg, are recorded as follows.

23.5	27.4	26.2	29.0	25.1	27.4	26.2	28.3	38.1	24.9

(b) For these data find

(c) Show that there is one outlier at the upper end of the data. How should the farmer decide whether to use this outlier in any further analysis of the data? **[3]**

- the mean,
- the sample standard deviation.

30. The maximum daytime temperature was recorded on each day in May 2016 at a weather station in Canada. The data are shown in the stem-and-leaf diagram below.

0	3
5	
10	0 0 1 3 3
15	0 1 1 2 3 4
20	0 0 1 2 2 3 3 4
25	0 0 1 2 2 2
30	0 0 2 2 3

key: 15 | 1 represents a temperature of 16 °C

(a) Describe the shape of the distribution.

(b) Find the interquartile range.

(c) Hence determine whether 3 is an outlier.

31.

(a) (i) The pre-release data shows that the total population of the 239 countries in the world in 2016 was 7174654290. The populations of a sample of 10 countries are given in Fig. 9.1.

Country	**Population**
Tuvalu	10 782
Equatorial Guinea	722 254
Somalia	10 428 043
Denmark	5 569 077
Burma	55 746 253
Norway	5 147 792
Botswana	2 155 784
Rwanda	12 337 138
Sint Maarten	39 689
Swaziland	1 419 623

Fig. 9.1

Show that the mean population per country for the whole world is much larger than the mean population per country for this sample. **[3]**

(ii) Rebecca takes a large number of different samples of 10 countries. She finds that the mean population per country is usually smaller for the sample than it is for the whole world. Explain whether this suggests that the sampling was not random. **[2]**

(b) Fig. 9.2 shows data for Norway.

Country	Population	GDP per capita (US$)	Health expenditure (% of GDP)
Norway	5 147 792	55 400	9.1

Fig. 9.2

Calculate Norway's health expenditure per person in US$. **[2]**

(c) As part of an investigation into the factors which might be associated with life expectancy, the scatter diagrams in Fig. 9.3 are drawn. The line of best fit and the corresponding value of the square of the correlation coefficient are also shown for each scatter diagram.

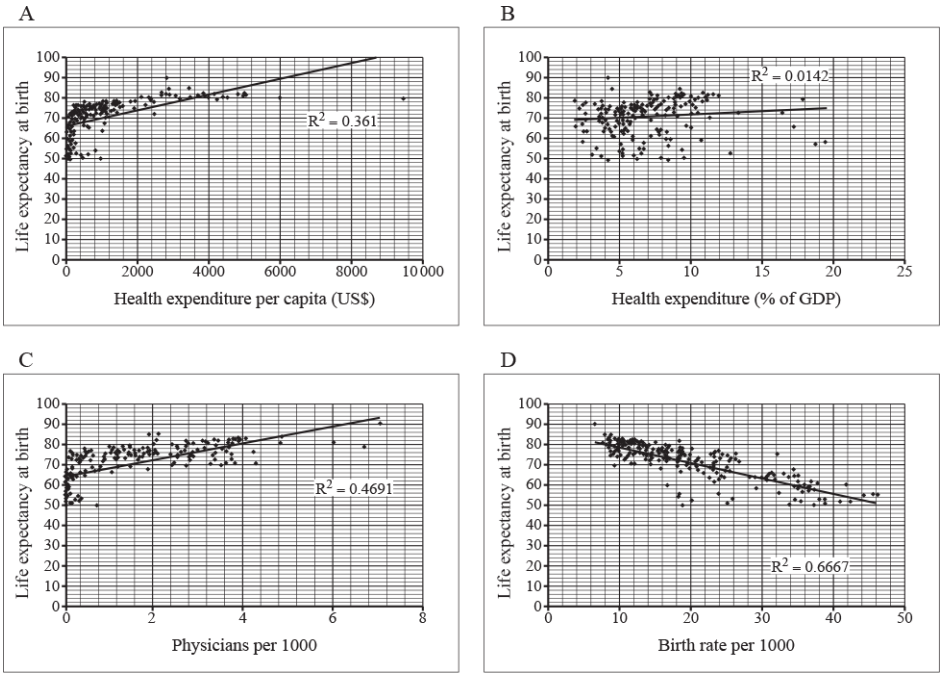


Fig. 9.3

(i) Which of the four factors appears to have the strongest positive association with life expectancy at birth? Give a reason for your answer. **[2]**

(ii) Explain why the line of best fit in scatter diagram B is not a good model for the relationship between the two variables. **[1]**

32. The numbers of units of electricity, x kWh (kilowatt-hours), used by 50 customers of an energy firm in a period of one month are summarised as follows.

$\Sigma x = 17\,100$ $\Sigma x^2 = 6115\,108$

(i) Calculate the mean and standard deviation of x.

(ii) The cost, £y, of the electricity used by each customer is given by the formula $y = 0.108x + 7.2$. Use your answers to part **(i)** to deduce the mean and standard deviation of the costs of the electricity used by these customers.

[3]

33(a). Fig. 10.1 shows a sample collected from the large data set.

BMI is defined as $\frac{\text{mass of person in kilograms}}{\text{square of person's height in metres}}$.

Sex	Age in years	Mass in kg	Height in cm	BMI
Male	38	77.6	164.8	28.57
Male	17	63.5	170.3	21.89
Male	18	68.0	172.3	22.91
Male	18	57.2	172.2	19.29
Male	19	77.6	191.2	21.23
Male	24	72.7	177.0	23.21
Male	25	92.5	177.9	29.23
Male	26	70.4	159.4	27.71
Male	31	77.5	174.0	25.60
Male	34	132.4	182.2	39.88
Male	38	115.0	186.4	33.10
Male	40	112.1	171.7	38.02

Fig. 10.1

Calculate the mass in kg of a person with a BMI of 23.56 a
your answer correct to 1 decimal place.

(b). Fig. 10.2 shows a scatter diagram of BMI against age for the data in the table. A line of best fit has also been drawn.

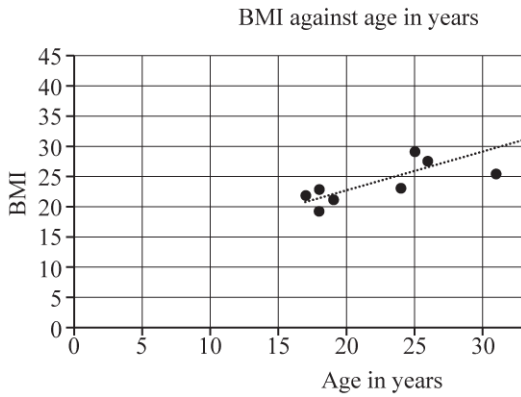


Fig. 10.2

Describe the correlation between age and BMI.

(c).

Use the line of best fit to estimate the BMI of a 30-year-old

(d).

Explain why it would not be sensible to use the line of best
year-old man.

(e).

Use your knowledge of the large data set to suggest **two** re
the table may not be representative of the population.

(f).

Once the data in the large data set had been cleaned there
selection. Describe how a sample of size 12 could be gene
sampling so that each of the 196 values could be selected

34. In a certain school there are 273 boys and 327 girls. A proportional stratified sample of 50 pupils is required.

How many girls should be in the sample?

35. Doug has a list of times taken by competitors in a ‘fun run’. He has grouped the data and calculated the frequency densities in order to draw a histogram to represent the information. Some of the data are presented in Fig. 2.

Time in minutes	15 −	20 −	25 −	35 −	45 − 60
Number of runners	12	23	59	71	
Frequency density	2.4		5.9	7.1	1.4

Fig. 2

[2]

(a) Write down the missing values in Fig. 2 above.

[1]

(b) Doug labels the horizontal axis on the histogram ‘time in minutes’ and the vertical axis ‘number of minutes per runner’. State which one of these labels is incorrect and write down a correct version.

36. At the start of the January term year 11 students at Amplesides College sat mock examinations in GCSE mathematics papers 4 and 5. A teacher collected the results and presented them in a scatter diagram, which is shown in Fig. 10. A line of best fit has been added.

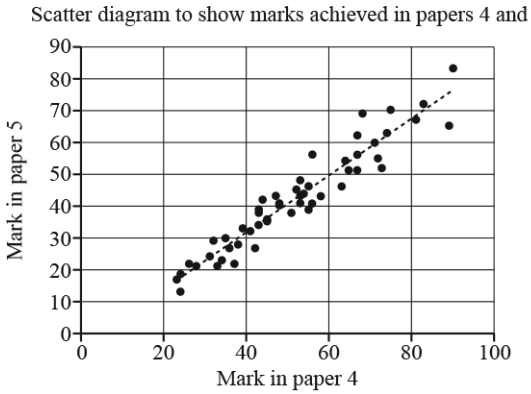


Fig. 10

The correlation coefficient for the data is 0.9566.

[2]

(a) Give **two** reasons why it is reasonable to model the relationship between the mark achieved in paper 4 and the mark achieved in paper 5 by a straight line.

The equation of the line of best fit is $y = 0.89x - 3.76$, where y is the mark achieved

in paper 5 and x is the mark achieved in paper 4.

[1]

(b) Tina achieved a mark of 83 in paper 4, but was absent for paper 5. Calculate an estimate of the mark she would have achieved in paper 5.

[2]

(c) Dave was absent for paper 4. He achieved a mark of 8 in paper 5. Calculate an estimate of the mark Dave would have achieved in paper 4.

