

GCSE STATISTICS TOPICS NOT COVERED IN ORDINARY MATHEMATICS

SAMPLING & BIAS

1. A town council wants to know what people think about how the council uses the council tax.

It will take a sample of all the people who pay council tax.

- (a) Which of the following sample frames would be most suitable?

- A The list of names in a telephone directory.
- B A list of all council tax payers.
- C A list of residents of one street.
- D The electoral roll.

(1)

The population of the town is very large.

The council will take a sample of 1000 people.

- (b) Which one of the following sampling methods is the most suitable?

- A Convenience sampling.
- B Systematic sampling.
- C Cluster sampling.

(1)

2. People have different reaction times when using either their left hand or their right hand. Megan wants to investigate this.

She is going to select a number of students from her school to use as a sample for the investigation.

- (a) What is the population for Megan's sample?

(1)

Megan selects a number of students from her class to use as a sample for the investigation.

- (b) Give a reason why this is not a good way of taking a sample.
- (c) Describe a better way of taking a sample that Megan could use.

(1)

(1)

3. 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.

(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)

SPEARMAN'S RANK CORRELATION COEFFICIENT

4. Ten schools entered a team of boys and a team of girls in a swimming gala.

The table shows the points scored by the boys' team and by the girls' team for each of the ten schools.

School	Boys' team points	Girls' team points	Boys' team rank	Girls' team rank		
A	19	15				
B	11	11				
C	16	7				
D	3	3				
E	13	5				
F	18	17				
G	7	4				
H	5	1				
I	14	2				
J	1	0				

- (a) Work out Spearman's rank correlation coefficient.
You may use the table for some of your working.

(4)

- (b) (i) Describe this correlation.

- (ii) Write down what you conclude about the relationship between the teams of boys and the teams of girls in the same school.

(2)

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)

Spearman's Rank Correlation Coefficient $1 - \frac{6 \sum d^2}{n(n^2 - 1)}$

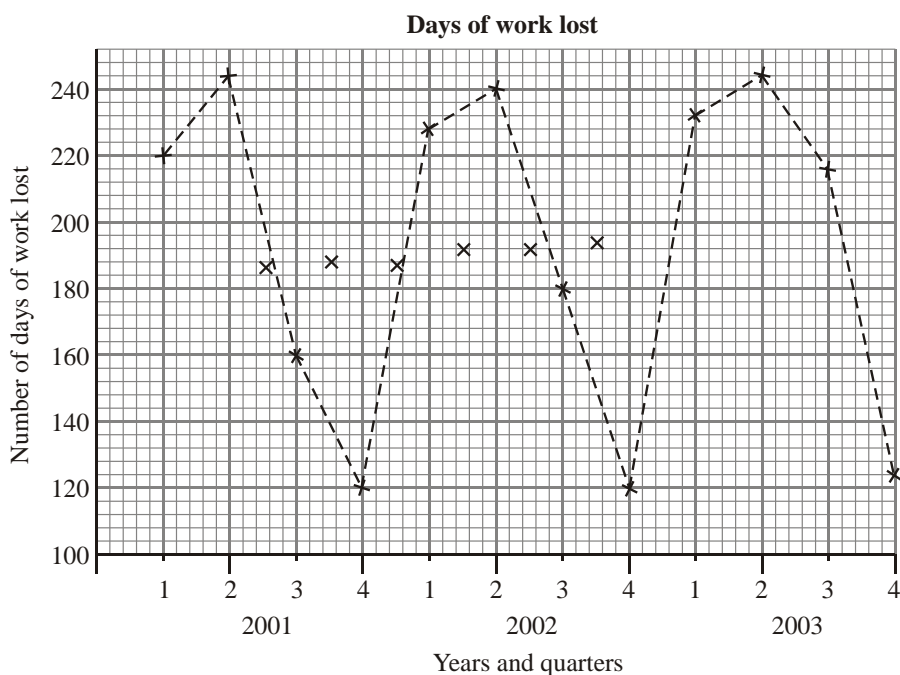
MOVING AVERAGES

6. The number of days of work lost per quarter due to illness by employees at Gale engineering is shown in the table.

Year	Quarter	Days
2001	1	220
	2	244
	3	160
	4	120
2002	1	228
	2	240
	3	180
	4	120
2003	1	232
	2	244
	3	216
	4	124

A time-series has been drawn below to show the data.

The first six of the 4-point moving averages have been plotted.

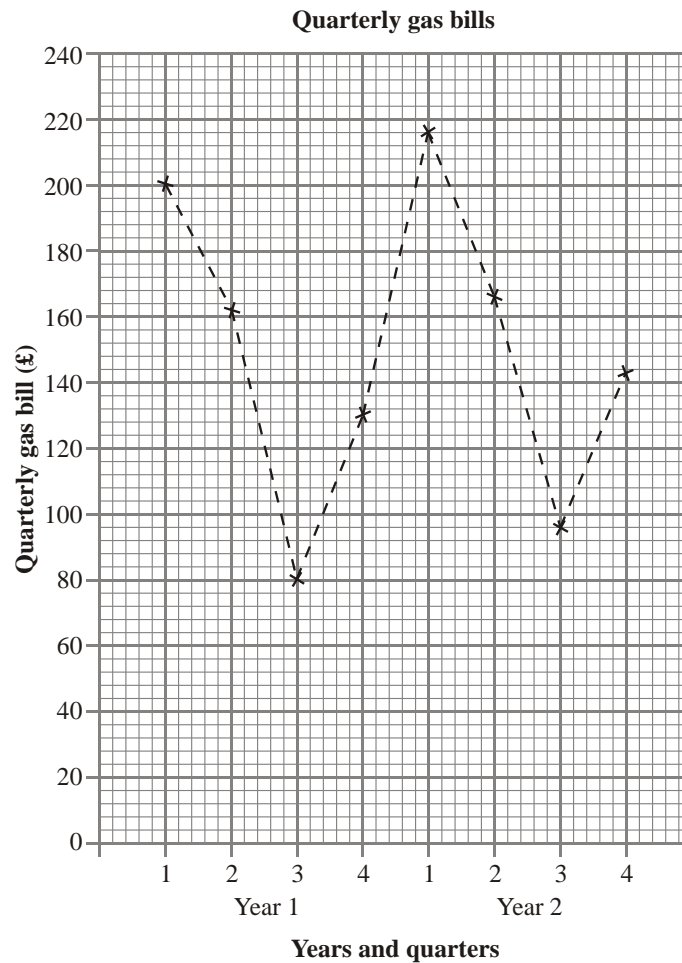


- (a) Why should the company consider 4-point moving averages? (1)
- (b) Calculate the remaining 4-point moving averages and plot them on the graph. (3)
- (c) On your diagram, draw a trend line for the moving averages. (1)
- The mean seasonal effect for the first quarter is +38.
- (d) Estimate the number of days of work likely to be lost in the first quarter of 2004. (2)
- The trend is for the number of days of work lost to increase.
- (e) Suggest a reason for this. (1)

7. The table shows information about the quarterly gas bill, in £'s, for Samira's house, over a period of two years.

	Quarter			
Year	1	2	3	4
1	£200	£162	£80	£130
2	£216	£166	£96	£142

The data has been plotted as a time series



- (a) The first three 4-point moving averages are £143, £147 and £148.

