Edexcel GCSE

Statistics

Paper 1H

Higher Tier

Friday 25 June 2004 – Morning

Time: 2 hours 30 minutes

Materials required for examination	**Items included with question papers**
Ruler graduated in centimetres and millimetres, protractor, compass, pen, HB pencil, eraser, electronic calculator.	Nil

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Answer **all** questions in the spaces provided in this question paper. **You must NOT write on the formulae page or any blank pages. Anything you write on these pages will gain NO credit.**
If you need more space to complete your answer to any question, use additional sheets.

Information for Candidates

The total mark for this paper is 100.
The marks for the various parts of questions are shown in round brackets: e.g. (**2**).
This question paper has eight questions in Section A and seven questions in Section B.

Advice to Candidates

Work steadily through the paper.
Do not spend too long on one question.
Show all stages in any calculations.
If you cannot answer a question, leave it and attempt the next one.
Return at the end to those you have left out.

Turn over

Edexcel
Success through qualifications

GCSE Statistics 1389

Higer Tier Formulae

You must not write on this page.
Anything you write on this page will gain NO credit.

Mean of a frequency distribution $\qquad = \dfrac{\sum fx}{\sum f}.$

Mean of a grouped frequency distribution $\qquad = \dfrac{\sum fx}{\sum f}$, where x is the mid-interval value.

Variance $\qquad = \dfrac{\sum (x - \overline{x})^2}{n}$

Standard deviation (set of numbers) $\qquad \sqrt{\left[\dfrac{\sum x^2}{n} - \left(\dfrac{\sum x}{n} \right)^2 \right]}$

or $\qquad \sqrt{\left[\dfrac{\sum (x - \overline{x})^2}{n} \right]}$

where $\overline{x}$ is the mean set of values.

Standard deviation (discrete frequency distribution) $\qquad \sqrt{\left[\dfrac{\sum fx^2}{\sum f} - \left(\dfrac{\sum fx}{\sum f} \right)^2 \right]}$

or $\qquad \sqrt{\left[\dfrac{\sum f(x - \overline{x})^2}{\sum f} \right]}$
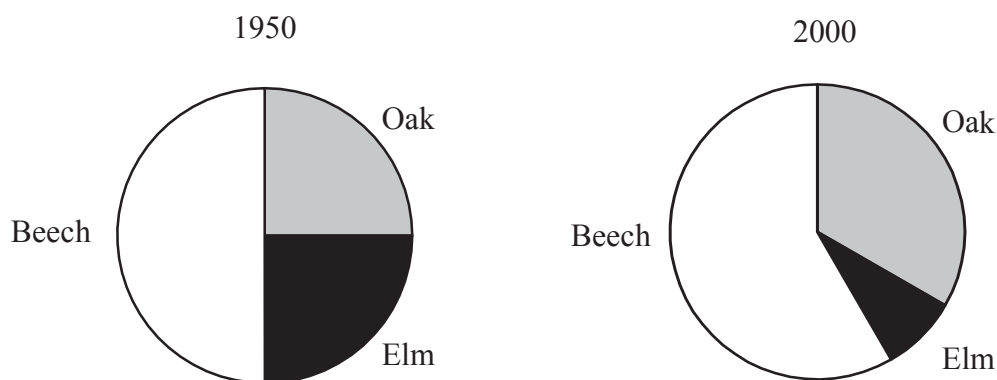
Spearman's rank correlation coefficient $\qquad 1 - \dfrac{6 \sum d^2}{n(n^2 - 1)}$

You must write down all stages of your working.

SECTION A

Answer ALL EIGHT questions.

1. The proportions of oak, elm and beech trees in Beckham's Wood in 1950 and 2000 are shown in the pie charts below.

Between 1950 and 2000 many of the elm trees in Britain were killed by Dutch elm disease.

(a) Write down one feature in the pie charts that suggests that Beckham's Wood could be in Britain.

....

....

(1)

(b) Emma says

"These pie charts show that the number of beech trees in Beckham's Wood has increased between 1950 and 2000."