(d) Explain why the estimate of Tina's mark for paper 5 is more reliable than the estimate of Dave's mark for paper 4. **[1]**

END OF QUESTION PAPER

Question			Answer/Indicative content	Marks	
1		i	$3\times\frac{5}{10}\times\frac{4}{9}\times\frac{5}{8}=\frac{300}{720}=\frac{5}{12}=(0.4167)$	M1	For 5/10 × 4/9
		i		M1	For × 5/8
		i		M1	For 3 × triple product
		i		A1	CAO (Fully simplified)
		i	**Or** $\frac{\binom{5}{2}\times\binom{5}{1}}{\binom{10}{3}}=\frac{10\times5}{120}=\frac{5}{12}$	M1*	For $\binom{5}{2}\times\binom{5}{1}$
		i		M1*	For $\binom{10}{3}$
		i		M1*dep	For whole fraction
		i		A1	CAO (Fully simplified) **Examiner's Comments** Candidates using the ${}^nC_r$ meth successful, as when using the many did not realise that they product by 3. A small minority instructions and either left a fra (usually 15/36) or gave the ans
		ii	$4\times\frac{7}{12}\times\left(\frac{5}{12}\right)^3+\left(\frac{5}{12}\right)^4$	M1FT	For first probability
		ii		M1FT	For $(5/12)^4$
		ii		M1FT	For sum of both correct probab
		ii	= 0.169 + 0.030 × 0.199 Or = $\frac{875}{5184}+\frac{625}{20736}=\frac{1375}{6912}$	A1	CAO Do not allow 0.2, unless fuller a **Examiner's Comments** Most candidates made a reaso their answer from part (i). How probability, or missed the coeff probability of 3 evenings, not re situation. Some candidates cal rather than *at least* 3, and thus minority of candidates used sta calculators to just write down a strategy, as a slip in copying th penalised, since no method wa

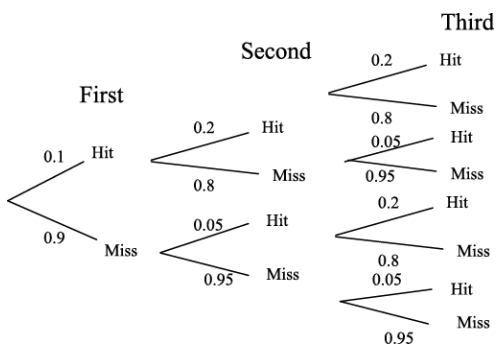
			Total	**8**	
2		i	$X \sim B(50, 0.1)$	M1	For $0.1^5 \times 0.9^{45}$
		i	$P(5 \text{ underweight}) = \binom{50}{5} \times 0.1^5 \times 0.9^{45} = 0.1849$	M1	For $\binom{50}{5} \times p^5 \times q^{45}$
		i		A1	CAO **Examiner's Comments** This question was very well an scoring all 3 marks. However, a have no idea about the binomia
		ii	$X \sim B(20, 0.1)$		
		ii	$P(X \geq 1) = 1 - P(X = 0)$	M1	For 0.1216 CAO
		ii	$= 1 - 0.1216 = 0.8784$	A1	**Examiner's Comments** Again another well answered q candidates did not read the que use $n = 50$ in their calculation.
		iii	$E(X) = 48 \times 0.8784 = 42.16 (= 42.2)$	M1	
		iii		A1	FT their probability from part (ii **Examiner's Comments** Full marks were available here part (ii), so many candidates m incorrect answer. However a la rounded their answer to the ne a mark. Others over-specified t mark. Other common errors we their answer to part (ii), or to us
			Total	**7**	
3		i	Because if people cannot make a correct identification, then the probability that they guess correctly will be 0.5	E1	For idea of a guess or 'chosen

		i	For ‘equally likely to guess right or wrong’ or ‘two outcomes with equal probability’ or ‘50:50 chance of success’ or ‘right one in two occasions on average’ or ‘two (equally likely) outcomes’ etc	E1	For idea of two outcomes **Examiner's Comments** The wording of the researcher’ confusion for some of the cand This was translated into some conclusions in all three parts o comprehension skills are requi unfortunately, these skills were Many candidates scored both r proportion lost either the first o mentioning ‘guess’ or only inclu question or not mentioning, in a possible outcomes. Some cand null hypothesis in words.
		ii	‘Because people may do better than they would by guessing’ or similar	B1	For idea of selecting correctly / **Examiner's Comments** The wording of the researcher’ confusion for some of the cand This was translated into some conclusions in all three parts o comprehension skills are requi unfortunately, these skills were This was not as well answered to distinguish between guessin between the two types of wate mark because they gave the re hypothesis as ‘13 people out o identified correctly’ which of co
		iii	$P(X \geq 13) = 1 - P(X \leq 12) = 1 - 0.8684 = 0.1316$	M1	For notation $P(X \geq 13)$ or $P(X >$
		iii		B1*	For 0.1316
		iii	$0.1316 > 0.05$	M1* dep	For comparison with 5%
		iii	So not significant	A1*	
		iii	There is insufficient evidence to suggest that people can make a correct identification.	E1* dep	NB Point probabilities score ze
		iii	ALTERNATIVE METHOD – follow method above unless some mention of CR seen		Must see some reference to CI

		iii	Critical region method UPPER TAIL $P(X \geq 14) = 1 - P(X \leq 13) = 1 - 0.9423 = 0.0577 < 5\%$	B1	For either probability
		iii	$P(X \geq 15) = 1 - P(X \leq 14) = 1 - 0.9793 = 0.0207 < 5\%$	M1*	For a correct comparison with 5
		iii	So critical region is {15,16,17,18,19,20}	M1* dep	cao dep on the two correct pro
		iii	13 not in CR so not significant	A1*	Must include '13 not in CR'
		iii	There is insufficient evidence to indicate that people can make a correct identification.	E1* dep on A1	Ignore any work on lower critic **Examiner's Comments** The wording of the researcher' confusion for some of the cand This was translated into some conclusions in all three parts o comprehension skills are requi unfortunately, these skills were The most successful way of ap was to compare $P(X \geq 13)$ with of the candidates, who used th mark due to not putting the exp question. Other candidates use $P(X \geq 12)$ or $P(X \geq 14)$. Candidates who used the critic gained the first two marks but t any more marks – usually beca critical region. Unfortunately so at the two probabilities necessa made no mention of the critical not gain any marks. It is pleasing to report, on the c candidates tried to use point pr full marks could be obtained by many candidates either compa comparison at all - such candid credit.

			Total	**8**	
4		i	Median = 3.32 kg	B1	
		i	Q1 (= 6.5th value) = 2.83 Q3 (= 19.5th value) = 3.71	B1	For Q1 or Q3
		i	Inter-quartile range = 3.71 − 2.83 = 0.88	B1	For IQR dep on both quartiles **Examiner's Comments** Most candidates successfully f instead of the 13th value some 13^{th} values. However, candidat finding the interquartile range. found correctly, but the upper q wrong, with an answer of 3.665 Occasionally candidates did no interquartile range, but instead their quartiles.
		ii		G1	For reasonably linear scale sho
		ii		G1	For boxes in approximately cor to right of centre For whiskers in approximately the box FT their median and quartiles i only for correct values
		ii		G1	**Examiner's Comments** The response to this question because they were faced with paper, most candidates though required. Very few had a scale failed to make it linear. Some c

					box and whisker plot and then relevant values. This did not ga instructs candidates to 'Draw a seems likely that many candida not think to use a ruler. Far too seen, with the sizes of the box the median not in proportion.
		iii	Lower limit 2.83 − (1.5 × 0.88) = 1.51	B1	For 1.51 FT
		iii	Upper limit 3.71 + (1.5 × 0.88) = 5.03	B1	For 5.03 FT
		iii	Exactly one baby weighs less than 1.51 kg and none weigh over 5.03 kg so there is exactly one outlier.	E1*	Dep on their 1.51 and 5.03
		iii	'Nothing to suggest that this baby is not a genuine data value so she should not be excluded' or 'This baby is premature and therefore should be excluded'.	E1* Dep	Any sensible comment in conte **Examiner's Comments** Many candidates correctly four the outliers. The most common were calculated using median errors were also seen. A few ca mean and standard deviation, correct, full marks were availab other of the two statistics was u It was necessary to check both only one outlier, but some cand Many candidates failed to give regarding the outlier, though th point about premature babies.
		iv	Median = 3.5 kg	B1	
		iv	Q1 = 50th value = 3.12 Q3 = 150th value = 3.84	B1	For Q1 or Q3 For IQR FT their quartiles
		iv	Inter-quartile range = 3.84 − 3.12 = 0.72	B1	**Examiner's Comments** As in part (i), the median was u candidates lost a mark due to i in finding the quartiles.
		v	Female babies have lower weight than male babies on the whole	E1	Allow 'on average' or
		v		FT	similar in place of 'on the whole
		v	Female babies have higher weight variation than male babies	E1	Allow 'more spread' or

					similar but not ‘higher range’ Condone less consistent **Examiner's Comments** Only about one third of candida
		v		FT	was given to those candidates medians and interquartile rang what they meant. Candidates w failed to get credit without a co average' or 'tend to be'. Similar or 'less spread' gained credit fo range’ was not awarded credit.
		vi	Male babies must weigh more than 4.34 kg		
		vi	Approx 10 male babies weigh more than this.	M1*	For 10 or 9 or 8
		vi	Probability $= \frac{10}{200} \times \frac{9}{199} = \frac{90}{39800} = \frac{9}{3980} = 0.00226$ or $\frac{9}{200} \times \frac{8}{199} = \frac{72}{39800} = 0.00181$	M1* dep	For first fraction multiplied by a binomial probability) CAO Allow their answer to min of 2 sf **Examiner's Comments**
		vi	or $\frac{8}{200} \times \frac{7}{199} = \frac{56}{39800} = \frac{7}{4975} = 0.00141$	A1	This part discriminated very we candidates. Many candidates r male babies weighed more tha then did not know how to proce (10/200) rather than multiplying who misread the scale but knew Special Case mark. A significa out this part altogether.
			Total	**18**	
5		i		G1	For first set of branches
		i		G1	For second set of branches (in
		i		G1	For third set of branches (indep
		i		G1	For labels **Examiner's Comments** The majority of tree diagrams w correct labelling. Weaker candi confused and made errors in th
		ii	P(Hits with at least one) = 1 − P(misses with all)	M1*	For 0.9 × 0.95 × 0.95

		ii	= 1 − (0.9 × 0.95 × 0.95) = 1 − 0.81225 = 0.18775	M1*dep	For 1 − ans
		ii		A1	CAO
		ii	ALTERNATIVE METHOD only if there is an attempt to add 7 probabilities		
		ii	At least three correct triple products	M1	
		ii	Attempt to add 7 triple products	M1	
		ii		A1	CAO
		ii	FURTHER ALTERNATIVE METHOD		
		ii	0.1 + 0.9 × 0.05	M1	
		ii	Above probability + 0.9 × 0.95 × 0.05	M1	
					CAO
					Examiner's Comments
		ii		A1	Many candidates employed the usually successfully, but a sign protracted method of listing all least one hit. Usually errors we approach.
		ii	P(Hits with exactly one)	M1	For two correct products
		ii	= (0.1 × 0.8 × 0.95) + (0.9 × 0.05 × 0.8) + (0.9 × 0.95 × 0.05)	M1	For all three correct products
		ii	$= 0.076 + 0.036 + 0.04275 = \frac{19}{250} + \frac{9}{250} + \frac{171}{4000}$	M1	For sum of all three correct pro
					CAO
					Examiner's Comments
		ii	$= \frac{619}{4000} = 0.15475$	A1	Most candidates found the corr calculated them correctly. A sm three. For those who got the tr through marks were available.
		iii	P(Hits with exactly one given hits with at least one) $= \frac{\text{P(Hits with exactly one and hits with at least one)}}{\text{P(Hits with at least one)}}$		
		iii		M1	For numerator FT
		iii	$= \frac{0.15475}{0.18775}$	M1	For denominator FT
					CAO
		iii	=0.8242	A1	**Examiner's Comments**
					Many of those who reached thi

