AQA Maths Statistics 1

Past Paper Pack

2006-2015

General Certificate of Education
January 2006
Advanced Subsidiary Examination

MATHEMATICS **MS/SS1B**
Unit Statistics 1B

STATISTICS
Unit Statistics 1B

Thursday 12 January 2006 1.30 pm to 3.00 pm

For this paper you must have:
- an 8-page answer book
- the **blue** AQA booklet of formulae and statistical tables
- an insert for use in Question 5 (enclosed)

You may use a graphics calculator.

Time allowed: 1 hour 30 minutes

Instructions
- Use blue or black ink or ball-point pen. Pencil should only be used for drawing.
- Write the information required on the front of your answer book. The *Examining Body* for this paper is AQA. The *Paper Reference* is MS/SS1B.
- Answer **all** questions.
- All necessary working should be shown; otherwise marks for method may be lost.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.
- Fill in the boxes at the top of the insert.

Information
- The maximum mark for this paper is 75.
- The marks for questions are shown in brackets.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, formulae may be quoted, without proof, from the booklet.

Answer **all** questions.

1 At a certain small restaurant, the waiting time is defined as the time between sitting down at a table and a waiter first arriving at the table. This waiting time is dependent upon the number of other customers already seated in the restaurant.

Alex is a customer who visited the restaurant on 10 separate days. The table shows, for each of these days, the number, x, of customers already seated and his waiting time, y minutes.

x	9	3	4	10	8	12	7	11	2	6
y	11	6	5	11	9	13	9	12	4	7

(a) Calculate the equation of the least squares regression line of y on x in the form $y = a + bx$. *(4 marks)*

(b) Give an interpretation, in context, for each of your values of a and b. *(2 marks)*

(c) Use your regression equation to estimate Alex's waiting time when the number of customers already seated in the restaurant is:

(i) 5;

(ii) 25. *(2 marks)*

(d) Comment on the likely reliability of **each** of your estimates in part (c), given that, for the regression line calculated in part (a), the values of the 10 residuals lie between +1.1 minutes and −1.1 minutes. *(3 marks)*

2 Xavier, Yuri and Zara attend a sports centre for their judo club's practice sessions. The probabilities of them arriving late are, independently, 0.3, 0.4 and 0.2 respectively.

(a) Calculate the probability that for a particular practice session:

(i) all three arrive late; *(1 mark)*

(ii) none of the three arrives late; *(2 marks)*

(iii) only Zara arrives late. *(2 marks)*

(b) Zara's friend, Wei, also attends the club's practice sessions. The probability that Wei arrives late is 0.9 when Zara arrives late, and is 0.25 when Zara does not arrive late.

Calculate the probability that for a particular practice session:

(i) both Zara and Wei arrive late; *(2 marks)*

(ii) either Zara or Wei, but not both, arrives late. *(3 marks)*

3 When an alarm is raised at a market town's fire station, the fire engine cannot leave until at least five fire-fighters arrive at the station. The call-out time, X minutes, is the time between an alarm being raised and the fire engine leaving the station.

The value of X was recorded on a random sample of 50 occasions. The results are summarised below, where $\bar{x}$ denotes the sample mean.

$$\sum x = 286.5 \qquad \sum (x - \bar{x})^2 = 45.16$$

(a) Find values for the mean and standard deviation of this sample of 50 call-out times. *(2 marks)*

(b) Hence construct a 99% confidence interval for the mean call-out time. *(4 marks)*

(c) The fire and rescue service claims that the station's mean call-out time is less than 5 minutes, whereas a parish councillor suggests that it is more than $6\frac{1}{2}$ minutes.

Comment on **each** of these claims. *(2 marks)*

4 The time, x seconds, spent by each of a random sample of 100 customers at an automatic teller machine (ATM) is recorded. The times are summarised in the table.

Time (seconds)	**Number of customers**
$20 < x \leqslant 30$	2
$30 < x \leqslant 40$	7
$40 < x \leqslant 60$	18
$60 < x \leqslant 80$	27
$80 < x \leqslant 100$	23
$100 < x \leqslant 120$	13
$120 < x \leqslant 150$	7
$150 < x \leqslant 180$	3
Total	**100**

(a) Calculate estimates for the mean and standard deviation of the time spent at the ATM by a customer. *(4 marks)*

(b) The mean time spent at the ATM by a random sample of **36** customers is denoted by $\overline{Y}$.

(i) State why the distribution of $\overline{Y}$ is approximately normal. *(1 mark)*

(ii) Write down estimated values for the mean and standard error of $\overline{Y}$. *(2 marks)*

(iii) Hence estimate the probability that $\overline{Y}$ is less than $1\frac{1}{2}$ minutes. *(3 marks)*

5 [**Figure 1**, printed on the insert, is provided for use in this question.]

The table shows the times, in seconds, taken by a random sample of 10 boys from a junior swimming club to swim 50 metres freestyle and 50 metres backstroke.

Boy	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Freestyle (x seconds)	30.2	32.8	25.1	31.8	31.2	35.6	32.4	38.0	36.1	34.1
Backstroke (y seconds)	33.5	35.4	37.4	27.2	34.7	38.2	37.7	41.4	42.3	38.4

(a) On **Figure 1**, complete the scatter diagram for these data. *(2 marks)*

(b) Hence:

(i) give **two** distinct comments on what your scatter diagram reveals; *(2 marks)*

(ii) state, **without calculation**, which of the following 3 values is most likely to be the value of the product moment correlation coefficient for the data in your scatter diagram.

0.912 0.088 0.462 *(1 mark)*

(c) In the sample of 10 boys, one boy is a junior-champion freestyle swimmer and one boy is a junior-champion backstroke swimmer.