Emma is **wrong**.

Explain why.

....

....

(1)

Do not write here

Page Total

2. A researcher wants to choose a random sample of 300 people from her town.

She considers choosing the people that live near her house.

(a) Write down one advantage and one disadvantage in this method of choosing her sample.

Advantage ....

....

Disadvantage ....

....

(2)

There are 30 000 people on the electoral register in her town.

(b) Describe how she should choose a systematic sample of 300 people from the electoral register.

....

....

....

(2)

Do not write here

3. The table below is part of a larger table that shows the examination achievements by pupils in 2001/ 2002.

Examination Achievements 2001/ 2002

| | Pupils in the last year of compulsory education | | | | |
| | Percentage achieving GCSE | | | | |
	5 or more grades A*-C	1-4 grades A*-C	Grades D-G only	No graded GCSEs	Total (=100%) (thousands)
United Kingdom	51.0	24.1	19.4	5.5	729.7
North East	43.9	24.8	24.8	6.5	33.3
North West	48.0	24.5	21.7	5.7	89.6
Yorkshire and the Humber	44.4	24.3	25.2	6.2	63.0
East Midlands	49.1	23.1	22.1	5.7	51.6
West Midlands	47.4	25.0	21.9	5.7	67.9
East	54.0	23.3	18.1	4.6	65.4
London	48.6	26.4	19.5	5.5	77.7
South East	55.5	22.3	17.2	5.0	95.9
South West	54.6	22.8	17.8	4.8	58.9

Source: Department of Education and Skills

(a) What percentage of pupils in the North West got 5 or more grades A*−C?

....
(1)

(b) Write down the total number of pupils in the United Kingdom in the last year of compulsory education in 2001/ 2002.

....
(1)

(c) Which region shown in the table had the lowest percentage of pupils with no graded GCSEs?

....
(1)

The following statement is wrong.

"19.5% of pupils getting grades D−G only, take their GCSE exams in London."

(d) Write down a correct version of this statement.

....

....
(1)

Page Total

4. This table below is part of a larger table that shows the percentages of male and female pupils achieving GCSEs in the United Kingdom in 2001/ 2002.

Examination Achievements 2001/ 2002

	Pupils in the last year of compulsory education				
	Percentage achieving GCSE				
	5 or more grades A*-C	1-4 grades A*-C	Grades D-G only	No graded GCSEs	Total (=100%) (thousands)
Males United Kingdom	45.7	24.6	23.1	6.5	372.1
Females United Kingdom	56.5	23.6	15.5	4.4	357.6

Source: Department of Education and Skills
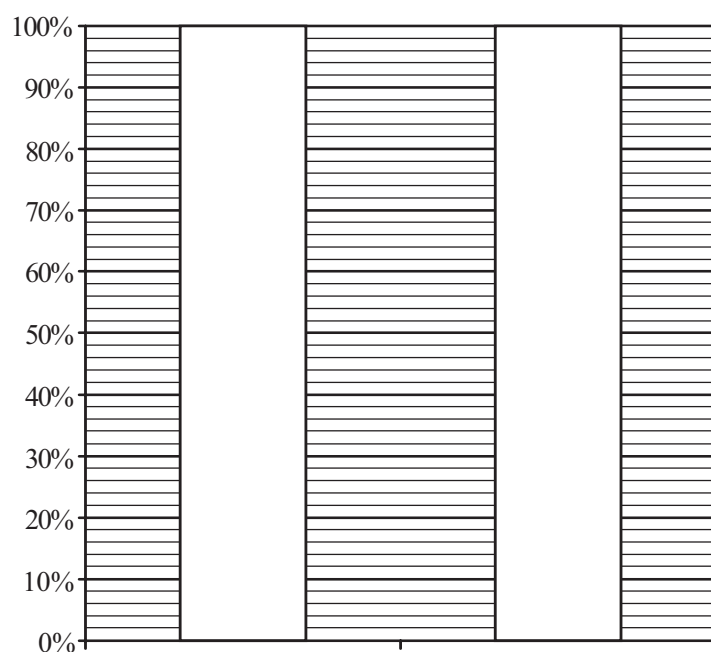
(a) The percentages for the male pupils do not add up to 100%. Explain why.