					However, there was considera conditional probability, often wi wrong numerator of P(at least candidates inverted the fractio
		iv	P(Hits three times overall) = (0.1 × 0.2 × 0.2) + (0.9 × 0.95 × 0.95 × 0.05 × 0.2 × 0.2)	M1	For 0.1 × 0.2 × 0.2 or 0.004 or
		iv		M1	For 0.9 × 0.95 × 0.95 × 0.05 ×
		iv	= 0.004 + 0.0016245	M1* Dep on both prev M1's	For sum of both
					CAO
					Examiner's Comments
		iv	= 0.0056245	A1	Approximately one third of can part. However many were conf successfully found the first pro second, or found additional pro second product often made err were often 0.1×0.2×0.2 rather
			Total	**18**	
6		i	$P(Y = 76) = P\left(\frac{75.5-76}{12} \le Z \le \frac{76.5-76}{12}\right)$ = P(−0.04166… < *Z* < 0.04166…) = Φ(0.04166…) − (1 − Φ(0.04166…)) = 2 × Φ(0.04166…) − 1 = 2 × 0.5167 − 1 = 0.0334	B1	For one correct continuity corre
		i		M1	For standardizing
		i		M1	For correctly structured probab
					CAO inc use of diff tables. Allo
					Examiner's Comments
		i		A1	Most candidates obtained a co significant number did not use corrections. Most used the diff Normal table correctly to provi relatively small number struggl calculation.

		ii	$P(Y \geq 80) = P\left(Z \geq \frac{79.5-76}{12}\right)$ = P(Z > 0.2917) = 1 − Φ(0.2917) = 1 − 0.6148 = 0.3852 = 0.385 to 3 sig fig	B1	For correct cc used
		ii		M1	For correct structure CAO do not allow 0.386
		ii		A1	**Examiner's Comments** Many candidates answered co was to omit or to provide an inc relatively small number did not correctly.
		iii	3 × 0.3852 × 0.6148^2 = 0.4368	M1	3 × their p × (1 − their p)^2 FT their p. Allow 2sf if working
		iii		A1	**Examiner's Comments** Most candidates knew and app many were dependent on the F number omitted the x3 from the
		iv	**EITHER:** P(Score ≥ k) = 0.1		
		iv	Φ^{-1} (0.9) = 1.282	B1	For 1.282
		iv		M1	Allow k − 0.5 used for k. Positiv
		iv	k = 76 + (1.282 × 12) = 91.38 or k = 76 + 0.5 + (1.282 × 12) = 91.88	A1	For 91.38 or 91.88
		iv	91.38 > 90.5 or 91.88 > 91	M1	Relevant comparison (e.g. diag
		iv	so lowest reported mark = 92	A1	
		iv	**OR** Trial and improvement method	M1	M1 for attempt to find P(Mark ≥
		iv	P(Mark ≥ 91) = P(Score ≥ 90.5) = 0.1135	A1	A1 for 0.1135
		iv	P(Mark ≥ 92) = P(Score ≥ 91.5) = 0.0982		M1 for comparisons
		iv	P(Mark ≥ 91) > 10% and P(Mark ≥ 92) < 10%	M1	**Examiner's Comments** 1.282 was identified by the maj to set up a correct equation an of these gave 92% as the final 91%. Others rearranged incorr first calculation. Few candidate understanding of the requireme
		iv	so lowest reported mark = 92	A1	

		v	P($Y \leq 50$) = 0.2		
		v		B1	For 50.5 used
		v		B1	For −0.8416. Condone − 0.842 Condone 0.8416 if numerator r
		v		M1	For structure. CAO **Examiner's Comments**
		v	μ = 50.5 + (12 × 0.8416) = 60.6	A1	In this part, the continuity corre incorrect value was used, by m +0.8416 leading to 60.0992 wh issue of over-specification was question.
			Total	**18**	
7		i	(*A*) P(Watched cyc but not fb) = $\frac{15}{250} = \frac{3}{50}$ = 0.06	B1	CAO (aef) **Examiner's Comments** This was answered very well, a gave the number that watched than the probability. A few had or 100.
		i	(*B*) P(Watched one or two) = $\frac{33+12+21+14+3+65}{250}$	M1	OR: $\frac{250-(64+38)}{250}$ =
		i	$= \frac{148}{250} = \frac{74}{125}$ = 0.592	A1	CAO (aef) **Examiner's Comments** Again this was very well-answe candidates making errors. The include those people who watc out one of the six who watched
		ii	P(Not watched fb|watched cyc) = $\frac{15}{67}$ = 1567= 0.224 (0.223880597…)	M1	CAO (aef) **Examiner's Comments**
		ii		A1	Approximately two thirds of car Of the rest, some were able to correct denominator but then fa numerator, often thinking that i candidates either did not recog

					probability question, or did not probability.
			Total	**5**	
8		i	0 | 6 1 | 5 8 2 | 1 5 8 3 | 1 1 3 5 8 9 Key 1 | 8 represents 18 people	G1	Stem (in either order) and leav
		i		G1	Sorted and aligned
		i		G1	Key **Examiner's Comments** Most candidates scored all thre accurately align the leaves or c and thus scored only 2 marks. than 2 out of 3.
		ii	Negative	B1	**Examiner's Comments** This was very well-answered w skew was positive.
		iii	Median = 29.5	B1	CAO
		iii	Mean = 26.7 (26.6666) or $26^2/_3$ or $^{80}/_3$ or 26.6	B1	CAO
		iii	Mode = 31	B1	CAO
		iii	The mode is not at all useful as it is just by chance that it is 31. Mark awarded for stating not useful and -not representative of data -does not represent Central Tendency -happened by chance (or similar) -comment about not appearing significantly more (only one repetition/only twice/etc) No mark for stating it would be useful OR NOT USEFUL because of -spread/range -sample size	E1	Allow any reasonable commen **Examiner's Comments** The mean, median and mode w although one or two candidates specification of the mean or rou the final mark for the comment quarter of candidates. Many ca descriptions of the usefulness commenting on this particular c stated incorrectly that the mode correctly stated that it was not

			-negatively skewed -unaffected by outliers -isn't close to mean and median		an incorrect reason such as be being negatively skewed; or no and/or median.
			Total	**8**	
9		i	$\binom{25}{15}$ $= 3268760$	M1	
		i		A1	**Examiner's Comments** This was very well-answered.
		ii	$\binom{13}{8} \times \binom{12}{7} = 1287 \times 792$ $= 1019304$	M1	For product of both correct com CAO
		ii		A1	**Examiner's Comments** This was again usually answer common error was to add the t multiply them.
		iii	1019304/3268760	M1	For their (ii) divided by their (i)
		iii	= 0.312 (0.311832) Allow fully simplified fraction11583/37145	A1FT	
		iii	**OR**		
		iii	$\binom{15}{8} \times \frac{13}{25} \times \frac{12}{24} \times \frac{11}{23} \times \frac{10}{22} \times \frac{9}{21} \times \frac{8}{20} \times \frac{7}{19}$ $\times \frac{6}{18} \times \frac{12}{17} \times \frac{11}{16} \times \frac{10}{15} \times \frac{9}{14} \times \frac{8}{13} \times \frac{7}{12} \times \frac{6}{11}$	M1	For product of fractions with co
		iii	= 0.312	A1	**Examiner's Comments** This was another well-answere had added in part (ii) still usual through. Candidates who tried (instead of simply dividing their answer to part (ii)) were rarely have the correct product of 15 multiplied this by any, let alone
			Total	**6**	
10		i	(A) $X \sim B(20, 0.78)$ P(Exactly 19 cured) = $\binom{20}{19} \times 0.78^{19} \times 0.22^{1}$	M1	For $0.78^{19} \times 0.22^{1}$

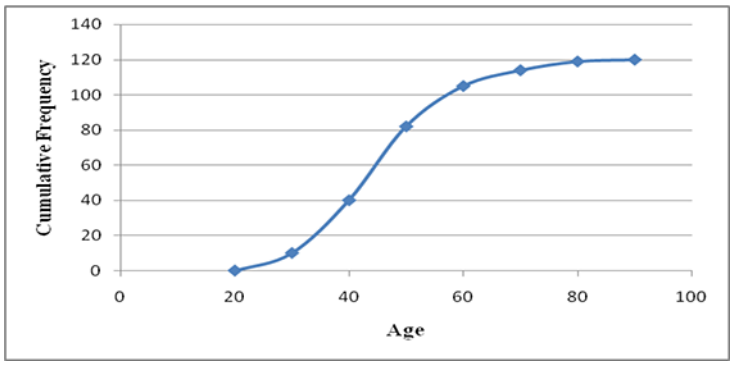
		i		M1	For $\binom{20}{19} \times p^{19} \times q^{1}$ CAO
		i	= 0.0392 (0.039197)	A1	**Examiner's Comments** This was generally very well-ar
		i	(*B*) P(Exactly 20 cured) = $\binom{20}{20} \times 0.78^{20} \times 0.22^{0} = 0.0069$	M1	For 0.78^{20} oe
		i	P(At most 18 cured) = 1 − (0.0069 + 0.0392)	M1	For P(19) + P(20) CAO
		i	= 0.954 (0.95385)	A1	**Examiner's Comments** Although around two-thirds of c correctly, some candidates incl and thus were only able to gair CAO
		i	(*C*) E(*X*) = *np* =20 × 0.78 = 15.6	B1	**Examiner's Comments** The majority of the candidates but a minority lost the mark wh answer to 15 or 16.
		ii	Let *X* ~ B(20, 0.78) Let *p* = probability of a patient being cured (for population)	B1	For definition of *p*
		ii	H_0: *p* = 0.78	B1	For H_0
		ii	H_1: *p*> 0.78	B1	For H_1
		ii	P(*X*≥ 19) = 0.0392 + 0.0069	B1	For **NOTATION** P(*X* ≥ 19) or P(*X* > 18) or 1 − P(X ≤ 18) or 1 − P(X < 19)