(i) Work out the last two 4-point moving averages.

(ii) Plot all five of the moving averages on the graph.

(4)

- (b) What do the moving averages show about the trend of the quarterly gas bills?

(1)

The time series shows that the quarterly gas bills are varying from the general trend.

- (c) (i) Write down what these variations are called.

(ii) Write down a reason for these variations.

(2)

STEM & LEAF DIAGRAMS

8. During one week a dentist recorded the time, to the nearest minute, that it took him to do a particular type of filling.

These are his results:

25 16 18 18 26 22 22 11 9 32 30 22 24 22 30

- (a) Complete the following ordered stem and leaf diagram.

Key 1 6 = 16	
0	9
1	1 6 8 8
2	
3	

(1)

- (b) Write down the median of these results.

(1)

- (c) Work out the mean of these results.

(2)

Another patient was treated and the filling took 24 minutes to do.

- (d) Write down in words, and without doing a calculation, the effect this will have on

(i) the median,

(ii) the mean.

(2)

- (e) How long should the dentist allow for appointments for this type of filling? Explain your answer.

(1)

9. 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)

9. (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)

GEOMETRIC MEAN

10. The population density (population/area in hectares) for each of 3 towns in Devon is given in the table below.

Area	Population Density (population/hectare)
Exeter	23.57
Plymouth	31.16
Brixham	13.89

Source: Registrar General http://www.devon.gov.uk/dris/admin/oth_dens.html

(a) Calculate the geometric mean for the population densities of Exeter, Plymouth and Brixham.

(2)

(b) Why should a geometric mean be used for this data?

(1)

INDEX NUMBERS

11. Part of the wage bills of a factory (in £1000s) for the years 1999 to 2003 are shown in the table.

Some index numbers are shown. The base year is the year 1999.

Year	1999	2000	2001	2002	2003
Wages (£1000s)	200	240		300	320
Index number	100	120	125	150	

(a) Work out the index number for 2003.

(2)

(b) Work out the wage bill of the factory in 2001.

(2)

The base year is changed to 2000.

(c) Work out the new index number for 2002.

(2)

12. 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)

13. Zenith Engineering makes stainless steel.

The raw materials for making stainless steel are steel, chromium and nickel.

A stainless steel is made with 74% steel, 18% chromium and 8% nickel.

The table shows the price, in pounds per ton, for chromium and nickel in the years 2000 and 2003.

	Metal prices	
	Year	
	2000	2003
Chromium (£/ton)	41 500	43 575
Nickel (£/ton)	8 000	8 240

- (a) Taking 2000 as the base year, calculate the index numbers for the price of chromium and nickel in 2003.

Metal	Index 2000	Index 2003
Steel	100	102
Chromium	100	
Nickel	100	

(3)

The index number for the price of steel in 2003 was 102.

Zenith Engineering wants to find out the weighted index number for the price of the raw materials for stainless steel.

- (b) Calculate this number.

(3)

- (c) Write down the percentage by which the price of raw materials for stainless steel went up between 2000 and 2003.

(1)

14. 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. (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)

STEM & LEAF DIAGRAMS; MEAN, MEDIAN & MODE

15. During one milking session a farmer records the amount of milk, in litres, collected from each of his cows. The results are shown below.

12 18 26 28 28 31 34 34 34 42 43

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

Amount of Milk	Key: 1 2 = 12 litres
1	
2	
3	
4	

- (b) Write down the modal amount of milk collected from the cows. (2)
- (c) Write down the median amount of milk collected from the cows. (1)
- (d) Work out the mean amount of milk collected from the cows. (1)
- (e) Work out the inter-quartile range for the amount of milk collected from the cows. (2)

Buttercup produced 12 litres of milk.

In the next milking session Buttercup produced more than 12 litres of milk.

The amount produced by the other cows remained the same.

- (f) What effect would this have on the mean amount of milk produced by all the cows? Explain your answer. (2)

FREQUENCY TABLES; HISTOGRAMS; VARIANCE & STANDARD DEVIATION

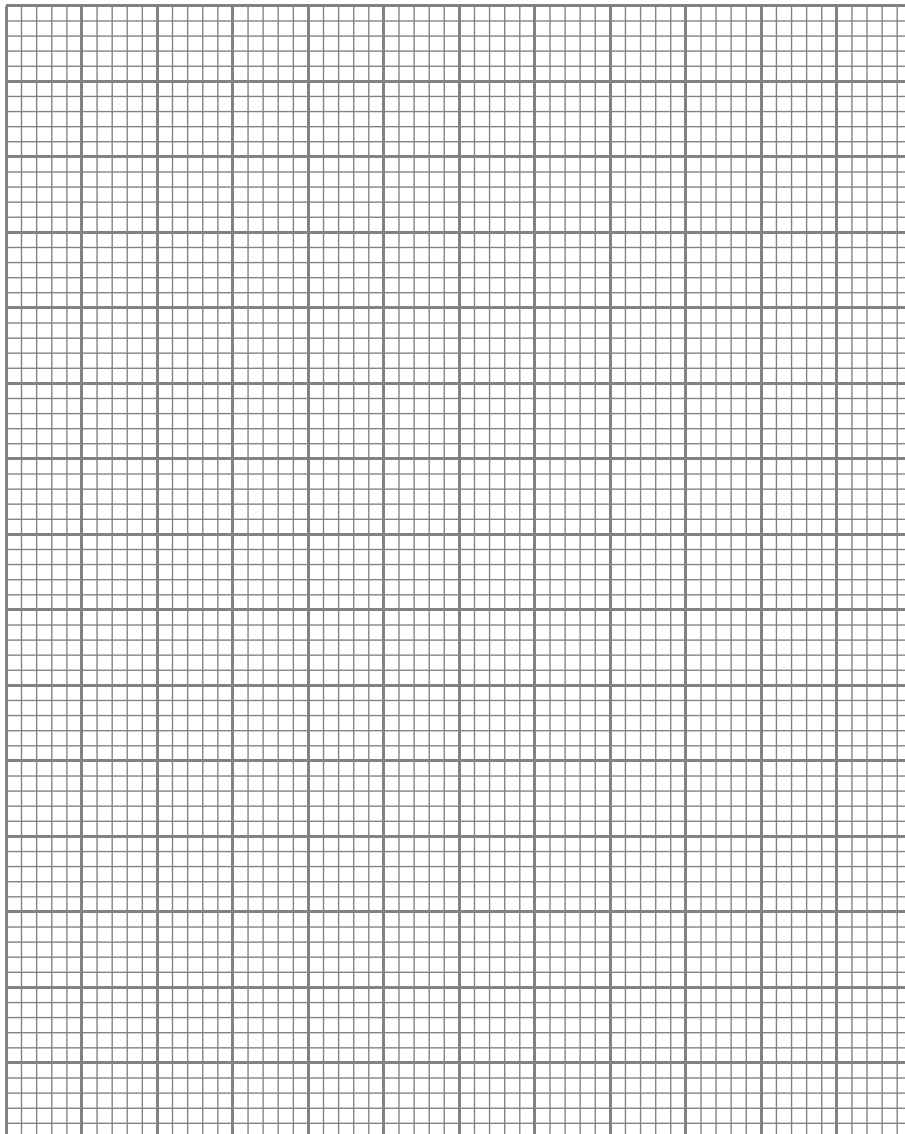
16. In an experiment a biologist records the length, x cm, of 125 worms.
The results are summarised in the table.