....

....

(1)

(b) Using the following grid, draw composite bar charts to show the data given in the table above.

Key	
5 or more A*-C grades	
1 - 4 A*-C grades	
Grades D - G only	
No graded GCSEs	

(3)

Do not write here

5. Two judges had to rank six gymnasts. Here are the rankings.

Gymnast	A	B	C	D	E	F
Judge 1	4	5	6	2	1	3
Judge 2	4	6	5	3	1	2

(a) Use this information to calculate Spearman's rank correlation coefficient.

....

(3)

(b) Interpret your answer to part (a).

....

....

....

(2)

Do not write here

Page Total

6. The following table shows the population of Hambleton in 1960 and in 1980.

Hambleton

Year	1960	1980
Population	6400	7040

(a) Taking 1960 as the base year, work out the population of Hambleton in 1980 as an index number.

....

(2)

Thorpe is a town near to Hambleton.

The population of Thorpe decreased by 2% between 1970 and 1990.

(b) Taking 1970 as the base year, complete the table below.

Thorpe

Year	1970	1990
Index number		

(2)

Do not write here

7. The table gives the mean and standard deviation of the marks in three examinations. The marks in each of these examinations are normally distributed.

Examination	Mean	Standard deviation
French	60	15
German	56	10
Science	50	12.5

Mary got a mark of 72 in the French examination and a mark of 66 in the German examination.

(a) Calculate Mary's standardised score for these two examinations.

Standardised French score ....

Standardised German score ....
(3)

(b) Did Mary do better in the French examination or in the German examination? Give a reason for your answer.

....

....
(1)

In the Science examination Mary had a standardised score of –0.56

(c) Calculate Mary's mark in the Science examination.

....
(2)

Do not write here

Page Total
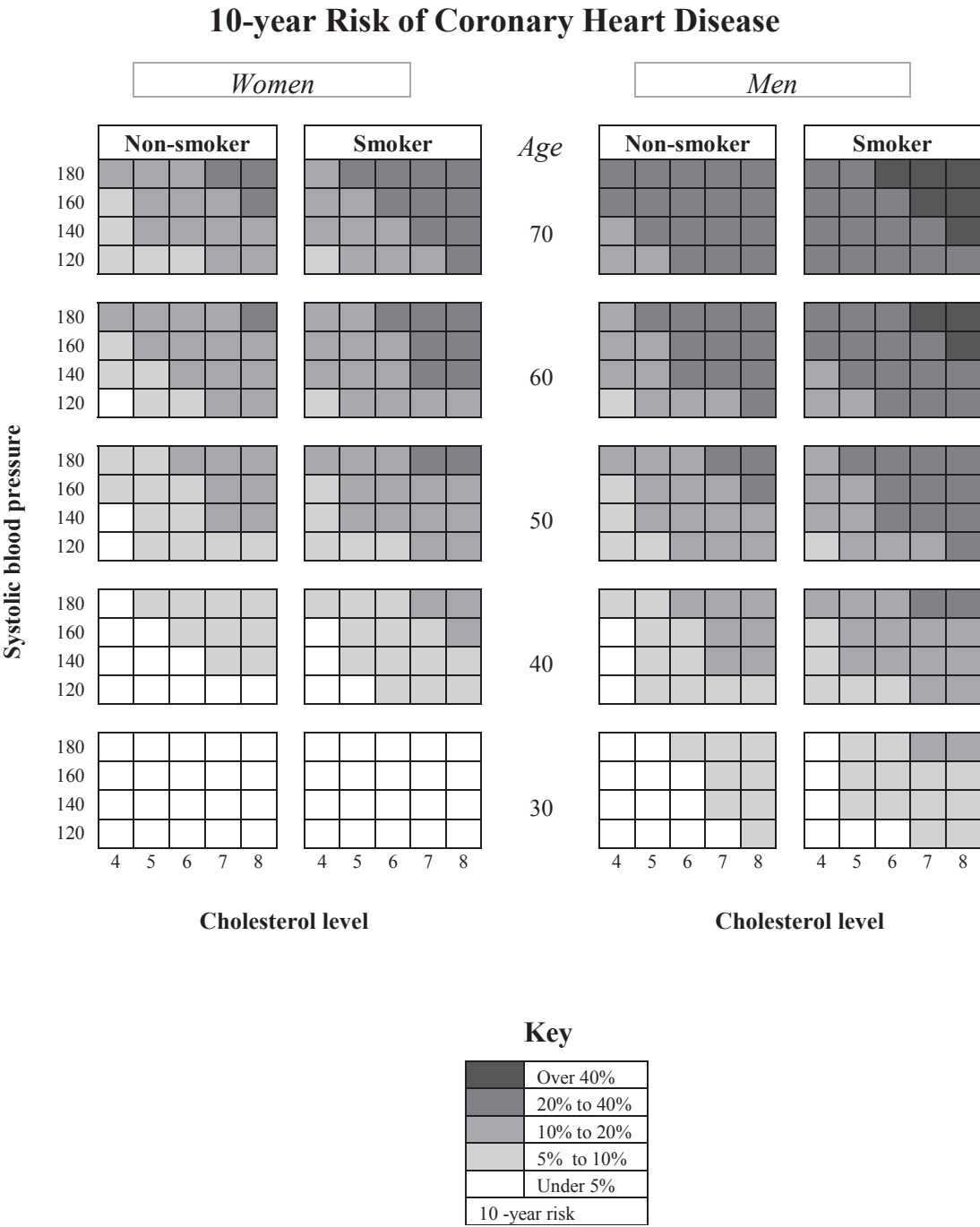
**8.** The diagram below is taken from an article about risks of coronary heart disease.