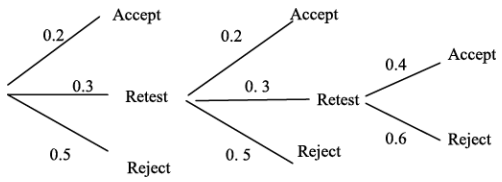
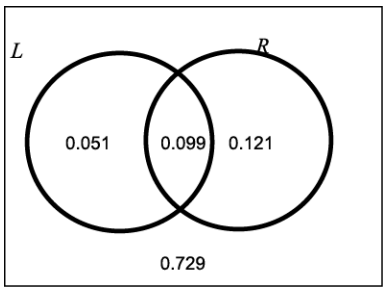
		ii	= 0.0461	B1*	CAO For 0.0461 allow 0.0462
		ii	0.0461 > 1%	M1* dep	For comparison with 1%
		ii	So not significant.	A1	
		ii	Conclude that there is not enough evidence to suggest that the new drug is more effective than the old one.	E1	
		ii	ALTERNATIVE METHOD FOR FINAL 5 MARKS		If combination of methods used mark.
		ii	P($X \geq 19$) = 0.0461 > 1%	3	For either probability
		ii	P($X \geq 20$) = 0.0069 < 1%	M1	For at least one comparison wi
		ii	So critical region is {20}	B1*	CAO dep on the two correct pr
		ii	(19 not in CR so) not significant.	A1* dep	Dep on correct CR Ignore any work on lower critic
		ii	Conclude that there is not enough evidence to suggest that the new drug is more effective than the old one.	E1* dep	**Examiner's Comments** In recent years, candidates hav hypothesis test questions than the case this year. Many fully c Most candidates scored the firs hypotheses, with most now kno The vast majority of successful method, finding P($X \geq 19$) and was pleasing to see that most answer in context and with an something to the effect of 'there suggest that…'. Those who trie method were less successful o use point probabilities, being a marks for the hypotheses. A fe and there full marks available f tables.

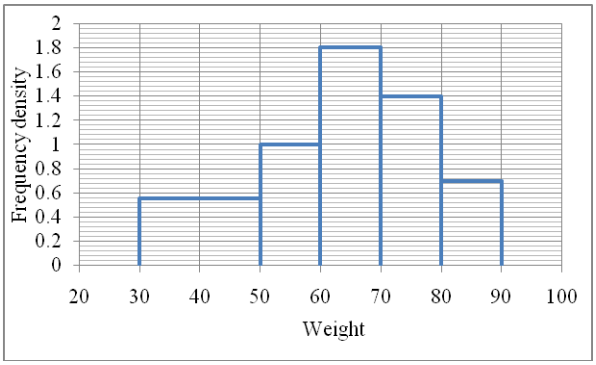
		iii	With a 5% significance level rather than a 1% level, the null hypothesis would have been rejected. OR: 'there would be enough evidence to suggest that the new drug is more effective than the old one.'	B1*	oe oe
		iii	This is because 0.0461 < 5%	B1* dep	**Examiner's Comments** Candidates who gained more c tended to gain full marks in this available if point probabilities w
			Total	**17**	
11		i	Inter-quartile range = 18.1 − 17.8 = 0.3	B1	
		i	Lower limit 17.8 − (1.5 × 0.3) (= 17.35)	M1	
		i	No outliers at lower end.	A1	dep on 17.35
		i	Upper limit 18.1 + (1.5 × 0.3) (= 18.55)	M1	
		i	(Max is 18.6) so at least one outlier at upper end.	A1	dep on 18.55
		i			**Examiner's Comments** Most candidates calculated the correctly to find the limits for ou median instead of the quartiles × inter-quartile range. Many ca on outliers at each end separat were outliers overall).
		ii	P(*A*) = P(All 3 have orange centres) = $\frac{7}{20}\times\frac{6}{19}\times\frac{5}{18}=\frac{7}{228}$	M1	For 7/20×
		ii		M1	For product of correct three fra Without extra terms

		ii	= 0.0307 (0.030702)	A1	CAO Allow full marks for fully simplif
		ii	P(*B*) = P(All 3 have same centres) = $\left(\frac{7}{20}\times\frac{6}{19}\times\frac{5}{18}\right)+\left(\frac{6}{20}\times\frac{5}{19}\times\frac{4}{18}\right)+\left(\frac{4}{20}\times\frac{3}{19}\times\frac{2}{18}\right)+\left(\frac{3}{20}\times\frac{2}{19}\times\frac{1}{18}\right)=$	M1	For at least two correct triple pr
		ii	=0.0307 + 0.0175 + 0.0035 + 0.0009	M1	For sum of all four correct
		ii	$=0.0526 \quad =\frac{1}{19}$ (0.052632)	A1	CAO Allow 0.053 or anything which **Examiner's Comments** This question was answered w struggled with earlier questions fractions (which were exact) wi specification. However many c up into the first part of the answ there was more space on the n forgot to take into consideration with $(7/20)^3$, gained only 1 mar through marks were available i
		ii	$\left(=\frac{7}{228}+\frac{1}{57}+\frac{1}{285}+\frac{1}{1140}\right)$		
		iii	$P(A \mid B)=\frac{0.0307..}{0.0526..}$	M1	For their 'A' divided by their 'B'
		iii	= 0.583 (= 0.58333)	A1	FT their answers to (ii) provide CAO **Examiner's Comments** Although over half of candidate 3 here, a number of candidates P(*A*|*B*) = P(*A*)/P(*B*) here, and ir P(*A*∩*B*) = P(*A*) × P(*B*), and the top and bottom of the fraction. understanding here of indepen disappointing number of candid an answer greater than 1 for P
		iii	P(*B* | *A*) = 1	B1	
		iv	P(All have orange centres) = 0.0307^2 = 0.00094 or $=\frac{49}{51984}$	M1	For their 0.0307^2 FT **Examiner's Comments** This was well-answered althou was 2 × P(A).
		iv	= (0.00094260)	A1	

		v	P(Has to select > 2) = 1 – P(Has to select ≤ 2) $= 1-\left(\frac{14}{20}+\left(\frac{6}{20}\times\frac{14}{19}\right)\right)$ = 1 – (0.7 + 0.221) = 1 – 0.921	M1	For $\left(\frac{6}{20}\times\frac{14}{19}\right)$
		v		M1	For 1 – sum of both
		v	= 0.079 (=0.078947)	A1	CAO
		v	OR		
		v	P(Has to select > 2) = P(First 2 both cherry) $=\left(\frac{6}{20}\times\frac{5}{19}\right)$	M2	For whole product
		v	$= 0.079 \quad =\frac{3}{38}$	A1	CAO
		v	OR		
		v	1 − (P(0 cherries) + P(1 cherry))= $1-\left(\frac{14}{20}\times\frac{13}{19}+\left(\frac{6}{20}\times\frac{14}{19}\right)+\left(\frac{14}{20}\times\frac{6}{19}\right)\right)$ = 1 − (0.4789 + 0.2211 + 0.2211) = 1 − 0.9209	M1	For any term
		v		M1	For 1 – sum of all three
		v	= 0.079	A1	CAO
		v	**OR**		
		v	$\left(\frac{6}{20}\times\frac{5}{19}\times\frac{14}{18}\right)+\left(\frac{6}{20}\times\frac{5}{19}\times\frac{4}{18}\times\frac{14}{17}\right)+$ $\left(\frac{6}{20}\times\frac{5}{19}\times\frac{4}{18}\times\frac{3}{17}\times\frac{14}{16}\right)+\left(\frac{6}{20}\times\frac{5}{19}\times\frac{4}{18}\times\frac{3}{17}\times\frac{2}{16}\times\frac{14}{15}\right)+$ $\left(\frac{6}{20}\times\frac{5}{19}\times\frac{4}{18}\times\frac{3}{17}\times\frac{2}{16}\times\frac{1}{15}\times\frac{14}{14}\right)$	M1	For any term
		v	$=\frac{7}{114}+\frac{14}{969}+\frac{7}{2584}+\frac{7}{19380}+\frac{1}{38760}$	M1	For sum of all five terms (all co
					CAO
					Examiner's Comments
		v	=0.079	A1	Only approximately a quarter o completely correct answer, alth correct answer on the way to a answers used the 6/20 × 5/19 r then compromising this with an
			Total	**19**	
12		i	Upper Bound: 20, 30, 40, 50, 60, 70, 80, 90 Cumulative Freq: 0, 10, 40, 82, 105, 114, 119, 120	B1	Cumulative frequencies All correct

		i		G1	For plotted points (Provided plotted at correct UCB positions)
		i		G1	For joining points (within ½ a square)
		i		G1	For scales For labels All marks dep on good attempt at cumulative frequency, but not cumulative fx's or other spurious values.
		i		G1	**Examiner's Comments** Many candidates gained full cr resulted in the loss of 2 marks but at mid-points. Only a few u Some candidates drew cumula number just plotted frequency candidates forgot to label their word "cumulative" on their verti
		ii	Median = 45	B1	Allow answers between 44 and Otherwise check curve. No marks if not using diagram.
		ii	Q1 = 37 Q3 = 53	B1	For Q3 or Q1 Allow Q1 between 37 and 38 w Allow Q3 between 52 and 54 w
		ii	Inter-quartile range = 53 − 37 = 16	B1	For IQR providing both Q1 and

		ii			Alternative version of tree diag **Examiner's Comments** This part was very well answer picking up the follow through m median and quartiles from thei
			Total	**8**	
13		i	Because P(L|R) ≠ P(L)	E1	If two or more methods given a award the mark Allow 0.45 ≠ 0.15 **Examiner's Comments**
		i			The majority of candidates who P(L∩R) = 0.099 ≠ P(L) × P(R) gave the simplest explanation the former, candidates had to but for the latter the symbolic r the probabilities were given in
		ii	P(L ∩ R) = P(L | R) × P(R) = 0.45 × 0.22	M1	For product CAO **Examiner's Comments**
		ii	= 0.099	A1	There were three common ans correctly obtained 0.099, but s wrong probabilities together to working was generally given b incorrect answers
		iii		G1	For two labelled intersecting ci labelling.
		iii		G1	For at least 2 correct probabilit FT their P(L ∩ R) from part (ii)
		iii		G1	For remaining probabilities. FT their P(L ∩ R) providing pro