x	Number of worms (f)		
$0 < x \leq 4$	36		
$4 < x \leq 6$	26		
$6 < x \leq 8$	20		
$8 < x \leq 10$	12		
$10 < x \leq 15$	21		
$15 < x \leq 20$	10		
Total	125		

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

- (a) (i) Show that an estimate for the mean length of the worms is 7.1 cm.
(ii) Work out an estimate for the standard deviation of the lengths of the worms.
(You may use $\sum fx^2 = 9089.75$)
- (b) Draw a histogram to represent the data in the table.

(6)



(3)

(c) (i) Shade the region in your histogram that represents worms whose length is less than or equal to the mean length of the worms.

(ii) Find the number of worms whose lengths are less than or equal to 7.1 cm.

(3)

(d) Are the lengths of these worms normally distributed? Explain your answer.

(1)

17. 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			

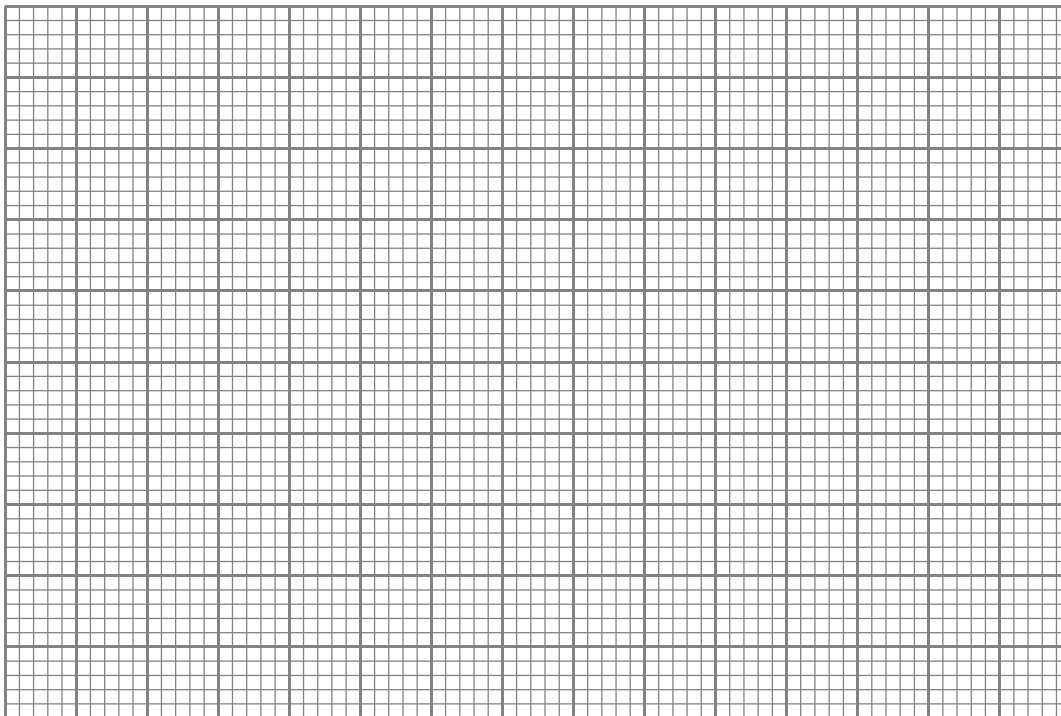
You may use the space provided in the table to answer this question.

(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)

QUALITY ASSURANCE

20. A machine in a factory fills bags with rice.

- (a) Give **one** reason why a quality assurance chart should be used?

(1)

The mean weight of rice in the bags is 260 g. The standard deviation of the weight of rice in the bags is 5 g.

- (b) Write down the target value for this process.

(1)

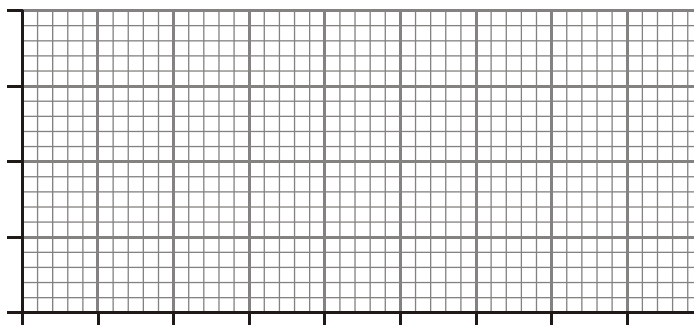
Samples were taken every hour for 8 hours and the mean weights of the samples are calculated.

The table shows the mean weights of the samples.

Sample number	1	2	3	4	5	6	7	8
Mean weight (g)	265	260	258	270	260	255	252	262

A quality assurance chart for the sample means is to be drawn. The allowable limits are 3 standard deviations above and 3 standard deviations below the target value.

- (c) Draw a quality assurance chart for the data.



(4)

- (d) Did the machine need to be reset during this period? Give a reason for your answer.

(2)

The probability that a bag of rice weighs less than 260 g is 0.3.

Four bags are chosen at random.