The article was published in a journal of the Royal Statistical Society in 2003.

10-year Risk of Coronary Heart Disease

Cholesterol level

Cholesterol level

Key

	Over 40%
	20% to 40%
	10% to 20%
	5% to 10%
	Under 5%
10 -year risk	

The diagram shows the risk of coronary heart disease for women and men.

The diagram shows that a woman with a systolic blood pressure of 120 and a cholesterol level of 4 has a 10-year risk of heart disease between 5% and 10% if she is 50 years old and a smoker.

(a) What is the risk if

    (i) the age of the woman is 50 and she is a non-smoker,

                                                 ....

    (ii) the age of the woman is 60 and she is a smoker?

                                                 ....

(2)

(b) Is the 10-year risk greater for men or for women? Give a reason for your answer.

....

....

....

(2)

(c) Write down two things that the diagram shows about the effects of age and smoking on the risk of coronary heart disease.

    (i) ....

    (ii) ....

(2)
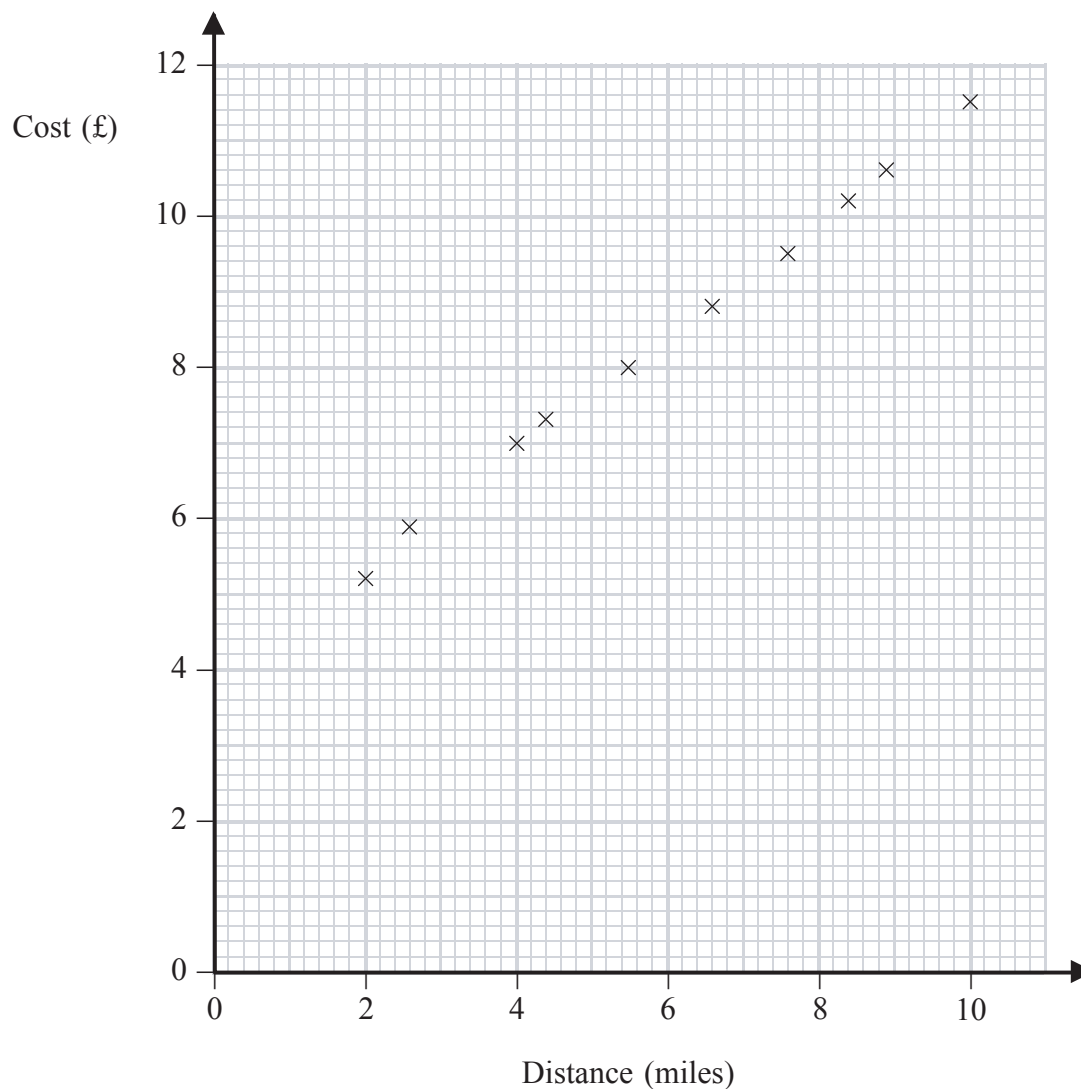
TOTAL FOR SECTION A: 35 MARKS

Do not write here

Page Total

SECTION B

Answer ALL SEVEN questions.