Identify the **two** most likely boys. *(2 marks)*

(d) **Removing the data for the two boys whom you identified in part (c):**

(i) calculate the value of the product moment correlation coefficient for the remaining 8 pairs of values of x and y; *(3 marks)*

(ii) comment, in context, on the value that you obtain. *(1 mark)*

6 Plastic clothes pegs are made in various colours.

The number of red pegs may be modelled by a binomial distribution with parameter p equal to 0.2.

The contents of packets of 50 pegs of mixed colours may be considered to be random samples.

(a) Determine the probability that a packet contains:

(i) less than or equal to 15 red pegs; *(2 marks)*

(ii) exactly 10 red pegs; *(2 marks)*

(iii) more than 5 but fewer than 15 red pegs. *(3 marks)*

(b) Sly, a student, claims to have counted the number of red pegs in each of 100 packets of 50 pegs. From his results the following values are calculated.

Mean number of red pegs per packet = 10.5

Variance of number of red pegs per packet = 20.41

Comment on the validity of Sly's claim. *(4 marks)*

7 (a) The weight, X grams, of soup in a carton may be modelled by a normal random variable with mean 406 and standard deviation 4.2.

Find the probability that the weight of soup in a carton:

(i) is less than 400 grams; *(3 marks)*

(ii) is between 402.5 grams and 407.5 grams. *(4 marks)*

(b) The weight, Y grams, of chopped tomatoes in a tin is a normal random variable with mean μ and standard deviation σ.

(i) Given that $\mathrm{P}(Y < 310) = 0.975$, explain why:

$$310 - \mu = 1.96\sigma$$

(3 marks)

(ii) Given that $\mathrm{P}(Y < 307.5) = 0.86$, find, to two decimal places, values for μ and σ. *(4 marks)*

END OF QUESTIONS

Surname		Other Names	
Centre Number		Candidate Number	
Candidate Signature			

General Certificate of Education
January 2006
Advanced Subsidiary Examination

MATHEMATICS **MS/SS1B**
Unit Statistics 1B

STATISTICS
Unit Statistics 1B

Insert

Thursday 12 January 2006 1.30 pm to 3.00 pm

Insert for use in **Question 5**.

Fill in the boxes at the top of this page.

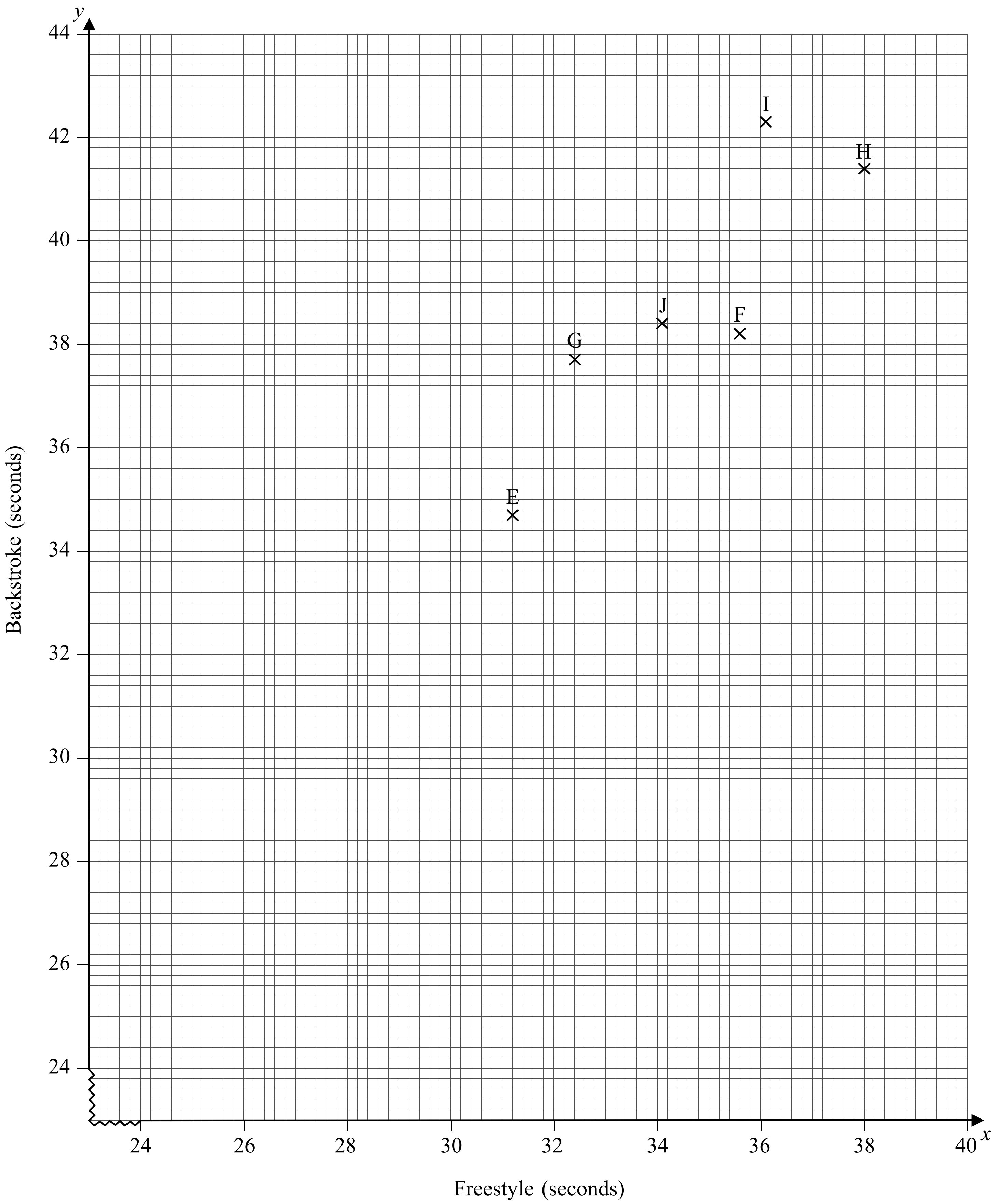
Fasten this insert securely to your answer book.