- (e) What is the probability that exactly one bag weighs less than 260 g?

You may use $p^4 + 4pq^3 + 6p^2q^2 + 4pq^3 + q^4$

(3)

21. 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.

(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 417 g and the standard deviation of the mean weights of the samples is 0.6 g.

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 417 g.

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

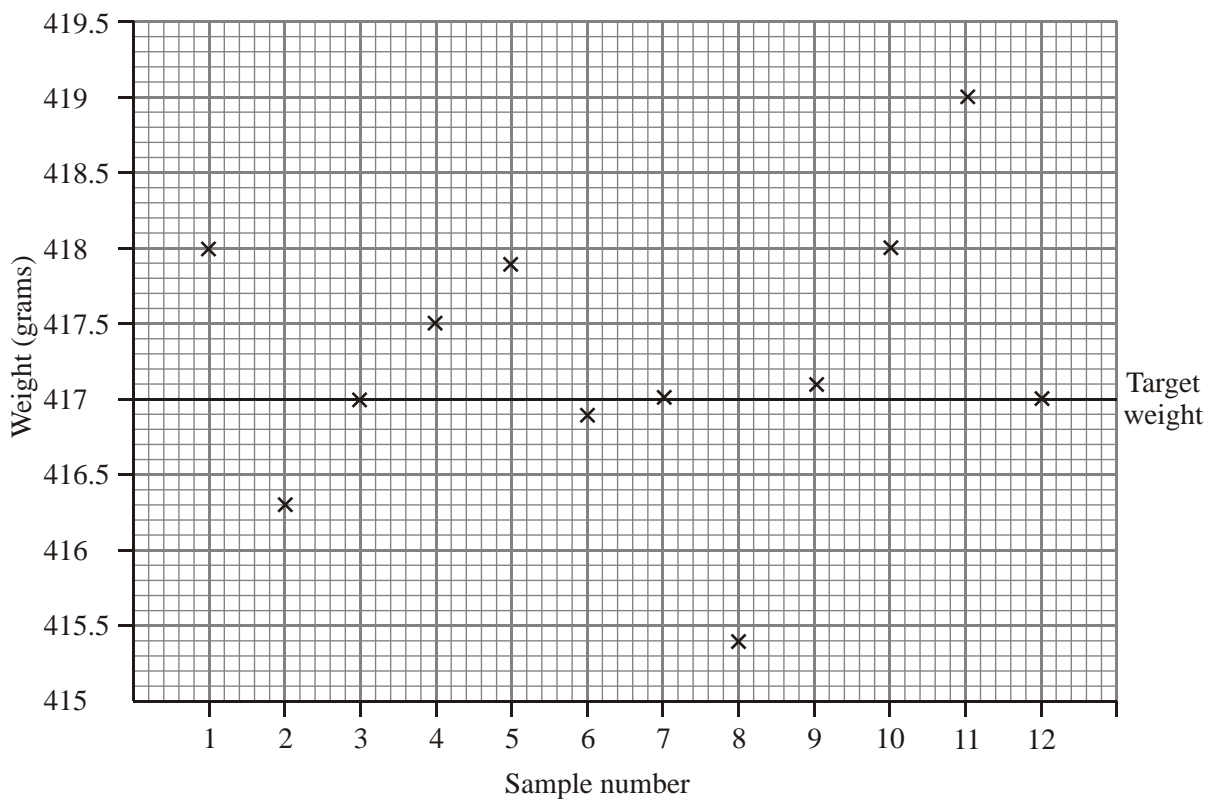
.....

.....

(1)

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)

22. A machine fills packets with roast peanuts.

Bob takes a sample of 10 packets every hour.

The mean weights of the samples are normally distributed with a mean of 106 grams and a standard deviation of 2 grams.

(a) Write down the percentage of the samples that are likely to have a sample mean of less than 100 grams. (1)

(b) Write down the percentage of the samples that are likely to have a sample mean of more than 110 grams. (1)

Bob sets up a quality assurance chart for means.

(c) Write down **one** reason why Bob uses a quality assurance chart. (1)

The chart has a lower action limit of 100 grams.

(d) One sample mean is less than 100 grams. What should Bob do? (1)

(e) Why should Bob also have an upper action limit? (1)

(f) What other sort of quality assurance chart is Bob likely to use? (1)

CONDITIONAL PROBABILITY & VENN DIAGRAMS

A travel agent organises several different tours to Germany, France and Switzerland. Each tour goes to 1, 2 or 3 of these countries.

A total of 200 people went on these tours.

Of these

130 people went on a tour to Germany,

131 people went on a tour to France,

122 people went on a tour to Switzerland,

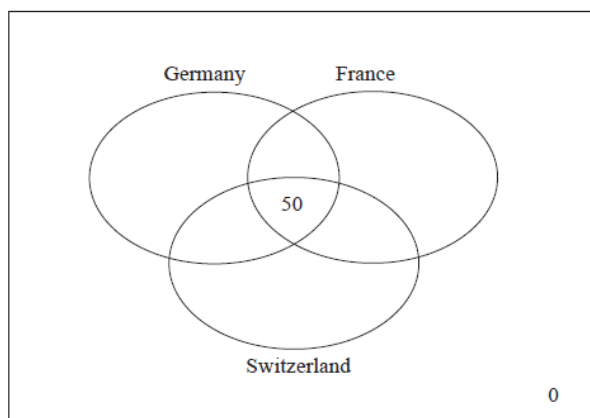
74 people went on a tour to Switzerland and France,

84 people went on a tour to France and Germany,

75 people went on a tour to Germany and Switzerland,

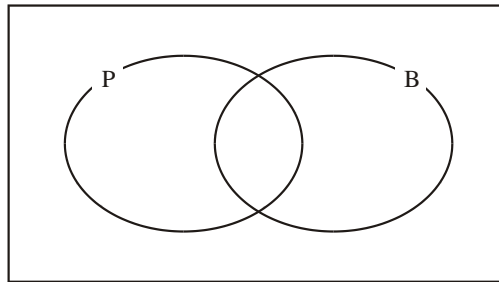
50 people went on a tour to all three countries.

(a) Complete the Venn diagram for this information.



23. There are 1000 students at a Cranborne College. 400 students at a college study Photography (P). 700 study Business (B). 200 study neither subject.

(a) Complete the Venn diagram to show this information.



(3)

One student is picked at random from those at Cranborne College.

(b) Write down the probability that the student is studying

- (i) Photography but not Business,
- (ii) Business and Photography.

(2)

Two students are picked at random.

(c) Calculate the probability that both are studying Photography.