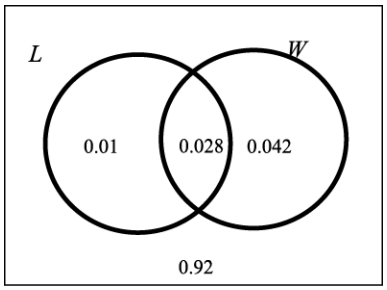
					Examiner's Comments Most candidates gained full cre through of a wrong answer to p to subtract P(L∩R) away from able to score one mark for the
			Total	**6**	
14		i	<table><tr><th>Weight</th><th>Frequency</th><th>Group Width</th><th>Frequency density</th></tr><tr><td>$30 \le w < 50$</td><td>11</td><td>20</td><td>0.55</td></tr><tr><td>$50 \le w < 60$</td><td>10</td><td>10</td><td>1</td></tr><tr><td>$60 \le w < 70$</td><td>18</td><td>10</td><td>1.8</td></tr><tr><td>$70 \le w < 80$</td><td>14</td><td>10</td><td>1.4</td></tr><tr><td>$80 \le w < 90$</td><td>7</td><td>10</td><td>0.7</td></tr></table>	M1	For fd's – at least 3 correct Acc as eg freq per 10g.
		i		A1	
		i		G1	linear scales on both axes and Vertical scale starting from zer mark for heights if broken)
		i		G1	width of bars
		i		G1	height of bars **Examiner's Comments** Most candidates found the frec usually then went on to draw th failed to start the frequency de axes. A few candidates used ir axis, which attracted a penalty scales on the vertical axis was some candidates vulnerable to positions. In particular the heig incorrectly plotted at 0.5 rather

		ii	Mean= $\frac{(40\times11)+(55\times10)+(65\times18)+(75\times14)+(85\times7)}{60}=\frac{3805}{60}$	M1	For midpoints Products are 440, 550, 1170, 1 CAO (exact answer 63.41666…) **Examiner's Comments**
		ii	= 63.4 (or 63.42)	A1	The calculation of the mean of cases accurately performed us calculation of the standard dev Whilst there were many correct factor in the frequencies and w Over specification of either or b some candidates to lose one m
		ii	$\sum x^2 f = (40^2 \times 11) + (55^2 \times 10) + (65^2 \times 18) + (75^2 \times 11) + (85^2 \times 7)$ = 253225		
		ii	$S_{xx} = 253225 - \frac{3805^2}{60} = 11924.6$	M1	For attempt at S_{xx} Should include sum of at least
		ii	$s = \sqrt{\frac{11924.6}{59}} = \sqrt{202.11} = 14.2$		At least 1dp required Use of mean 63.4 leading to ar 12051.4 gets full credit.
		ii		A1	
		ii			63.42 leads to 14.2014… Do not FT their incorrect mean (exact answer14.2166…)
		iii	$\bar{x} - 2s = 63.4 - (2 \times 14.2) = 35$	M1	For either No marks in (iii) unless usin $\bar{x}+$
		iii	$\bar{x} + 2s = 63.4 + (2 \times 14.2) = 91.8$	A1	For both (FT) Must include an element of dou **Examiner's Comments**
		iii	So there are probably some outliers at the lower end, but none at the upper end	E1	Most candidates scored at leas many omitted the fact that ther the top end of the data and/or s some outliers present at the bo mark.

		iv	Mean $= \frac{3624.5}{50} = 72.5$g (or exact answer 72.49g)	B1	CAO Ignore units
		iv	$S_{xx} = 265416 - \frac{3624.5^2}{50} = 2676$	M1	For S_{xx}
		iv	$s = \sqrt{\frac{2676}{49}} = \sqrt{54.61} = 7.39$g	A1	CAO ignore units Allow 7.4 but NOT 7.3 (unless **Examiner's Comments** This was generally very well ar
		v	Variety A have lower average than Variety B oe	E1	FT their means Do not condone lower central t FT their sd **Examiner's Comments**
		v	Variety A have higher variation than Variety B oe	E1	For this type of question candid discuss 'average' and 'variation that the mean of A is lower tha any credit.
			Total	**17**	
15		i	(*A*) $X \sim$ B(15, 0.85)	M1	For $0.85^{12} \times 0.15^3$
		i	P(exactly 12 germinate) $= \binom{15}{12} \times 0.85^{12} \times 0.15^3$	M1	For $\binom{15}{12} \times p^{12} \times q^3$
		i	= 0.2184	A1	CAO
		i	**OR**	OR	
		i	from tables: 0.3958 − 0.1773	M2	For 0.3958 − 0.1773 CAO
		i	= 0.2185	A1	**Examiner's Comments** This was generally very well ar
		ii	(*B*) P(X<12) = P($X \leq 11$) = 0.1773	M1	For P($X \leq 11$) or P(≤11) (With no extras)
		ii		A1	CAO (as final answer) May see alternative method: 0.3958 − 0.2185 = 0.1773 0.3958 − their wrong answer to **Examiner's Comments** Although most candidates answ

					P(X ≤ 12) rather than P(X ≤ 11 probability but then subtracted
		iii	Let p = probability of a seed germinating (for the population)	B1	For definition of p in context
		iii			
		iii	H_0: $p = 0.85$	B1	For H_0
		iii	H_1: $p < 0.85$	B1	For H_1
					Dep on < 0.85 used in H_1 Do not allow just 'Germination
		iii	H_1 has this form because the test is to investigate whether the proportion of seeds which germinate is lower.	E1	**Examiner's Comments** Most candidates wrote down th correct notation. It is encouragi candidates gave a correct defi previous years.
		iv	Let X ~ B(20, 0.85)		
		iv	P(X ≤ 13) = 0.0219	M1*	For probability (provided not as
		iv			P(X = 13)) Ignore notation
		iv	0.0219 > 1%	M1*dep	For comparison
		iv	So not enough evidence to reject H_0.	A1*	For not significant oe
		iv	Not significant.		
		iv	Conclude that there is not enough evidence to indicate that the proportion of seeds which have germinated has decreased.	E1*dep	For conclusion in context Must mention decrease, not jus
		iv	ALTERNATIVE METHOD – follow method above unless some mention of CR seen Critical region method LOWER TAIL		
		iv	P(X ≤ 13) = 0.0219 > 1%	M1	For either probability
		iv	P(X ≤ 12) = 0.0059 < 1%		
		iv	So critical region is {0,1,2,3,4,5,6,7,8,9,10,11,12}	A1	cao dep on at least one correct
		iv	13 not in CR so not significant	A1*	
		iv	There is insufficient evidence to indicate that the proportion of seeds which have germinated has decreased.	E1*dep	**Examiner's Comments** Those candidates who calculat more successful than those usi

					Those who used the latter meth wrong, thereby losing credit. In more clearly than in previous s context. There was also rather than in the past.
		v	33 < 35	M1	For comparison
		v	So there is sufficient evidence to reject H_0	A1*	
		v			
		v	Conclude that there is enough evidence to indicate that the proportion of seeds which have germinated has decreased.	E1*dep	For conclusion in context Must mention decrease, not jus **Examiner's Comments**
		v			Many candidates, despite havi correctly, reverted to point prob calculator to find P(X = 33). Th Others made a correct compar always sure what this meant in
		vi	For $n = 3$, P($X \leq 0$) = 0.0034 < 0.01	M1	For P($X \leq 0$) = 0.0034
		vi	For $n = 2$, P($X \leq 0$) = 0.0225 > 0.01	M1	For P($X \leq 0$) = 0.0225
		vi	So the least value of n for which the critical region is not empty and thus H_0 could be rejected is 3.	A1	CAO
		vi			
		vi	ALTERNATIVE METHOD using logs		
		vi	$0.15^n < 0.01$	M1	
		vi	$n >$ log 0.01 / log 0.15	M1	
		vi	$n > 2.427$		
		vi	Least $n = 3$	A1	**Examiner's Comments** Most candidates who knew hov down 'for n = 3, P(X = 0) = 0.0 not then justify their answer by and thus only gained one mark successful attempts using loga
			Total	**19**	

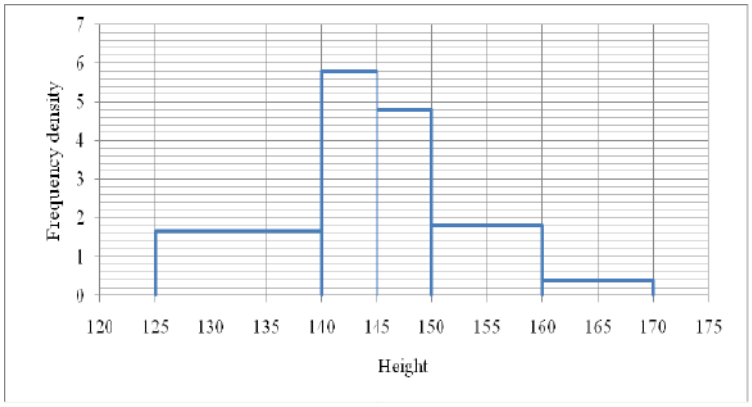
16		i	$P(L \cap W) = P(L \mid W) \times P(W) = 0.4 \times 0.07 = 0.028$	M1	For $P(L \mid W) \times P(W)$ cao
		i		A1	**Examiner's Comments** This question was well answer candidates scoring both marks
		ii		B1	For two labelled intersecting ci
		ii		B1	For at least 2 correct probabilit For remaining probabilities
		ii		B1	**Examiner's Comments** Almost all candidates gained th intersecting circles. Many cand part (i) into the intersection but from P(W) so put 0.07 instead the other part of the circle labe number of candidates gained f correct 0.01 and 0.92 in the oth correctly.
		iii	$P(L \cap W) = 0.028$, $P(L) \times P(W) = 0.038 \times 0.07 = 0.00266$	M1	For correct use of $P(L) \times P(W)$ If P(L) wrong, max M1A0E0. No marks if P(W) wrong
		iii		A1	For 0.00266 Allow 'they are dependent'
		iii	Not equal so not independent	E1* dep on	Do not award E1 if $P(L \cap W)$ wrong
		iii		M1	**Examiner's Comments** The vast majority of candidates independence by comparing P However most of these did not and many had P(W) wrong, de question. A small number of ca P(L) and these were more ofte
			Total	**8**	