Turn over for Figure 1

Turn over ▶

Figure 1 (for use in Question 5)

Scatter Diagram for Freestyle and Backstroke Swimming Times



General Certificate of Education
June 2006
Advanced Subsidiary Examination

MATHEMATICS **MS/SS1B**
Unit Statistics 1B

STATISTICS
Unit Statistics 1B

Wednesday 24 May 2006 1.30 pm to 3.00 pm

For this paper you must have: • an 8-page answer book • the **blue** AQA booklet of formulae and statistical tables You may use a graphics calculator.

Time allowed: 1 hour 30 minutes

Instructions
- Use blue or black ink or ball-point pen. Pencil should only be used for drawing.
- Write the information required on the front of your answer book. The *Examining Body* for this paper is AQA. The *Paper Reference* is MS/SS1B.
- Answer **all** questions.
- Show all necessary working; otherwise marks for method may be lost.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

Information
- The maximum mark for this paper is 75.
- The marks for questions are shown in brackets.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.

Answer **all** questions.

1 The table shows, for each of a random sample of 8 paperback fiction books, the number of pages, x, and the recommended retail price, £y, to the nearest 10p.

x	223	276	374	433	564	612	704	766
y	6.50	4.00	5.50	8.00	4.50	5.00	8.00	5.50

(a) (i) Calculate the value of the product moment correlation coefficient between x and y. *(3 marks)*

(ii) Interpret your value in the context of this question. *(2 marks)*

(iii) Suggest one other variable, in addition to the number of pages, which may affect the recommended retail price of a paperback fiction book. *(1 mark)*

(b) The same 8 books were later included in a book sale. The value of the product moment correlation coefficient between the number of pages and the sale price was 0.959, correct to three decimal places.

What can be concluded from this value? *(2 marks)*

2 The heights of sunflowers may be assumed to be normally distributed with a mean of 185 cm and a standard deviation of 10 cm.

(a) Determine the probability that the height of a randomly selected sunflower:

(i) is less than 200 cm; *(3 marks)*

(ii) is more than 175 cm; *(3 marks)*

(iii) is between 175 cm and 200 cm. *(2 marks)*

(b) Determine the probability that the mean height of a random sample of 4 sunflowers is more than 190 cm. *(4 marks)*

3 A new car tyre is fitted to a wheel. The tyre is inflated to its recommended pressure of 265 kPa and the wheel left unused. At 3-month intervals thereafter, the tyre pressure is measured with the following results:

Time after fitting (x months)	0	3	6	9	12	15	18	21	24
Tyre pressure (y kPa)	265	250	240	235	225	215	210	195	180

(a) (i) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(ii) Interpret in context the value for the gradient of your line. *(2 marks)*

(iii) Comment on the value for the intercept with the y-axis of your line. *(2 marks)*

(b) The tyre manufacturer states that, when one of these new tyres is fitted to the wheel of a car and then inflated to 265 kPa, a suitable regression equation is of the form

$$y = 265 + bx$$

The manufacturer also states that, as the car is used, the tyre pressure will decrease at twice the rate of that found in part (a).

(i) Suggest a suitable value for b. *(2 marks)*

(ii) One of these new tyres is fitted to the wheel of a car and inflated to 265 kPa. The car is then used for 8 months, after which the tyre pressure is checked for the first time.

Show that, accepting the manufacturer's statements, the tyre pressure can be expected to have fallen below its minimum safety value of 220 kPa. *(2 marks)*

Turn over for the next question

4 The weights of packets of sultanas may be assumed to be normally distributed with a standard deviation of 6 grams.

The weights of a random sample of 10 packets were as follows:

498	496	499	511	503	505	510	509	513	508

(a) (i) Construct a 99% confidence interval for the mean weight of packets of sultanas, giving the limits to one decimal place. *(5 marks)*

(ii) State why, in calculating your confidence interval, use of the Central Limit Theorem was **not** necessary. *(1 mark)*

(iii) On each packet it states 'Contents 500 grams'.

Comment on this statement using **both** the given sample **and** your confidence interval. *(3 marks)*

(b) Given that the mean weight of all packets of sultanas is 500 grams, state the probability that a 99% confidence interval for the mean, calculated from a random sample of packets, will **not** contain 500 grams. *(1 mark)*

5 Kirk and Les regularly play each other at darts.

(a) The probability that Kirk wins any game is 0.3, and the outcome of each game is independent of the outcome of every other game.

Find the probability that, in a match of 15 games, Kirk wins:

(i) exactly 5 games; *(3 marks)*

(ii) fewer than half of the games; *(3 marks)*

(iii) more than 2 but fewer than 7 games. *(3 marks)*

(b) Kirk attends darts coaching sessions for three months. He then claims that he has a probability of 0.4 of winning any game, and that the outcome of each game is independent of the outcome of every other game.

(i) Assuming this claim to be true, calculate the mean and standard deviation for the number of games won by Kirk in a match of 15 games. *(3 marks)*

(ii) To assess Kirk's claim, Les keeps a record of the number of games won by Kirk in a series of 10 matches, each of 15 games, with the following results:

8 5 6 3 9 12 4 2 6 5

Calculate the mean and standard deviation of these values. *(2 marks)*

(iii) Hence comment on the validity of Kirk's claim. *(3 marks)*

Turn over for the next question

6 A housing estate consists of 320 houses: 120 detached and 200 semi-detached. The numbers of children living in these houses are shown in the table.