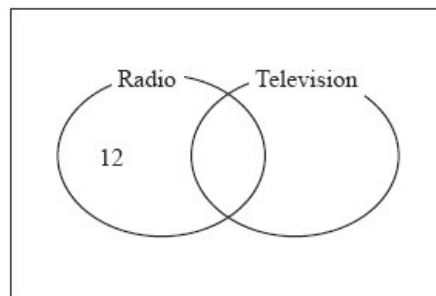
(2)

(d) Find the probability that a student picked at random is studying Photography, given that he or she is studying Business.

(2)

24. 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.



(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)

25. There are 30 people eating in a restaurant.

Each person chooses a first course of fish or meat or vegetarian, and a second course of cake or ice cream or cheese.

The numbers of people eating each combination of first and second course are given in the table.

		Choice of meals		
		Second course		
		Cake	Ice cream	Cheese
First course	Fish	4	4	3
	Meat	5	3	4
	Vegetarian	2	3	2

One of the 30 people is chosen at random.

(a) Write down the probability that this person eats

(i) fish for their first course and cake for their second course,

(ii) ice cream for their second course, given their first course was vegetarian.

(2)

John says that people are equally likely to have fish or meat or vegetarian for their first course.

(b) Does the data support this idea? Give a reason for your answer.

(2)

BINOMIAL DISTRIBUTION

26. Barry has two fair dice. Each dice has 6 faces. Barry rolls both dice.

He adds the numbers on the top of the two dice to get his total score.

(a) Write down all the ways in which he can score 7.

(2)

(b) Write down the probability that he scores 7.

(1)

Barry rolls both dice together 80 times.

(c) Estimate how many times he scores 7.

(1)

Megan throws darts at a target. She can hit the target with 7 out of 9 throws. Megan throws three darts.

(d) Assuming a binomial distribution, work out the probability that she hits the target exactly twice. You may use $(p + q)^3 = p^3 + 3p^2q + 3pq^2 + q^3$.

(2)

(e) In order to assume a binomial distribution, what do you have to assume about the probability of hitting the target with each of the three darts?

(1)

The assumption may not be realistic.

(f) Write down **one** reason why.

(1)

Q26 Part II

Gordon is going to throw five stones in turn at the target.

Gordon can hit the target with a probability of 0.8 with any one of these five stones.

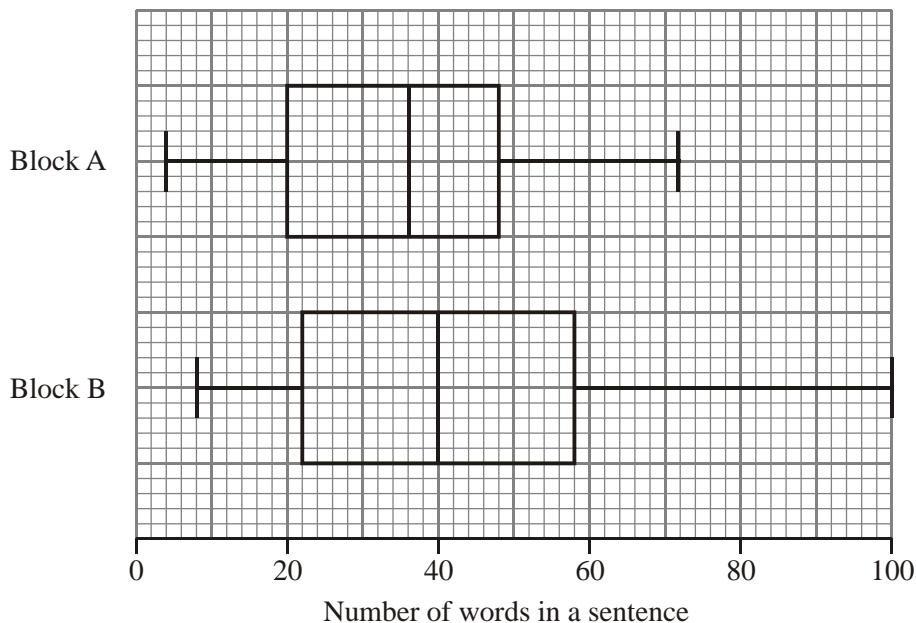
- (d) Name the probability distribution that models the number of times he will hit the target in the five throws.
- (e) Work out the probability that he will hit the target with only one of the five stones.
[You may use $(p+q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$]
- (f) Work out the most likely number of times he will hit the target.
[You may use $(p+q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$]

BOX PLOTS & OUTLIERS

An Outlier is any value that is bigger than or smaller than $(1.5 \times \text{IQR})$

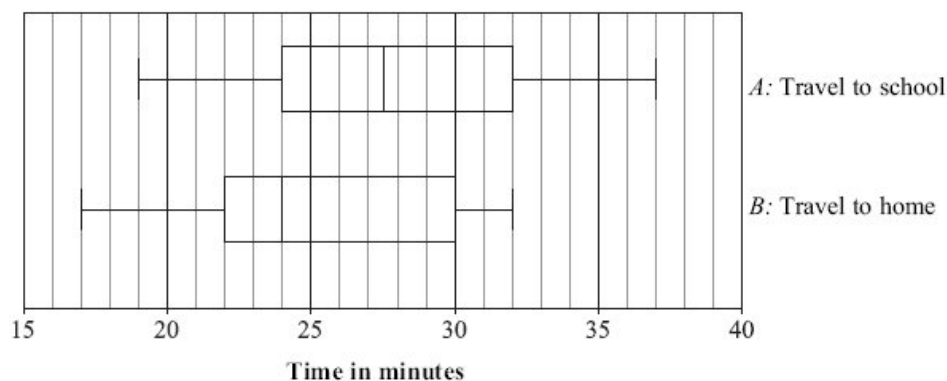
27. Peter takes a random sample of sentences from Book A and from Book B. He counts the number of words per sentence.

The box plots summarise his results.



- (a) Write down the median number of words per sentence for Book B. (1)
- (b) Work out the inter-quartile range of the number of words per sentence for Book A. (1)
- Peter says that these books were written by different authors.
- (c) Write down two ways the box plots show this. (2)

28. The box plots give information about the times, in minutes, some students take to travel from home to school (A), and from school to home (B).



- (a) Work out the range for the times taken to travel to school. (2)
- (b) Write down the median time taken to travel home. (1)
- (c) Write down the shortest time taken to travel home from school. (1)

The students say that it takes longer to travel to school in the morning than it does to travel home in the evening.