9. The scatter diagram shows the distance, x miles, and the cost, £y, of each of ten taxi journeys in Huddersgate.

Cost (£)

Distance (miles)

The table below gives the distance, x miles, and the cost, £y, for each of the ten journeys shown in the scatter diagram.

x miles	£y
8.4	10.20
6.6	8.80
4.0	7.00
2.6	5.90
5.5	8.00
8.9	10.60
2.0	5.20
10.0	11.50
4.4	7.30
7.6	9.50

(a) What type of correlation does this scatter diagram show?

....

(1)

(b) Calculate the coordinates of the mean point $(\overline{x}, \overline{y})$ for the data.

(........................ , ...)

(2)

(c) On the scatter diagram

(i)   plot the point $(\overline{x}, \overline{y})$,

(ii) draw the line of best fit.

(2)

The equation of the line of best fit has the form $y = ax + b$.

(d) Using your line of best fit, find the value of a and the value of b.

$a =$ ....

$b =$ ....

(3)

(e) Explain what the value of a and the value of b represent in this context.

....

....

....

....

(2)

(f) Give a reason why the equation $y = ax + b$ may not be used reliably to predict the cost of a journey of 100 miles.

....

....

....

(1)

Page Total

10. A farmer records the weight of plums, to the nearest kg, that were produced by each of her 15 plum trees. The results are given below.