	Number of children				
	None	**One**	**Two**	**At least three**	**Total**
Detached house	24	32	41	23	120
Semi-detached house	40	37	88	35	200
Total	64	69	129	58	320

A house on the estate is selected at random.

D denotes the event 'the house is detached'.
R denotes the event 'no children live in the house'.
S denotes the event 'one child lives in the house'.
T denotes the event 'two children live in the house'.
(D' denotes the event 'not D'.)

(a) Find:

(i) $P(D)$; *(1 mark)*

(ii) $P(D \cap R)$; *(1 mark)*

(iii) $P(D \cup T)$; *(2 marks)*

(iv) $P(D \mid R)$; *(2 marks)*

(v) $P(R \mid D')$. *(3 marks)*

(b) (i) Name two of the events D, R, S and T that are mutually exclusive. *(1 mark)*

(ii) Determine whether the events D and R are independent. Justify your answer. *(2 marks)*

(c) Define, in the context of this question, the event:

(i) $D' \cup T$; *(2 marks)*

(ii) $D \cap (R \cup S)$. *(2 marks)*

END OF QUESTIONS

General Certificate of Education
January 2007
Advanced Subsidiary Examination

MATHEMATICS **MS/SS1B**
Unit Statistics 1B

STATISTICS
Unit Statistics 1B

Tuesday 23 January 2007 1.30 pm to 3.00 pm

For this paper you must have:
- an 8-page answer book
- the **blue** AQA booklet of formulae and statistical tables
- an insert for use in Question 7 (enclosed).

You may use a graphics calculator.

Time allowed: 1 hour 30 minutes

Instructions
- Use blue or black ink or ball-point pen. Pencil should only be used for drawing.
- Write the information required on the front of your answer book. The *Examining Body* for this paper is AQA. The *Paper Reference* is MS/SS1B.
- Answer **all** questions.
- Show all necessary working; otherwise marks for method may be lost.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.
- Fill in the boxes at the top of the insert.

Information
- The maximum mark for this paper is 75.
- The marks for questions are shown in brackets.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.

Answer **all** questions.

1 The times, in seconds, taken by 20 people to solve a simple numerical puzzle were

17	19	22	26	28	31	34	36	38	39
41	42	43	47	50	51	53	55	57	58

(a) Calculate the mean and the standard deviation of these times. *(3 marks)*

(b) In fact, 23 people solved the puzzle. However, 3 of them failed to solve it within the allotted time of 60 seconds.

Calculate the median and the interquartile range of the times taken by all 23 people. *(4 marks)*

(c) For the times taken by all 23 people, explain why:

(i) the mode is **not** an appropriate numerical measure;

(ii) the range is **not** an appropriate numerical measure. *(2 marks)*

2 A hotel has 50 single rooms, 16 of which are on the ground floor. The hotel offers guests a choice of a full English breakfast, a continental breakfast or no breakfast. The probabilities of these choices being made are 0.45, 0.25 and 0.30 respectively. It may be assumed that the choice of breakfast is independent from guest to guest.

(a) On a particular morning there are 16 guests, each occupying a single room on the ground floor. Calculate the probability that exactly 5 of these guests require a full English breakfast. *(3 marks)*

(b) On a particular morning when there are 50 guests, each occupying a single room, determine the probability that:

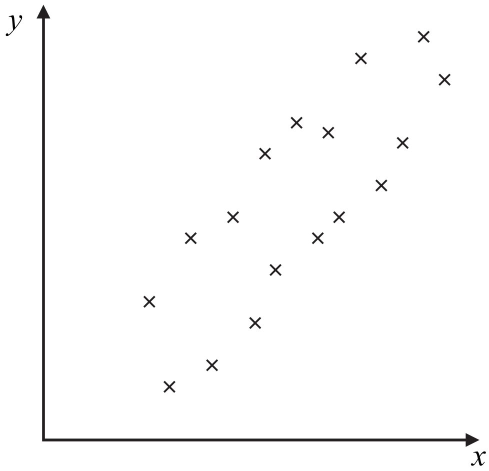
(i) at most 12 of these guests require a continental breakfast; *(2 marks)*

(ii) more than 10 but fewer than 20 of these guests require no breakfast. *(3 marks)*

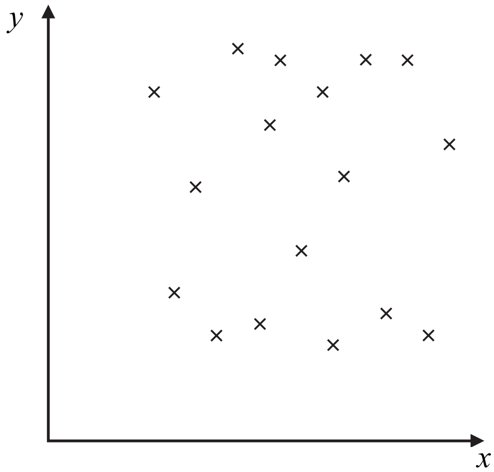
(c) When there are 40 guests, each occupying a single room, calculate the mean and the standard deviation for the number of these guests requiring breakfast. *(4 marks)*

3 Estimate, **without undertaking any calculations**, the value of the product moment correlation coefficient between the variables x and y in each of the three scatter diagrams.