- (d) Give **one** way that the box plots support this claim. (1)
- (e) Write down which of the box plots shows the most skewness. Describe this skewness. (2)

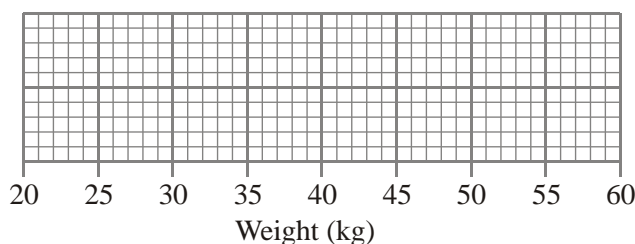
29. The weights, to the nearest kilogram, of 19 pigs were

36 38 30 31 38 43 55 38 37 30 48 41 33 25 34 43 37 40 36

- (a) Find the inter-quartile range of the weights. (2)
- (b) Find any weights that are outliers. (2)

The median of the data is 37 kg.

- (c) Draw a box plot for this data.



- (d) Name a distribution that could be used to model the weight of these pigs. Give a reason for your choice. (2)

The weight of a full-grown pig is about 45 kg.

- (e) What does this suggest about the 19 pigs? (1)

BAR CHARTS & FREQUENCY DIAGRAMS

30. 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

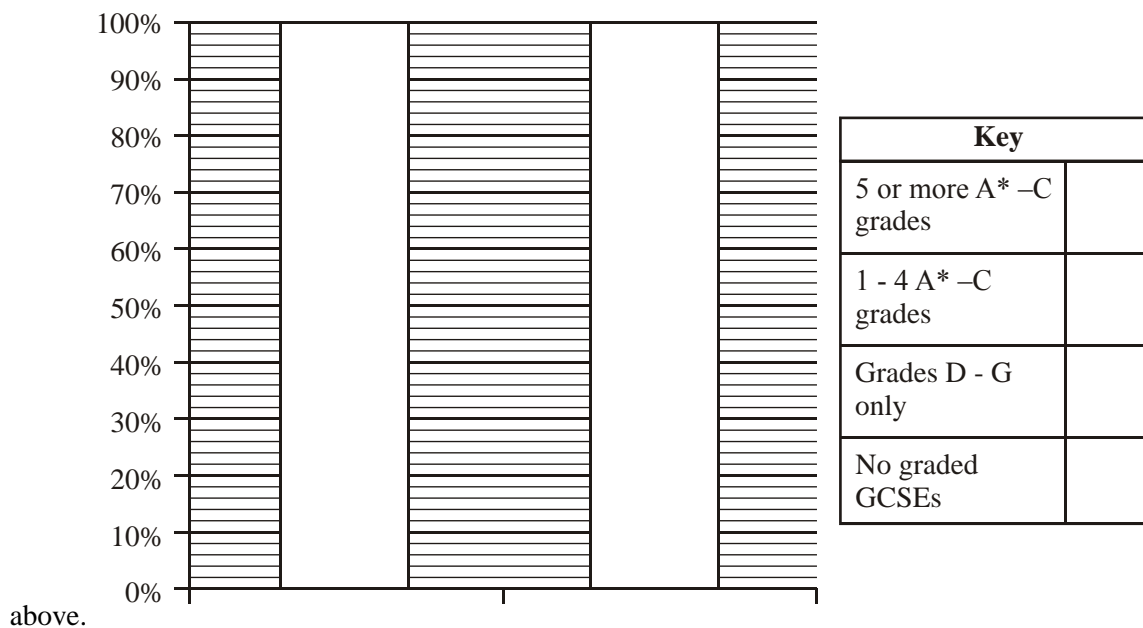
	Pupils in the last year of compulsory education				Total (=100%) (thousands)
	Percentage achieving GCSE				
	5 or more grades A*-C	1-4 grades A*-C	Grades D-G only	No graded GCSEs	
Males United Kingdom	45.7	24.6	23.1	6.5	372.1
Females United Kingdom	56.5	23.6	15.5	4.4	357.6

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



(3)

(Total 4 marks)

31. 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				Total (=100%) (thousands)
		Percentage achieving GCSE				
		5 or more grades A*-C	1-4 grades A*-C	Grades D-G only	No graded GCSEs	
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)

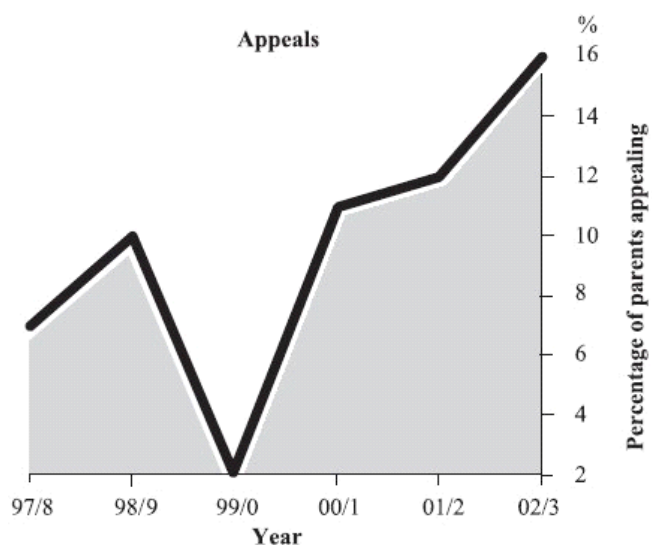
The following statement is wrong.

“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)

32. Parents can appeal against the secondary school their children are sent to.

This graph, from the *Times Educational Supplement*, shows information about appeals from 1997 to 2002.