17		i	Positive	B1	CAO **Examiner's Comments** Approximately 95% of candida
		ii	Mean = 5.064 allow 5.1 with working 126.6/25 or 5.06 without	B1	Allow B1 for RMSD = 1.297 or
		ii	SD = 1.324 allow 1.3 with working or 1.32 without	B2	**Examiner's Comments** There were many fully correct a formulas rather than using the calculators. A few candidates f and these gained a method ma not to use the key and thus get This error was severely penalis for a follow through.
		iii	$\bar{x} - 2s = 5.064 - 2 \times 1.324 = 2.416$	B1FT	FT their mean and sd
		iii	$\bar{x} + 2s = 5.064 + 2 \times 1.324 = 7.712$	M1	for $\bar{x} + 2s$ but withhold final E mar are no outliers.
		iii		A1FT	For upper limit Incorrect statement such as 7.6 Do not award E1 if calculation
		iii	So there is an outlier.	E1	**Examiner's Comments** The limits for outliers were wid most candidates. Even those w deviation were able to gain 3 o through correctly. Some candid despite often having got part (i made errors, losing some if not
			Total	**8**	
18		i	4 + ½ of 18 = 4 + 9 = 13	M1	For ½ of 18 cao
		i		A1	**Examiner's Comments** On the whole, this question wa common incorrect answer was frequently. A small minority of c calculated incorrectly (both 11

		ii	(Median) = 50.5th value	M1	For 50.5 seen
		ii	$\text{Est} = 140 + \left(\frac{25.5}{29}\right)\times 5$ or $= 140 + \left(\frac{50.5-25}{54-25}\right)\times 5$	M1	For attempt to find this value
					Examiner's Comments
		ii	= 144.4	A1	Only about 10% of candidates answers. Many scored SC2 for 50.5th, value. Those that did sta 50.5th value often just gave the interpolation. Many candidates specification.
		iii		M1	For fd's - at least 3 correct

**

Height	Frequency	Group width	Frequency density
$125 \le x \le 140$	25	15	1.67
$140 < x \le 145$	29	5	5.80
$145 < x \le 150$	24	5	4.80
$150 < x \le 160$	18	10	1.80
$160 < x \le 170$	4	10	0.40

		iii		A1	Accept any suitable unit for fd s to at least one dp allow 1.66 bu
		iii		G1	linear scales on both axes and
		iii		W1	width of bars
					height of bars
					Examiner's Comments
		iii		H1	The histogram was generally c previous years. Most candidate frequency densities correctly, a axes correctly. A fairly commor

					frequency density down to 1.6 errors with careless drawing of incorrect heights.
		iv	4 boys	3	
		iv	0.6 × 15	M1	For 0.6 × 15
		iv	= 9 girls	A1	For 9 girls cao
		iv	So 5 more girls	A1	**Examiner's Comments** Roughly 90% of candidates sc
		v	Frequencies and midpoints for girls are Height 132.5 142.5 147.5 155 167.5 Frequency 18 23 31 19 9	B1	For at least three frequencies c
		v		B1	At least three midpoints correc
		v	So mean =	M1	For attempt at Σ*xf*
		v	$\frac{(132.5\times18)+(142.5\times23)+(147.5\times31)+(155\times19)+(167.5\times9)}{100}$ $=\frac{(2385)+(3277.5)+(4572.5)+(2945)+(1507.5)}{100}$	M1* Dep on M1	For division by 100 Cao NB Watch for over-specificatio
		v		A1	**Examiner's Comments** Many candidates did everythin answer as 146.875 or 146.88 t over-specifying. The scheme a frequencies and the mid-points to gain 4 marks. The most com mid-point as 165 rather than 16
			Total	**18**	
19		i	$\left(\frac{5}{6}\right)^2\times\frac{1}{6}=\frac{25}{216}(=0.116)$	M1	For 5/6 (or 1 – 1/6) seen
		i		M1	For whole product cao
		i		A1	**Examiner's Comments** This was well answered with th marks. Some calculated $(1/6)^3$ scored 1 for ⅝.

		ii	$1-\left(\frac{5}{6}\right)^{10}=1-0.1615=0.8385$	M1	For $(5/6)^{10}$(without extra terms)
					cao
					Examiner's Comments
		ii		A1	Only about 25% of candidates either misread or misinterprete the probability of needing exac candidates read the question c calculating all the probabilities usually successfully. Very few mark scheme.
			Total	**5**	
20		i	$X \sim$ B(10, 0.35)	M1	or 0.35^5 x 0.65^5
		i	$P(5 \text{ accessing internet}) = \binom{10}{5} \times 0.35^5 \times 0.65^5$	M1	For $\binom{10}{5} \times p^5 \times q^5$
		i	= 0.1536	A1	cao
		i	**OR**		
		i	from tables = 0.9051 – 0.7515 = 0.1536	M2	For 0.9051 – 0.7515
		i		A1	cao
		i	$P(X \geq 5) = 1 - P(X \leq 4)$		
		i	= 1 – 0.7515	M1	For 0.7515
					cao
					Examiner's Comments
		i	= 0.2485	A1	Again many fully correct respo usually used the correct table b was 1 - P($X \leq$ 5) rather than 1 - the lengthy method of finding t more and then adding, sometir cases with errors.
		i	$E(X) = np = 10 \times 0.35$	M1	For 10 × 0.35
		i	= 3.5	A1	cao
					Examiner's Comments

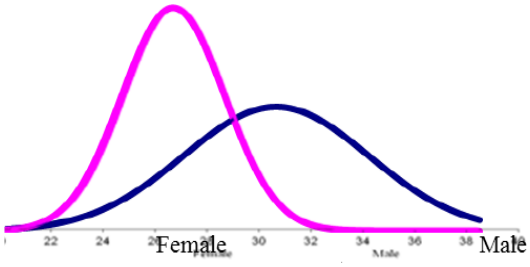
					Almost all candidates multiplie them either rounded to 4 or tru second mark.
		ii	Let $X \sim B(20, 0.35)$	9	
		ii	Let p = probability of a customer using the internet (for population)	B1	For definition of p in context
		ii	H_0: $p = 0.35$	B1	For H_0
		ii	H_1: $p \neq 0.35$	B1	For H_1
		ii	H_1 has this form because the test is to investigate whether the proportion is different, (rather than lower or higher).	E1	
		ii	$P(X \geq 10)$	B1	For notation $P(X \geq 10)$ or $P(X >$ no incorrect notation)
		ii	$= 1 - 0.8782 = 0.1218$	B1*	For 0.1218 Allow 0.12
		ii	$> 2.5\%$	M1* dep	For comparison with 2.5%
		ii	So not significant.	A1*	

		ii	Conclude that there is not enough evidence to indicate that the probability is different. (Must state 'probability', not just 'p')	E1* dep on A1	
		ii	ALTERNATIVE METHOD FOR FINAL 5 MARKS		
		ii	Critical region method		
		ii	LOWER TAIL $P(X \leq 2) = 0.0121 < 2.5\%$ $P(X \leq 3) = 0.0444 > 2.5\%$	B1	For either probability
		ii	UPPER TAIL $P(X \geq 11) = 1 - P(X \leq 10) = 1 - 0.9468 = 0.0532 > 2.5\%$ $P(X \geq 12) = 1 - P(X \leq 11) = 1 - 0.9804 = 0.0196 < 2.5\%$	B1	For either probability
		ii	So critical region is {0, 1, 2, 12, 13, 14, 15, 16, 17, 18, 19, 20}	M1*	For either probability cao dep c comparison with 2.5% **Examiner's Comments** Most candidates were able to i test and were able to correctly hypotheses. However, some ca others failed to explain why it w weaker candidates used poor n hypotheses. Rather more cand method than finding $P(X \geq 10)$. H probability method were genera tried to find the critical region o so lost the final three marks. U number of candidates made co 2.5%, or omitted the comparisc the last three marks. A disappo thus losing all of the final 5 ma the majority of candidates did h with probabilities, is needed wh Conclusions, for those who get However care should be taken findings including an *element o* Those answering by the critica aware that '10 is not in CR' is n add 'insufficient evidence to rej go on to give an answer in con
		ii	So not significant	A1*	
		ii	Conclude that there is not enough evidence to indicate that the probability is different.	E1* dep on A1	

		iii	0.0022 < 2.5%	2	
		iii	So reject H_0, Significant.	B1	For either reject H_0 or significa
					Dep on good attempt at correc
					Examiner's Comments
		iii	Conclude that there is enough evidence to indicate that the probability is different.	E1* dep	Under half of the candidature s question. Many did not attempt compared with 5% even thoug with 2.5% in part (ii). A further correctly state their conclusion question.
			Total	**18**	
21		i	$P(X > 135) = P\left(Z > \frac{135-130.5}{\sqrt{11.84}}\right)$	M1	For standardising. Penalise use of "continuity cor
		i	= P(Z> 1.308) = 1 − Φ(1.308) = 1 − 0.9045	M1	For correct structure i.e. finding the area to the right
					CAO inc use of diff tables Allow 0.0954 and 0.0956 If numerator reversed, give BO
		i	= 0.0955 (3s.f.)	A1	**Examiner's Comments**
					Was found to be easy, though exact and approximate values thereby losing a mark.
		ii	From tables Φ^{-1} (0.99) = 2.326	B1	±2.326 or (better) seen, not ±2
					For sensible equation in k with Note that use of z = 0.8389 fro 0.8389 is clearly a **probability**
		ii	$\frac{k-130.5}{\sqrt{11.84}} = 2.326$ oe	M1	Allow use of −2.326 (or their ne reversed. Condone use of σ = 11.84 if als Condone use of *"k* ± 0.5" for *k* 0/3 for trial and improvement
					CAO Allow 138.504. Accept 138.5 Do not accept final answers of
		ii	$k = 130.5 + \left(2.326 \times \sqrt{11.84}\right) = 138.50$	A1	**Examiner's Comments**
					Was found hard. Very few can and cubing gave double and tri

		iii	$\text{P(Wing length} = 131) = \text{P}\left(\frac{130.5-130.5}{\sqrt{11.84}} \le Z \le \frac{131.5-130.5}{\sqrt{11.84}}\right)$	B1	For both limits correct, soi. e.g. use of 0.5 in probability ca limit.
		iii	= P(0 < *Z* < 0.2906) = Φ(0. 2906) − Φ(0) = 0.6143 − 0.5	M1	For **correct structure** using th **Finding the area between the 130.5** Condone use of σ = 11.84 if als
		iii	= 0.1143	A1	CAO inc use of diff tables Allow 0.1145 Allow 0.114 www
		iv	H_0: μ = 130.5 H_1: μ > 130.5	B1	For both correct Hypotheses in words must refe Do not allow other symbols unl population mean
		iv	Where μ denotes the **mean wing length** (in the population) (of Scandinavian male blackbirds).	B1	For definition of *μ* in context. Do not allow "sample mean wir or "mean wing length of Englis
		iv	$\text{Test statistic} = \frac{132.4-130.5}{\sqrt{11.84}/\sqrt{20}} = \frac{1.90}{0.7694}$	M1*	must include $\sqrt{20}$ Condone use of σ = 11.84 if als part (iii). Condone numerator reversed f (max 4/8)
		iv	= 2.469	A1	Allow 2.47
		iv	Upper 5% level 1 tailed critical value of *z* = 1.645	B1	For 1.645. Must be positive. B0 if −1.645 seen. No further A marks from here i
		iv	2.469 > 1.645	M1dep*	For sensible comparison leadir
		iv	The result is significant. There is sufficient evidence to reject H_0	A1	For correct conclusion. e.g. for FT only candidate's test statisti
		iv	There is sufficient evidence to **suggest** that the mean wing length (of this population of birds) is greater (than 130.5 mm).	A1	For **non-assertive** conclusion result Condone use of "average" for " FT only candidate's test statisti
		v	With a 10% significance level rather than a 5% significance level, Advantage: One is less likely to accept the null hypothesis when it is false.	E1	Accept equivalent wording.
		v	Disadvantage: One is more likely to reject the null hypothesis when it is true.	E1	Note — Unless stated otherwis relates to an advantage and th disadvantage.
			Total	**19**	