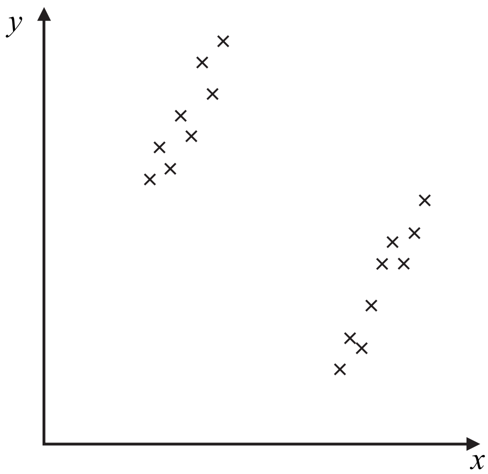
(a)



(b)



(c)



(5 marks)

Turn over for the next question

4 A very popular play has been performed at a London theatre on each of 6 evenings per week for about a year. Over the past 13 weeks (78 performances), records have been kept of the proceeds from the sales of programmes at each performance. An analysis of these records has found that the mean was £184 and the standard deviation was £32.

(a) Assuming that the 78 performances may be considered to be a random sample, construct a 90% confidence interval for the mean proceeds from the sales of programmes at an evening performance of this play. *(4 marks)*

(b) Comment on the likely validity of the assumption in part (a) when constructing a confidence interval for the mean proceeds from the sales of programmes at an evening performance of:

(i) this particular play;

(ii) any play. *(3 marks)*

5 Dafydd, Eli and Fabio are members of an amateur cycling club that holds a time trial each Sunday during the summer. The independent probabilities that Dafydd, Eli and Fabio take part in any one of these trials are 0.6, 0.7 and 0.8 respectively.

Find the probability that, on a particular Sunday during the summer:

(a) none of the three cyclists takes part; *(2 marks)*

(b) Fabio is the only one of the three cyclists to take part; *(2 marks)*

(c) exactly one of the three cyclists takes part; *(3 marks)*

(d) either one or two of the three cyclists take part. *(3 marks)*

6 When Monica walks to work from home, she uses either route A or route B.

(a) Her journey time, X minutes, by route A may be assumed to be normally distributed with a mean of 37 and a standard deviation of 8.

Determine:

(i) $\mathrm{P}(X < 45)$; *(3 marks)*

(ii) $\mathrm{P}(30 < X < 45)$. *(3 marks)*

(b) Her journey time, Y minutes, by route B may be assumed to be normally distributed with a mean of 40 and a standard deviation of σ.

Given that $\mathrm{P}(Y > 45) = 0.12$, calculate the value of σ. *(4 marks)*

(c) If Monica leaves home at 8.15 am to walk to work hoping to arrive by 9.00 am, state, with a reason, which route she should take. *(2 marks)*

(d) When Monica travels to work from home by car, her journey time, W minutes, has a mean of 18 and a standard deviation of 12.

Estimate the probability that, for a random sample of 36 journeys to work from home by car, Monica's mean time is more than 20 minutes. *(4 marks)*

(e) Indicate where, if anywhere, in this question you needed to make use of the Central Limit Theorem. *(1 mark)*

Turn over for the next question

7 [**Figure 1**, printed on the insert, is provided for use in this question.]

Stan is a retired academic who supplements his pension by mowing lawns for customers who live nearby.

As part of a review of his charges for this work, he measures the areas, $x\,\text{m}^2$, of a random sample of eight of his customers' lawns and notes the times, y minutes, that it takes him to mow these lawns. His results are shown in the table.

Customer	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
x	360	140	860	600	1180	540	260	480
y	50	25	135	70	140	90	55	70

(a) On **Figure 1**, plot a scatter diagram of these data. *(2 marks)*

(b) Calculate the equation of the least squares regression line of y on x. Draw your line on **Figure 1**. *(6 marks)*

(c) Calculate the value of the residual for Customer H and indicate how your value is confirmed by your scatter diagram. *(3 marks)*

(d) Given that Stan charges £12 per hour, estimate the charge for mowing a customer's lawn that has an area of $560\,\text{m}^2$. *(4 marks)*

END OF QUESTIONS

Surname		Other Names							
Centre Number						Candidate Number			
Candidate Signature									

General Certificate of Education
January 2007
Advanced Subsidiary Examination

MATHEMATICS **MS/SS1B**
Unit Statistics 1B

STATISTICS
Unit Statistics 1B

Insert

Insert for use in **Question 7**.

Fill in the boxes at the top of this page.

Fasten this insert securely to your answer book.