80	42	68	72	84	92	77	89
68	69	66	86	73	81	65	

(a) Complete the ordered stem and leaf diagram below.

Weight of Plums   Key: 8 | 0 = 80 kg

```
4 |
5 |
6 |
7 |
8 |
9 |
```

(2)

(b) For this data

(i) find the median,

....

(ii) work out the inter-quartile range.

....

(3)

The farmer thinks the tree that produced 42 kg of plums is diseased.

(c) Is 42 an outlier for this data? Give a reason for your answer.

....

....

....

(2)

The farmer removes 42 from her results.
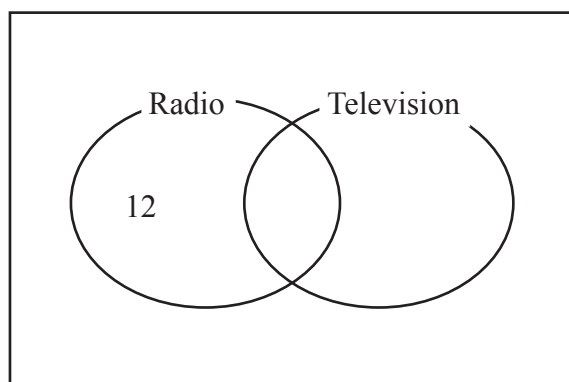
(d) Work out the median for the remaining 14 results.

....

(2)

11. A health trust does a survey to see how many of its hospital beds have access to a radio and/or a television. It finds that 21% have access to televisions only, 12% have access to radios only and 2% have access to neither.

(a) Complete the Venn diagram.

Radio   Television

12

(2)

(b) Find the probability that a bed chosen at random will have access to a radio or a television but not both.

....
(1)

(c) Find the probability that a bed chosen at random will **not** have access to a television.

....
(1)

(d) Given that a patient is in a bed with access to a radio, what is the probability that the bed also has access to a television?

....
(2)

Do not write here

Page Total

12. (a) Explain what a census is.

....

....

(1)

(b) Give two reasons why a person may take a sample rather than a census.

(i) ....

....

(ii) ....

....

(2)

Ashton Girls school wants to take a sample of 100 of its students to find out what they think about compulsory games. There are five year-groups in the school. Each year-group has the same number of students.

Ashton Girls school wants the sample to be representative of the whole school.

(c) Describe how the school should select the sample.

....

....

....

....

(2)

The school is to ask selected students to fill in a questionnaire.

(d) Give two reasons why a pilot survey should be carried out first.

(i) ....

....

(ii) ....

....

(2)

One question on the questionnaire is

"Do you agree that compulsory games are good for you?"

(e) This question is biased. Give one reason why.

....

(1)

It is suggested that it might be better to interview the students rather than asking them to fill in a questionnaire.

(f) Write down one advantage and one disadvantage of interviewing the students.

Advantage ....

....

....

Disadvantage ....

....

....

(2)

Do not write here

13. On a production line in a factory, baked beans in tomato sauce are put in tins.

The label on each tin says that the contents weigh 415 g.

(a) Give two reasons why it is not practical to check the weight of the contents of each tin.

(i) ....

....

(ii) ....

....

(2)

Samples of tins are taken at intervals and the weights of the contents are found.

It has been found that the mean weight of the samples is 417g and the standard deviation of the mean weights of the samples is 0.6g.

The mean weights of the samples are normally distributed.

(b) Between what limits would you expect 99.8% of the sample means to lie?

....

(3)

The target weight of the contents is set at 417g.

(c) Using your answer to part (b), give a reason why the target weight is 417g rather than 415g.

....