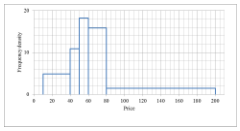
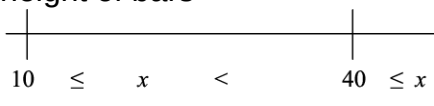
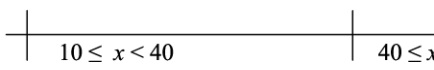
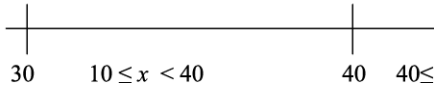
22		a	0.758 > 0.279 So there is sufficient evidence of correlation (in the population)	**M1(AO1.1)** **A1(AO2.2b)** **[2]**	Oe but not evidence of positive correlation.
		b	E.g. diagram shows positive correlation overall, but the data consists of two distinct clusters. E.g. neither of the two clusters show evidence of correlation	**B1(AO2.3)** **B1(AO2.2b)** **[2]**	Accept other suitable correct comments
			Total	**4**	
23		i	*(A)* P(X < 30)		
		i	$= \mathrm{P}\left(Z < \frac{30-30.7}{3.5}\right)$	M1	For standardising
		i	P(Z < −0.20)		
		i	= Φ(−0.20)		
		i	= 1 − Φ(0.20) = (1 − 0.5793)	M1	For correct structure CAO
		i	= 0.4207	A1	**Examiner's Comments** The majority of candidates corr value then found the correct pr made good use of diagrams to
		i	*(B)* P(25 < X < 35)		
		i	$= \mathrm{P}\left(\frac{25-30.7}{3.5} < Z < \frac{35-30.7}{3.5}\right)$	M1	Correctly standardising both.
		i	= P(−1.629 < X < 1.229) = Φ(1.229) − Φ(−1.629)		
		i	= 0.8904 − (1 − 0.9483) = 0.8904 − 0.0517	M1	For correct structure Use of differences column requ
		i	= 0.8387	A1	**Examiner's Comments** For a question that had potenti standardising, approximating to incorrectly, this question was a general done rather better than

		ii	P(all 5 weigh at least 30kg)		
		ii	= 0.5793^5	M1	Allow FT (1 − their (i)(*A*))^5 or [their P($X \geq 30$)]^5 FT only (1 − their (i)(*A*))^5 **Examiner's Comments**
		ii	= 0.0652	A1	The binomial situation, within a Normal distribution, was dealt The errors that did occur were found p(X < 30) whereas this p weigh *more than* 30 kg.
		iii	P(weight > 30) = 0.05		
		iii	$P(Z > \frac{30-26.8}{\sigma}) = 0.05$		
		iii	Φ^{-1} (0.95) = 1.645	B1	For 1.645. B0 for 1 - 1.645 or 0.1645
		iii	$\frac{30-26.8}{\sigma} = 1.645$	M1*	For equation as seen or equiva
		iii	$\sigma = \frac{30-26.8}{1.645} = 1.945$ kg	M1dep*	Rearranging for σ
		iii		A1	CAO **Examiner's Comments** Few candidates used the wron correctly identifying that +1.645 use as 5% weighed *more than*
		iv		G1	For two Normal shapes includi behaviour with horizontal axis a
		iv		G1	For means, shown explicitly or

		iv		G1	For lower max height for Male For visibly greater width for Ma **Examiner's Comments**
		iv		G1	This part of the question tested found dealing with the correct s constraints on the curves difficu with interpretation of the requir
			Total	**16**	
24		i	Number of ways = 5! = 120	B1	**Examiner's Comments** This was generally well answer
		ii	Probability = 2/120	M1	For division by their 120 CAO **Examiner's Comments**
		ii	= 1/60 or 0.0167 or 0.01 $\dot{6}$	A1	Again this was generally well a candidates truncated their deci rounding. The use of fractions
		iii	$\frac{1}{5}\times\frac{1}{4}\times\frac{1}{3}=\frac{1}{60}$ or 0.0167 or 0.01 $\dot{6}$	B1	Condone 2/120 for B1 **Examiner's Comments** Only around two thirds of cand The most common error was to
		iv	$\frac{3}{5}\times\frac{2}{4}\times\frac{1}{3}=\frac{1}{10}$	M1	For 3/5 ×
		iv		A1	CAO For division by $^5C_3$ **Examiner's Comments**
		iv	or $1/^5C_3 = 1/10$	M1	Again about two thirds of candi many scoring the marks for the not, multiplied by 6, rather than $^1/_3$.
		iv		A1	
			Total	**6**	
25		i	P(Does not lose any match) = 0.83 = 0.512 = $^{64}/_{125}$	B1	**Examiner's Comments** Many candidates thought that r hence gave the common wrong

					tried to consider combinations success. The fact that the ques should have been a clue to the approach.
		ii	P(Wins all 3 or draws all 3 or loses all 3) = $0.5^3 + 0.3^3 + 0.2^3$	M1	Including addition **Examiner's Comments** This was generally well answe interpreted this as 'find three s did, and listed them but with no
		ii	$= 0.16 = {}^4/_{25}$	A1	
		iii	P(all three outcomes occur) = $3! \times 0.5 \times 0.3 \times 0.2$	M1*	
		iii	$= 0.18$	A1	
		iii	Required probability = 1 − '0.18' − 0.16	*M1 dep	
		iii	$= 0.66 = {}^{33}/_{50}$	A1	CAO
		iii	**OR:**		
		iii	P(WWW') + P(DDD') + P(LLL')	M1	For any one product (no need f
		iii	$3 \times 0.5^2 \times 0.5 + 3 \times 0.3^2 \times 0.7 + 3 \times 0.2^2 \times 0.8$	M1	For '3 ×'
		iii	0.375 + 0.189 + 0.096	M1	For sum of three correct terms And no incorrect terms
		iii	0.66	A1	CAO
		iii	**OR:**		
		iii	P(WWD) + P(WWL) + P(DDW) + P(DDL) + P(LLW) + P(LLD)	M1	For any one product (no need f
		iii	$3\times0.5^2 \times 0.3 + 3 \times 0.5^2 \times 0.2 + 3 \times 0.3^2 \times 0.5 + 3 \times 0.3^2 \times 0.2 + 3 \times 0.2^2 \times 0.5 + 3 \times 0.2^2 \times 0.3$	M1	For '3 ×'
		iii	0.225 + 0.15 + 0.135 + 0.054 + 0.06 + 0.036	M1	For sum of six correct terms (n And no incorrect terms
		iii	0.66	A1	CAO **Examiner's Comments** Only a small proportion of cand approach (the first method in th who did, many forgot to multipl lists or tree diagrams to show F not multiply by 3, so the most c rather than the correct 0.66.
			Total	**7**	

26		i	Mean = $\frac{(25\times147)+(45\times109)+(55\times182)+(70\times317)+(140\times175)}{930}$ $\frac{750\times7+1250\times22+1750\times26+2500\times18+4000\times7}{80}$ $=\frac{151250}{80}$ = (£)70.19 or (£)70.2	M1	For midpoints (at least 3 correc (allow 25.005, 45.005 etc leadi CAO (exact answer 70.19355…) Correct answers obtained from functions gain full marks
		i		A1	Condone answer of (£)70.20
		i	$\Sigma x^2 f$ = $(25^2 \times 147) + (45^2 \times 109) + (55^2 \times 182) + (70^2 \times 317) + (140^2 \times 175)$ = 91875 + 220725 + 550550 + 1553300 + 3430000 = 5846450 $S_{xx} = 5846450 - \frac{65280^2}{930}$ = 1264215.161 or $5846450 - 930 \times 70.19^2$	M1	For attempt at S_{xx} Should include sum of at least Do not FT their incorrect mean
		i	$s = \sqrt{\frac{1264215}{929}} = \sqrt{1360.83}$ = 36.89 or (£)36.9 Allow any answer between 36.87 and 36.90 without checking working	A1	(exact answer 36.88949…) Condone answer of (£)36.90 If both mean and sd overspecif **Examiner's Comments** This part was fairly well answe gaining full credit. A few had no used correct midpoints, althoug occasionally used figures such deviation proved more difficult a variety of wrong methods see functions on their calculator to being the recommended metho specified either or both of their mark.
		ii	100/120 × 175 = 145.83	M1*	For 175/120 **Examiner's Comments** Candidates found this part rath almost half scored full marks. T they were after was the bigges some were then unsure what to once they had found it. Some r the most common mistake of th

					needed to do) and others failed percentage, just giving the fina
		ii	145.83/930 = 0.1568	*M1dep	
		ii	So 15.7%	A1	
		iii	<table><tr><th>Price</th><th>Frequency</th><th>Group width</th><th>Frequency density</th></tr><tr><td>$10 \le x \le 40$</td><td>147</td><td>30</td><td>4.90</td></tr><tr><td>$40 < x \le 50$</td><td>109</td><td>10</td><td>10.90</td></tr><tr><td>$50 < x \le 60$</td><td>182</td><td>10</td><td>18.20</td></tr><tr><td>$60 < x \le 80$</td><td>317</td><td>20</td><td>15.85</td></tr><tr><td>$80 < x \le 200$</td><td>175</td><td>120</td><td>1.46</td></tr></table>	M1	For fds - at least 3 correct Accept any suitable unit for fd s
					Allow 15.9 and 1.5 and condon
					Examiner's Comments
		iii		A1	This part was again well answe candidates gaining at least 4 m were seen, but none very comm were: using frequency rather th non-linear scale on one of the a axis), stopping the horizontal a horizontal axis 'Class width'.
		iii		B1	linear scales on both axes and horizontal axis labelled *x*) Vertical scale starting from zer mark for heights if broken)
		iii		B1	width of bars (within half a squa (NO GAPS ALLOWED)
		iii	NB If not using fd's only mark available is B1 for width of bars		
		iii	Heights must be within ≤ 1 square of overlay (only for scales 2cm = 4 units (blue) or 5 units (red)) – otherwise check heights. Note that you must make sure that the overlay is aligned correctly with the vertical axis.	B1	height of bars   **BUT** 