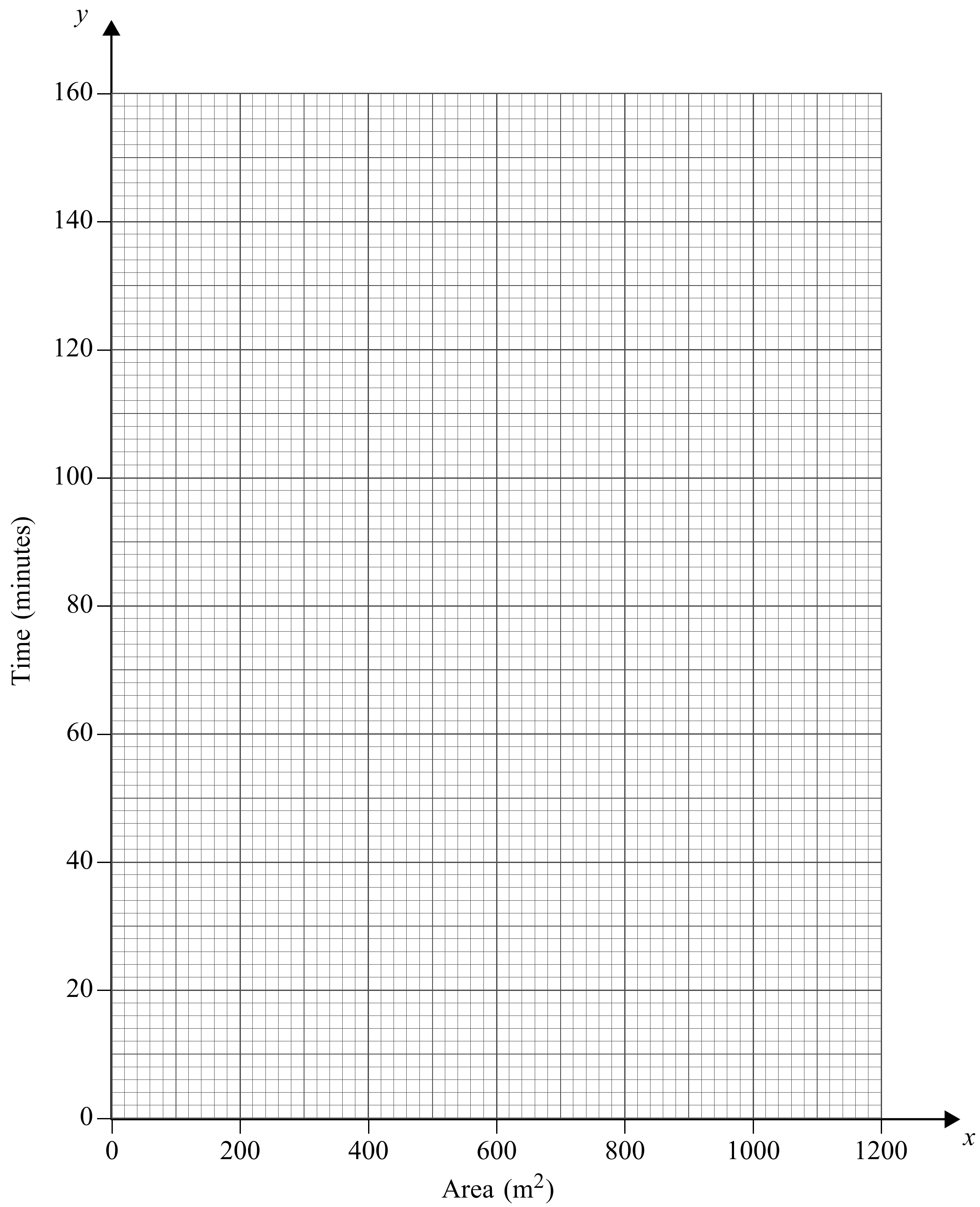
Turn over for Figure 1

Turn over ▶

P89732/Jan07/MS/SS1B 6/6/6/

Figure 1 (for use in Question 7)

Lawn Areas and Mowing Times



General Certificate of Education
June 2007
Advanced Subsidiary Examination

MATHEMATICS **MS/SS1B**
Unit Statistics 1B

STATISTICS
Unit Statistics 1B

Thursday 14 June 2007 1.30 pm to 3.00 pm

For this paper you must have: • an 8-page answer book • the **blue** AQA booklet of formulae and statistical tables. You may use a graphics calculator.

Time allowed: 1 hour 30 minutes

Instructions
- Use blue or black ink or ball-point pen. Pencil should only be used for drawing.
- Write the information required on the front of your answer book. The *Examining Body* for this paper is AQA. The *Paper Reference* is MS/SS1B.
- Answer **all** questions.
- Show all necessary working; otherwise marks for method may be lost.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

Information
- The maximum mark for this paper is 75.
- The marks for questions are shown in brackets.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.

Answer **all** questions.

1 The table shows the length, in centimetres, and maximum diameter, in centimetres, of each of 10 honeydew melons selected at random from those on display at a market stall.

Length	24	25	19	28	27	21	35	23	32	26
Maximum diameter	18	14	16	11	13	14	12	16	15	14

(a) Calculate the value of the product moment correlation coefficient. *(3 marks)*

(b) Interpret your value in the context of this question. *(2 marks)*

2 The British and Irish Lions 2005 rugby squad contained 50 players. The nationalities and playing positions of these players are shown in the table.

		Nationality			
		English	**Welsh**	**Scottish**	**Irish**
Playing position	**Forward**	14	5	2	6
	Back	8	7	2	6

(a) A player was selected at random from the squad for a radio interview. Calculate the probability that the player was:

(i) a Welsh back; *(1 mark)*

(ii) English; *(2 marks)*

(iii) not English; *(1 mark)*

(iv) Irish, given that the player was a back; *(2 marks)*

(v) a forward, given that the player was not Scottish. *(2 marks)*

(b) Four players were selected at random from the squad to visit a school. Calculate the probability that all four players were English. *(3 marks)*

3 (a) A sample of 50 washed baking potatoes was selected at random from a large batch. The weights of the 50 potatoes were found to have a mean of 234 grams and a standard deviation of 25.1 grams.

Construct a 95% confidence interval for the mean weight of potatoes in the batch. *(4 marks)*

(b) The batch of potatoes is purchased by a market stallholder. He sells them to his customers by allowing them to choose any 5 potatoes for £1.

Give a reason why such chosen potatoes are unlikely to represent a random sample from the batch. *(1 mark)*

4 A library allows each member to have up to 15 books on loan at any one time.

The table shows the numbers of books currently on loan to a random sample of 95 members of the library.