....
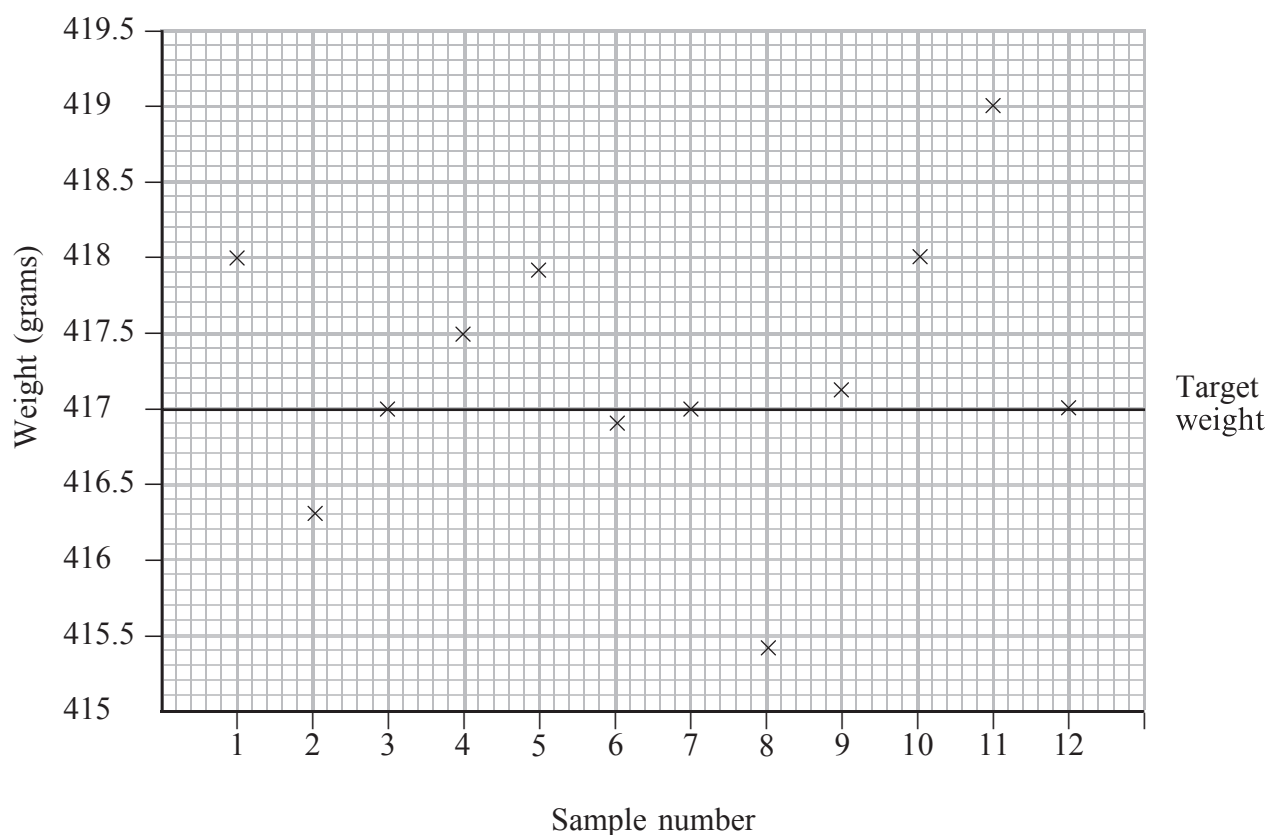
(1)

Do not write here

A sample of tins is taken each half hour during a six-hour shift and the mean weight of the contents is found.

The mean weight of the samples are plotted on the chart below.

The allowable limits for the weights of the samples are ±3 standard deviations from the target weight.

(d) Comment on any action that would have been taken during the six-hour shift.

....

....

(2)

Do not write here

Page Total

14. In an experiment a psychologist records the times, x seconds, for 50 people to complete a puzzle.

The results are summarised in the table.

x	Number of people (f)			
$0 < x \leq 40$	10			
$40 < x \leq 60$	9			
$60 < x \leq 80$	12			
$80 < x \leq 100$	9			
$100 < x \leq 180$	10			
	Total			

You may use the space provided in the table to answer this question.

(a) (i) Work out an estimate for the mean time.