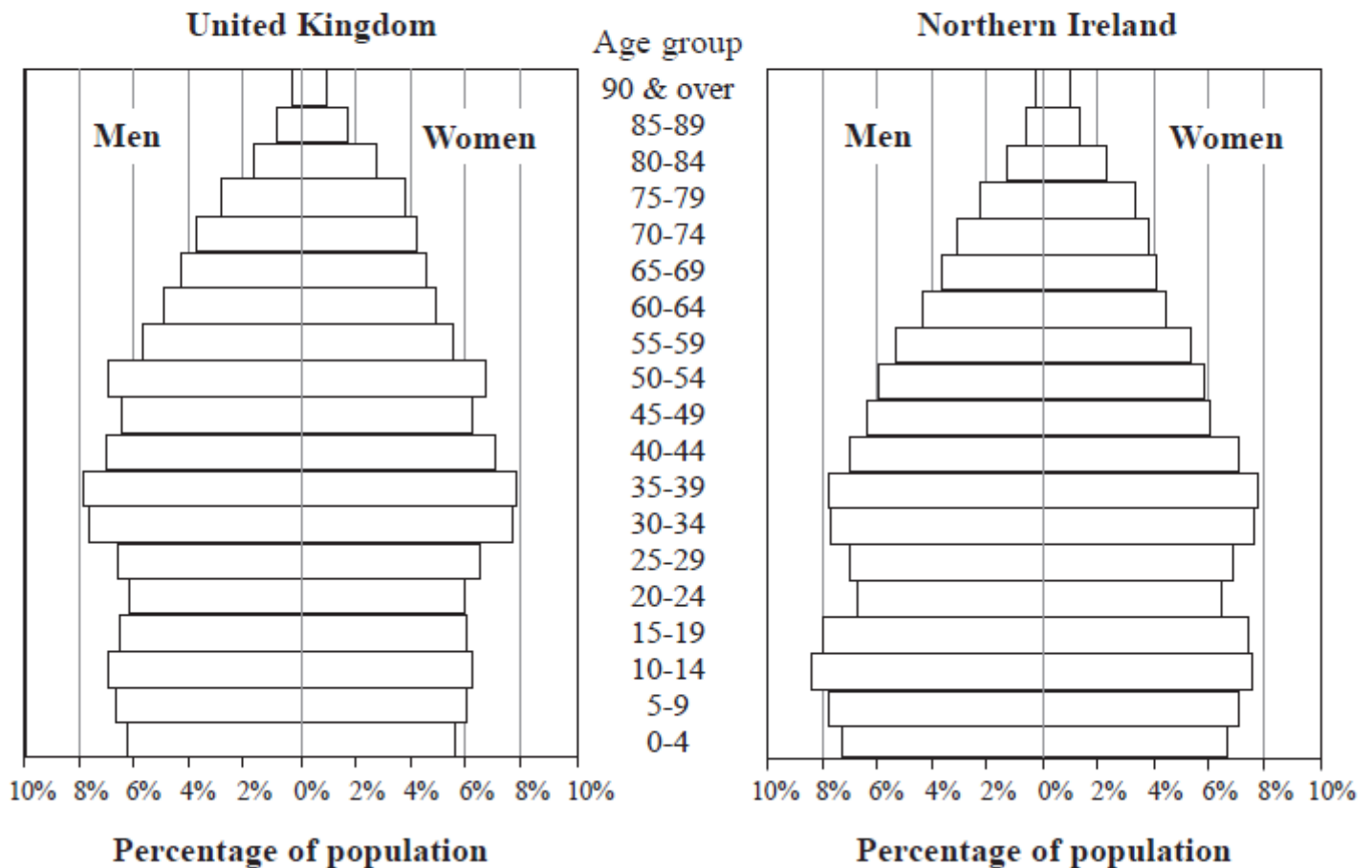
- (a) Write down **two** reasons why this graph could be misleading. (2)
- (b) Does the graph show that the number of parents appealing in 2000/2001 was greater than the number appealing in the year 1999/2000? Give a reason for your answer. (1)

Source: *Sendist*

POPULATION PYRAMIDS

3. The two population pyramids show the percentages of men and women in each age group in the United Kingdom, and in Northern Ireland.

The percentages of men and women in each age group



- (a) Write down the age group which has the greatest percentage of women in the United Kingdom.

6% of the men in Northern Ireland are in one age group.

- (b) Write down this age group.

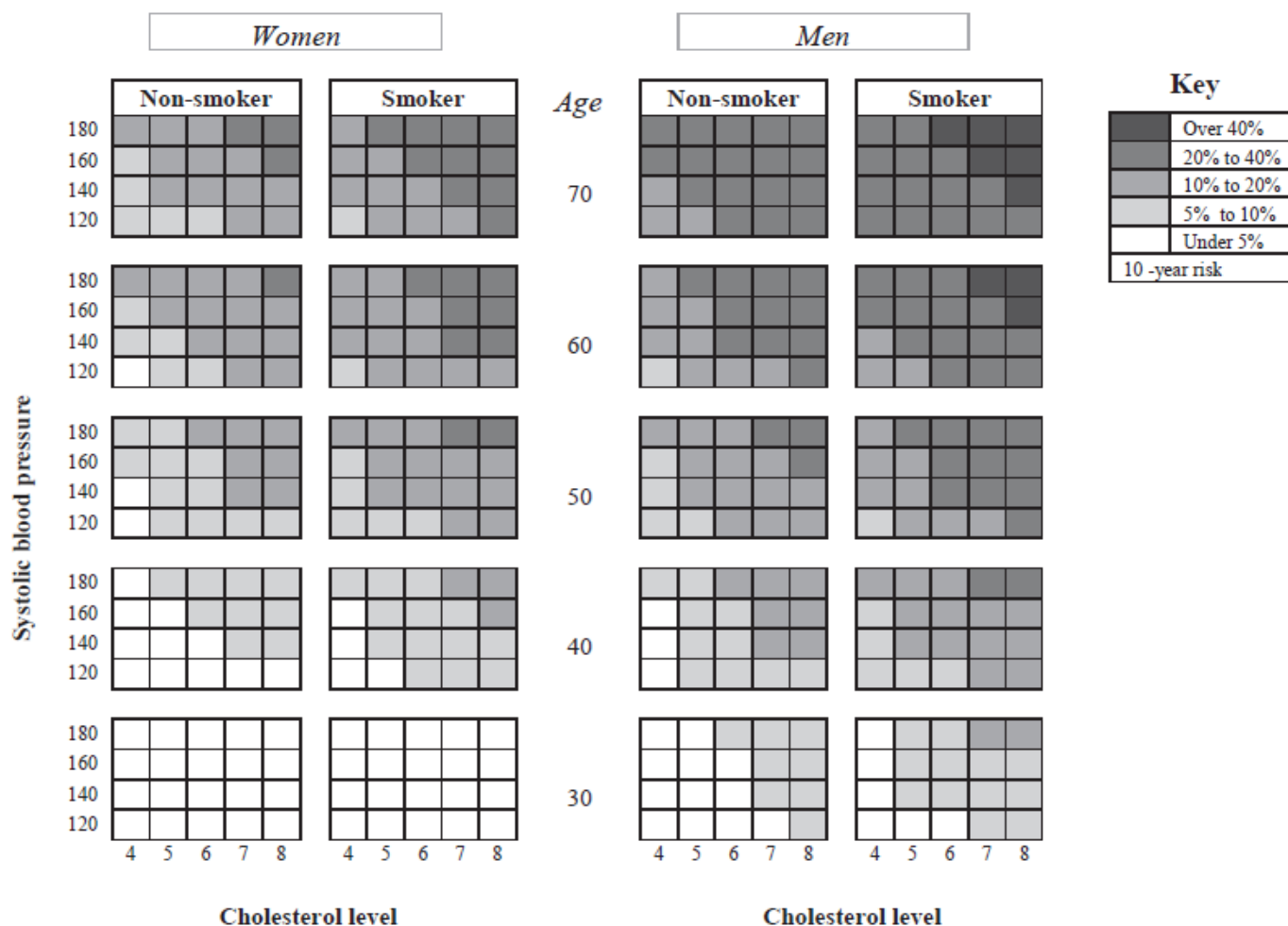
- (c) Compare the percentage of people up to the age of 19 in the United Kingdom with the percentage of people up to the age of 19 in Northern Ireland.