Number of books on loan	0	1	2	3	4	5–9	10–14	15
Number of members	4	13	24	17	15	11	5	6

(a) For these data:

(i) state values for the mode and range; *(2 marks)*

(ii) determine values for the median and interquartile range; *(4 marks)*

(iii) calculate estimates of the mean and standard deviation. *(4 marks)*

(b) Making reference to your answers to part (a), give a reason for preferring:

(i) the median and interquartile range to the mean and standard deviation for summarising the given data; *(1 mark)*

(ii) the mean and standard deviation to the mode and range for summarising the given data. *(1 mark)*

5 Bob, a gardener, measures the time taken, y minutes, for 60 grams of weedkiller pellets to dissolve in 10 litres of water at different set temperatures, x °C. His results are shown in the table.

x	16	20	24	28	32	36	40	44	48	52	56
y	4.7	4.3	3.8	3.5	3.0	2.7	2.4	2.0	1.8	1.6	1.1

(a) State why the explanatory variable is temperature. *(1 mark)*

(b) Calculate the equation of the least squares regression line $y = a + bx$. *(4 marks)*

(c) (i) Interpret, in the context of this question, your value for b. *(2 marks)*

(ii) Explain why no sensible practical interpretation can be given for your value of a. *(2 marks)*

(d) (i) Estimate the time taken to dissolve 60 grams of weedkiller pellets in 10 litres of water at 30 °C. *(2 marks)*

(ii) Show why the equation cannot be used to make a valid estimate of the time taken to dissolve 60 grams of weedkiller pellets in 10 litres of water at 75 °C. *(2 marks)*

6 Each weekday, Monday to Friday, Trina catches a train from her local station. She claims that the probability that the train arrives on time at the station is 0.4, and that the train's arrival time is independent from day to day.

(a) Assuming her claims to be true, determine the probability that the train arrives on time at the station:

(i) on at most 3 days during a 2-week period (10 days); *(2 marks)*

(ii) on more than 10 days but fewer than 20 days during an 8-week period. *(3 marks)*

(b) (i) Assuming Trina's claims to be true, determine the mean and standard deviation for the number of times during a week (5 days) that the train arrives on time at the station. *(3 marks)*

(ii) Each week, for a period of 13 weeks, Trina's travelling colleague, Suzie, records the number of times that the train arrives on time at the station. Suzie's results are

2 2 4 1 2 3 3 2 2 0 3 2 0

Calculate the mean and standard deviation of these values. *(3 marks)*

(iii) Hence comment on the likely validity of Trina's claims. *(2 marks)*

7 (a) Electra is employed by E & G Ltd to install electricity meters in new houses on an estate. Her time, X minutes, to install a meter may be assumed to be normally distributed with a mean of 48 and a standard deviation of 20.

Determine:

(i) $\mathrm{P}(X < 60)$; *(2 marks)*

(ii) $\mathrm{P}(30 < X < 60)$; *(3 marks)*

(iii) the time, k minutes, such that $\mathrm{P}(X < k) = 0.9$. *(4 marks)*

(b) Gazali is employed by E & G Ltd to install gas meters in the same new houses. His time, Y minutes, to install a meter has a mean of 37 and a standard deviation of 25.

(i) Explain why Y is unlikely to be normally distributed. *(2 marks)*

(ii) State why $\overline{Y}$, the mean of a random sample of 35 gas meter installations, is likely to be approximately normally distributed. *(1 mark)*

(iii) Determine $\mathrm{P}(\overline{Y} > 40)$. *(4 marks)*

END OF QUESTIONS

General Certificate of Education
January 2008
Advanced Subsidiary Examination

ASSESSMENT and QUALIFICATIONS ALLIANCE

MATHEMATICS **MS/SS1B**
Unit Statistics 1B

STATISTICS
Unit Statistics 1B

Tuesday 22 January 2008 1.30 pm to 3.00 pm

For this paper you must have:
- an 8-page answer book
- the blue AQA booklet of formulae and statistical tables
- an insert for use in Question 4 (enclosed).

You may use a graphics calculator.

Time allowed: 1 hour 30 minutes

Instructions
- Use blue or black ink or ball-point pen. Pencil should only be used for drawing.
- Write the information required on the front of your answer book. The *Examining Body* for this paper is AQA. The *Paper Reference* is MS/SS1B.
- Answer **all** questions.
- Show all necessary working; otherwise marks for method may be lost.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.
- Fill in the boxes at the top of the insert.

Information
- The maximum mark for this paper is 75.
- The marks for questions are shown in brackets.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.

Answer **all** questions.

1 In large-scale tree-felling operations, a machine cuts down trees, strips off the branches and then cuts the trunks into logs of length X metres for transporting to a sawmill.

It may be assumed that values of X are normally distributed with mean μ and standard deviation 0.16, where μ can be set to a specific value.

(a) Given that μ is set to 3.3, determine:

(i) $P(X < 3.5)$; *(3 marks)*

(ii) $P(X > 3.0)$; *(3 marks)*

(iii) $P(3.0 < X < 3.5)$. *(2 marks)*

(b) The sawmill now requires a batch of logs such that there is a probability of 0.025 that any given log will have a length less than 3.1 metres.

Determine, to two decimal places, the new value of μ. *(4 marks)*

2 The head and body length, x millimetres, and tail length, y millimetres, of each of a sample of 20 adult dormice were measured. The following statistics are derived from the results.

$$S_{xx} = 1280.55 \qquad S_{yy} = 281.8 \qquad S_{xy} = 416.3$$

(a) Calculate the value of the product moment correlation coefficient between x and y. *(2 marks)*

(b) Interpret your value in the context of this question. *(2 marks)*

(c) Write down the value of the product moment correlation coefficient if the measurements had been recorded in centimetres. *(1 mark)*

(d) Give a reason why it is not generally advisable to calculate the value of the product moment correlation coefficient without first viewing a scatter diagram of the data. Illustrate your answer with a sketch. *(2 marks)*

3 The height, in metres, of adult male African elephants may be assumed to be normally distributed with mean μ and standard deviation 0.20.

The heights of a sample of 12 such elephants were measured with the following results, in metres.