....
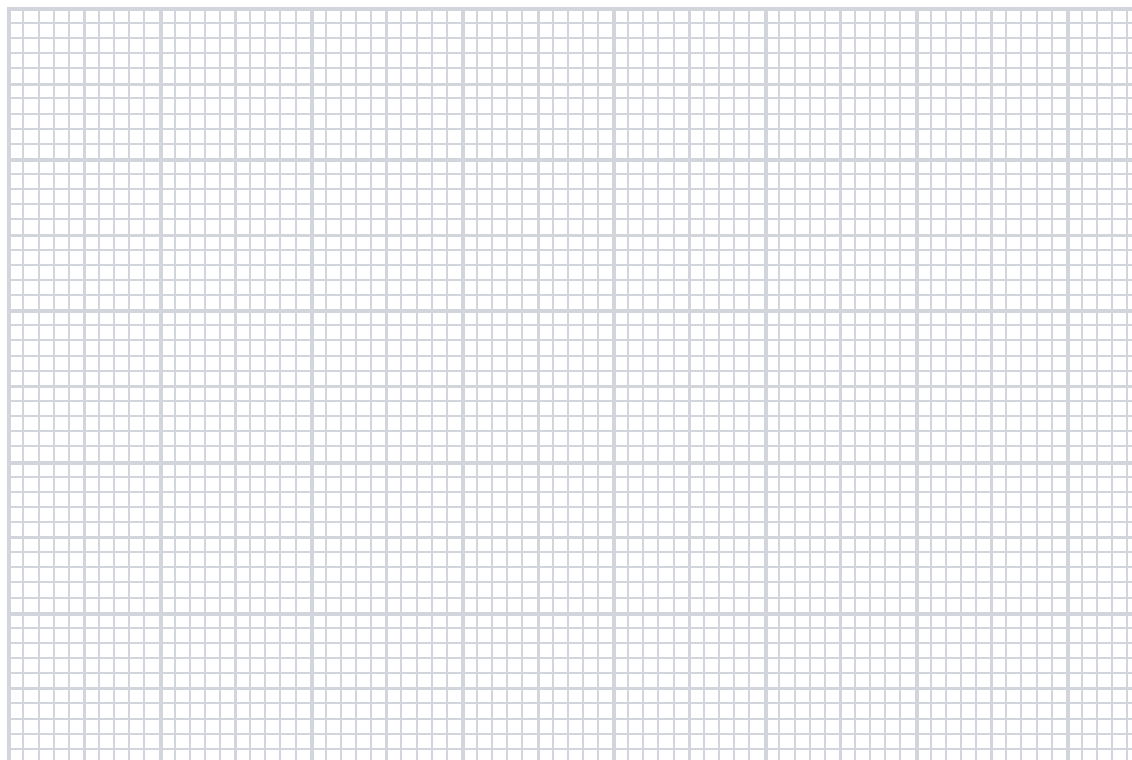
(ii) Show that the standard deviation for the times is approximately 40 seconds.

(5)

(b) Draw a histogram to represent the data in the table.

(3)

(c) (i) Shade the region in your histogram that is within two standard deviations of the mean.

(ii) Find the proportion of people represented by this region.

....
(4)

It is claimed that the time to complete the puzzle is normally distributed.

(d) Comment on the validity of this claim.

....

....

....
(2)

Do not write here

Page Total

15. John is going on a 5-day holiday to Costa Packet. The travel brochure says that on average 5 out of every 7 days are sunny at Costa Packet. Each day's weather is independent of the weather on all preceding days.

(a) (i) Name the probability distribution that would model the number of sunny days at Costa Packet.

....

There are two values, n and p, that you need to use in this probability distribution.

(ii) Write down the value of n and the value of p.

$n = $ ....

$p = $ ....

(2)

(b) Calculate the probability that John has 2, or less, sunny days on his holiday?

You may use $(p + q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$.

....

(3)

(c) What is the most likely number of sunny days that John will get on his holiday? Show your working.

....

(2)

TOTAL FOR SECTION B: 65 MARKS

TOTAL FOR PAPER: 100 MARKS

END

Page Total

BLANK PAGE

BLANK PAGE