					Examiner's Comments This part was again well answe candidates gaining at least 4 m were seen, but none very comm were: using frequency rather th non-linear scale on one of the a axis), stopping the horizontal a horizontal axis 'Class width'.
		iv	Positive skewness	B1	Allow +ve **Examiner's Comments** Over 90% of candidates scored
		v	Area for men from 100 to 200 = 100 × 2 = 200 200/990 = 0.202	M1	
		v	So 20.2%	A1	
		v	Cannot be certain as both figures are estimates	E1	Independent **Examiner's Comments** A good number of candidates a question was answered better a similar calculation. Of those v most started with 240/990 or 2 explanation over certainty was candidates achieving this mark first 2 marks.
		vi	Men's running shoes have a lower average price than women's (as their mean is only £68.83 compared to £70.19). Or equivalent for women	E1	FT their mean Do NOT condone lower centra FT their SD **Examiner's Comments** Although this is essentially a si candidates scored zero. Candi acceptable comparisons, with r "central tendency" when compa few discussing averages. A mo candidates were able to provid differences in the standard dev central tendency was somethin of candidates were unable to c grammatically correct sentence
		vi	Men's running shoes have a more variation in price than women's (as their sd is £42.93 compared to £36.89). Or equivalent for women	E1	
			Total	**18**	

27		a	**(i)** the cumulative frequencies have been plotted against the mid-points of the class intervals, mis-plotting [at centre of each class] reduces estimate (by 2.5) oe	**B1 (AO 2.4)** **B1 (AO 2.4)** **[2]**	**Examiner's Comments** Candidates who did well in this cumulative frequencies had be intervals instead of at the uppe
			(ii) grouped data has been used grouping has slightly reduced the error introduced by misplotting (because the error is less than 2.5)	**B1 (AO 2.4)** **B1 (AO 2.4)** **[2]**	or eg Hodge used the graph (instead of the raw data) **Examiner's Comments** Candidates understood that gr accuracy of the result and com Candidates who did less well m the points had been joined by s
		b	percentage unemployment is often estimated oe	**E1 (AO 2.4)** **[1]**	allow data (on percentage unemployment) is not available for all countries **in Europe** oe **Examiner's Comments** Candidates who did less well b geographical or economic idea to issues related to the estimati
		c	there are many other countries in the pre-release material; it is very unlikely that a random sample would only include European countries.	**E1 (AO 2.4)** **[1]**	**Examiner's Comments** Candidates who did well on pa with the pre-release material m
		d	negative correlation / association (may be embedded) comparison of *p*-value with 0.05 or 0.01 or other appropriate significance level and supporting comment	**B1 (AO 2.2b)** **B1 (AO 2.2b)**	if **B0B0** allow **SC2** for eg comment on no significant association justified by

				[2]	comparison of *p*-value with appropriate significance level (eg 0.025) **Examiner's Comments** Candidates who did well comm correlation. They compared the level and then made an approp did less well compared the corr value or did not comment on th
		e	(even though this is interpolation), the scatter / weak correlation / presence of an outlier would suggest that the use of of a line of best fit is inappropriate	**E1 (AO 2.2b)** **[2]**	allow explanation based on the value for Kosovo being an outlier or on it lying in the (large) gap in the scatter **Examiner's Comments** Candidates who did well comm scatter or the position of 35.3 r justify their comment.
			Total	**9**	
28		a	$1 - 0.51^2$ $= 0.7399$	**M1(AO3.1b)** **A1(AO1.1)** **[2]**	Accept 0.74 or 0.740
		b	$1 - 0.49^2 - 0.38^2 - 0.1^2 - 0.03^2$ $= 0.6046$	**M1(AO3.1b)** **M1(AO1.1)** **A1(AO1.1)**	For squaring probabilities OR products of pairs For complementary event OR doubling products of pairs $2 \times (0.1862 + 0.049 + 0.0147 + 0.038 + 0.0114 + 0.003)$

				[3]	
			Total	**5**	
29		a	Allocate numbers 001 to 200 to the trees Choose 10 (3 digit) random numbers	**B1(AO1.2)** **B1(AO2.4)** **[2]**	e.g. use calculator to get 10 different random numbers
		b	Mean = 27.61kg SD = 4.04 kg (3sf)	**B1(AO1.1)** **B1(AO1.1)** **[2]**	**BC** **BC**
		c	Upper limit = 27.61 + 2 × 4.04 = 35.69 So the value of 38.1 is an outlier This value should be investigated to check if it is genuine. If so, it should not be removed from the data	**M1(AO1.1)** **A1(AO1.1)** **B1(AO2.2b)** **[3]**	For mean + 2 × sd OR UQ + 1.5 IQR = 28.3 + 1.5 × 3.2 = 33.1 OR e.g. If the value is not representative of the other 199 trees because e.g. this tree is a different type it should be ignored
			Total	**7**	
30		a	Symmetrical with one possible outlier	**B1(AO1.2)** **[1]**	or negative skew
		b	24^{th} value − 8^{th} value 27 – 16 = 11	**M1(AO1.1)** **A1(AO1.1)** **[2]**	
		c	16 − 1.5 × 11 = −0.5 3 > − 0.5 so it is not an outlier.	**M1(AO1.1)** **A1(AO2.2a)**	Check for outliers using their Q_1 − 1.5 × IQR

				[2]	
			Total	**5**	
31		a	**(a) i** Mean for world = 7 174 654 290 ÷ 239 = 30 million (approx.) Mean for sample is 9.36 million so much smaller	**M1(AO 1.1)** **A1(AO 1.1)** **B1(AO 1.1)** **[3]**	Any degree of accuracy
		b	**(a) ii** There are a lot of small countries in the world. Therefore there is no reason to suppose the sample is not random,	**B1(AO 2.4)** **E1(AO 2.2b)** **[2]**	
		c	**(b)** 0.091 × 55 400 =[$] 5041.40	**M1(AO 3.3)** **A1(AO 1.1)** **[2]**	Condone missing units. Answer can be rounded to 5041 or 5040
		d	**(c)** i Identify physicians per 1000 as highest R^2 with positive correlation and positive correlation	**B1(AO 2.2b)** **E1(AO 2.4)** **[2]**	NB graph C e.g. Reject D as it shows negative correlation
		e	**(c)** ii There is hardly any correlation.	**E1(AO 3.5b)** **[1]**	
			Total	**10**	
32		i	Mean $= \frac{17100}{50} = 342$ $Sxx = 6115108 - \frac{17100^2}{50} = 266908$ $s = \sqrt{\frac{266908}{49}} = \sqrt{5447.10} = 73.8 \ (73.8044\ldots)$	**B1** **M1**	Ignore units CAO For S*xx* M1 for 6115108 − 50 x their me

				A1 **[3]**	BUT NOTE M0 if their $S_{xx} < 0$ CAO ignore units M1A0 for RMSD = 73.1 (73.062 **Examiner's Comments** Nearly all candidates worked o candidates also found the stan specified the answer thus losin candidates made an error in th or worked out the RMSD. It wa candidates recalling and using
		ii	New mean = (0.108 × 342) + 7.2 = £44.14 New sd = 0.108 × 73.8 = £7.97 Using RMSD gives £7.89 Using variance gives 588.29	**B1** **M1** **A1** **[3]**	FT their mean Allow £44.1 or b FT their sd (unless negative)fo NB If candidate 'starts again' o CAO Do not penalise lack of units in Deduct at most 1 mark overall specification of either mean or **Examiner's Comments** Most candidates used the trans mean. Many candidates also o deviation, with a pleasingly sm 7.2 to their final answer. Some specification, although there w the whole question for this erro most candidates giving answer accuracy.
			Total	**6**	
33	a		Mass = 23.56×1.816^2 =77.7 [kg]	M1 (AO1.1) A1 (AO1.1) [2]	Allow M1 for cm used here CAO **Examiner's Comments** Most candidates were able to f some did not realise the height others used the height in cm ra part (b) was generally describe description such as 'as age inc was usually done correctly, but would require extrapolation, wh candidates did correctly explain an age of 45, and so the line of while others gave an incomplet at (or about) age 60 meant the used.

					For part (e), a good number of that the sample data included r ages used was restricted. Ther answers to part (f); some candi that every *n*th value should be but none explained the need fo
	b		Positive [correlation]	B1 (AO1.1) [1]	Ignore, eg, weak or strong
	c		28 ≤ BMI < 30	B1 (AO3.4) [1]	
	d		extrapolation	B1 (AO3.5b) [1]	Or equivalent
	e		eg females in population wider age range in population scatter diagram for whole population shows very weak negative correlation	B1 (AO2.2b) B1 (AO2.2b) [2]	any two distinct comments
	f		Eg Generate a random number, *n*, between 1 and (eg 20) inclusive and select the *n*th item in the data set OR select every *m*th (eg 16th value). For valid solution.	M1 (AO1.2) A1 (AO1.1) [2]	Random no between 1 and 9 and every 17th item would also work.
			Total	**9**	
34			$\frac{327}{600} \times 50$ 27 cao	**M1(AO1.1b)** **A1(AO1.1b)** **[2]**	may be implied by 27.25
			Total	**2**	
35		a	Frequency density 4.6 Number of runners 21	**B1(AO2.2a)** **B1(AO1.1)** **[2]**	

					Examiner's Comments This part was usually done cor omitted one or both answers. M were sometimes seen instead
		b	Vertical axis should be labelled "number of runners per minute"	**E1(AO2.3)** **[1]**	Or frequency density **Examiner's Comments** The most common wrong answ axis should be labelled 'frequen
			Total	**3**	
36		a	the pattern of data points suggest a straight line oe the value of *r* indicates that the fit is close oe	**E1 (AO3.3)** **E1 (AO2.4)** **[2]**	
		b	70	**B1 (AO3.4)** **[1]**	from 0.89×83 – 3.76 (= 70.11)
		c	(10 + 3.76)÷ 0.89 13	**M1 (AO3.4)** **A1 (AO1.1)** **[2]**	(= 13.21 …)
		d	The approximation for Tina's mark is obtained by interpolation, whereas the approximation for Dave's mark is from extrapolation	**E1** **(AO3.5b)** **[1]**	allow eg should use the equation for *x* on y to estimate Dave's mark
			Total	**6**	

Mark scheme