CHOROPLETH MAPS

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



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

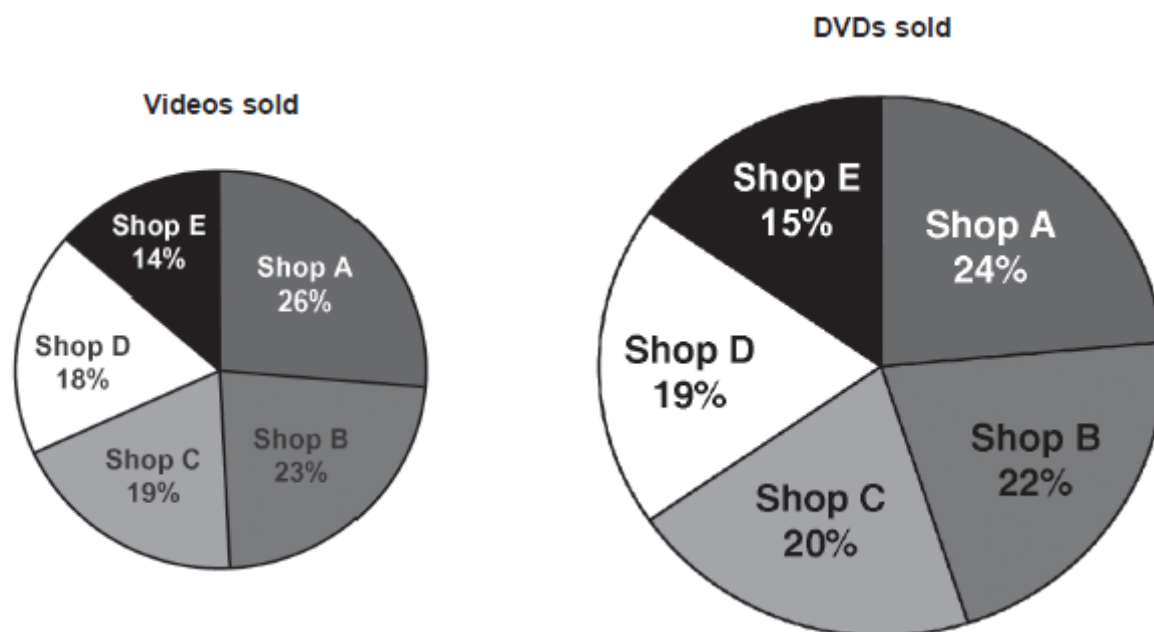
The diagram shows that a woman with a systolic blood pressure of 120 and a cholesterol level of 4 has a 10-year risk of heart disease between 5% and 10% if she is 50 years old and a smoker.

- (a) What is the risk if
- the age of the woman is 50 and she is a non-smoker,
 - the age of the woman is 60 and she is a smoker?
- (b) Is the 10-year risk greater for men or for women? Give a reason for your answer.
- (c) Write down two things that the diagram shows about the effects of age and smoking on the risk of coronary heart disease.

COMPARATIVE PIE CHARTS

6. Five shops in Whitehaven sell both videos and DVDs.

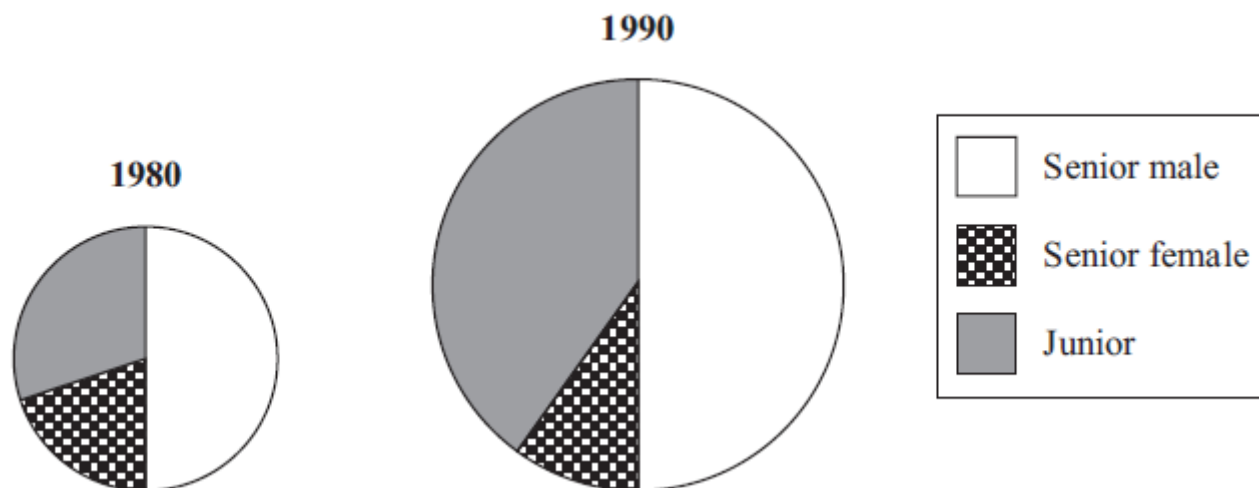
The comparative pie charts show the percentage of each that were sold by the 5 shops in one week.



- (a) Comment on the sales of videos compared to DVDs in Whitehaven. Give reasons for your opinion.
- (b) Discuss any relationship between the percentages of videos sold and the percentages of DVDs sold in each of the 5 shops.

2. The comparative pie charts show some information about the players at Seaton squash club in 1980 and in 1990.

The three types of players at Seaton squash club are Senior male, Senior female and Junior.



- (a) What has happened to the number of Senior male players at Seaton squash club between 1980 and 1990? Give a reason for your answer.

The table shows the numbers of players in 2005.

	Number of players
Senior male	197
Senior female	108
Junior	81

A stratified sample of 40 players was taken from the 386 players of the club.

- (b) Explain why a **stratified** sample was taken.
- (c) Work out the number of Junior players who were selected for the stratified sample.

END OF QUESTIONS