3.37 3.45 2.93 3.42 3.49 3.67 2.96 3.57 3.36 2.89 3.22 2.91

(a) Stating a necessary assumption, construct a 98% confidence interval for μ. *(6 marks)*

(b) The mean height of adult male **Asian** elephants is known to be 2.90 metres.

Using your confidence interval, state, with a reason, what can be concluded about the mean heights of adult males in these two types of elephant. *(2 marks)*

4 [**Figure 1**, printed on the insert, is provided for use in this question.]

Roseen is a self-employed decorator who wishes to estimate the times that it will take her to decorate bedrooms based upon their floor areas. She records the floor area, $x\,\text{m}^2$, and the decorating time, y hours, for each of 10 bedrooms she has recently decorated.

x	11.0	22.0	7.5	21.0	13.0	16.5	14.0	16.0	18.5	20.5
y	15.0	35.0	16.0	23.5	24.0	17.5	14.5	27.5	22.5	34.5

(a) On **Figure 1**, plot a scatter diagram of these data. *(2 marks)*

(b) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(c) Draw your regression line on **Figure 1**. *(2 marks)*

(d) (i) Use your regression equation to estimate the time that Roseen will take to decorate a bedroom with a floor area of $15\,\text{m}^2$. *(2 marks)*

(ii) Making reference to **Figure 1**, comment on the likely reliability of your estimate in part (d)(i). *(2 marks)*

Turn over ▶

5 A health club has a number of facilities which include a gym and a sauna. Andrew and his wife, Heidi, visit the health club together on Tuesday evenings.

On any visit, Andrew uses either the gym or the sauna or both, but no other facilities. The probability that he uses the gym, $P(G)$, is 0.70. The probability that he uses the sauna, $P(S)$, is 0.55. The probability that he uses both the gym and the sauna is 0.25.

(a) Calculate the probability that, on a particular visit:

(i) he does not use the gym; *(1 mark)*

(ii) he uses the gym but not the sauna; *(2 marks)*

(iii) he uses either the gym or the sauna but not both. *(2 marks)*

(b) Assuming that Andrew's decision on what facility to use is independent from visit to visit, calculate the probability that, during a month in which there are exactly four Tuesdays, he does not use the gym. *(2 marks)*

(c) The probability that Heidi uses the gym when Andrew uses the gym is 0.6, but is only 0.1 when he does not use the gym.

Calculate the probability that, on a particular visit, Heidi uses the gym. *(3 marks)*

(d) On any visit, Heidi uses **exactly one** of the club's facilities.

The probability that she uses the sauna is 0.35.

Calculate the probability that, on a particular visit, she uses a facility other than the gym or the sauna. *(2 marks)*

6 For each of the Premiership football seasons 2004/05 and 2005/06, a count is made of the number of goals scored in each of the 380 matches. The results are shown in the table.

Number of goals scored in a match	**Number of matches**	
	2004/05	**2005/06**
0	30	32
1	79	82
2	99	95
3	68	78
4	60	48
5	24	30
6	11	9
7	6	6
8	2	0
9	1	0
Total	**380**	**380**

(a) For the number of goals scored in a match during the **2004/05** season:

(i) determine the median and the interquartile range; *(4 marks)*

(ii) calculate the mean and the standard deviation. *(4 marks)*

(b) Two statistics students, Jole and Katie, independently analyse the data on the number of goals scored in a match during the **2005/06** season.

- Jole determines correctly that the median is 2 and that the interquartile range is also 2.
- Katie calculates correctly, to two decimal places, that the mean is 2.48 and that the standard deviation is 1.59.

(i) Use your answers from part (a), together with Jole's and Katie's results, to compare briefly the two seasons with regard to the average and the spread of the number of goals scored in a match. *(2 marks)*

(ii) Jole claims that Katie's results must be wrong as 95% of values always lie within 2 standard deviations of the mean and $(2.48 - 2 \times 1.59) < 0$ which is nonsense.

Explain why Jole's claim is incorrect. (You are not expected to confirm Katie's results.) *(2 marks)*

Turn over ▶

7 A travel agency in Tunisia offers customers a 3-day tour into the Sahara desert by either coach or minibus.

(a) The agency accepts bookings from 50 customers for seats on the coach. The probability that a customer, who has booked a seat on the coach, will **not** turn up to claim the seat is 0.08, and may be assumed to be independent of the behaviour of other customers.

Determine the probability that, of the customers who have booked a seat on the coach:

(i) two or more will **not** turn up;

(ii) three or more will **not** turn up. *(4 marks)*

(b) The agency accepts bookings from 15 customers for seats on the minibus. The probability that a customer, who has booked a seat on the minibus, will **not** turn up to claim the seat is 0.025, and may be assumed to be independent of the behaviour of other customers.

Calculate the probability that, of the customers who have booked a seat on the minibus:

(i) all will turn up;

(ii) one or more will **not** turn up. *(4 marks)*

(c) The coach has 48 seats and the minibus has 14 seats. If 14 or fewer customers who have booked seats on the minibus turn up, they will be allocated a seat on the minibus. If all 15 customers who have booked seats on the minibus turn up, one will be allocated a seat on the coach. This will leave only 47 seats available for the 50 customers who have booked seats on the coach.

Use your results from parts (a) and (b) to calculate the probability that there will be seats available on the coach for all those who turn up having booked such seats. *(4 marks)*

END OF QUESTIONS

Surname		Other Names							
Centre Number						Candidate Number			
Candidate Signature									

General Certificate of Education
January 2008
Advanced Subsidiary Examination

MATHEMATICS **MS/SS1B**
Unit Statistics 1B

STATISTICS
Unit Statistics 1B

Insert

Insert for use in **Question 4**.

Fill in the boxes at the top of this page.

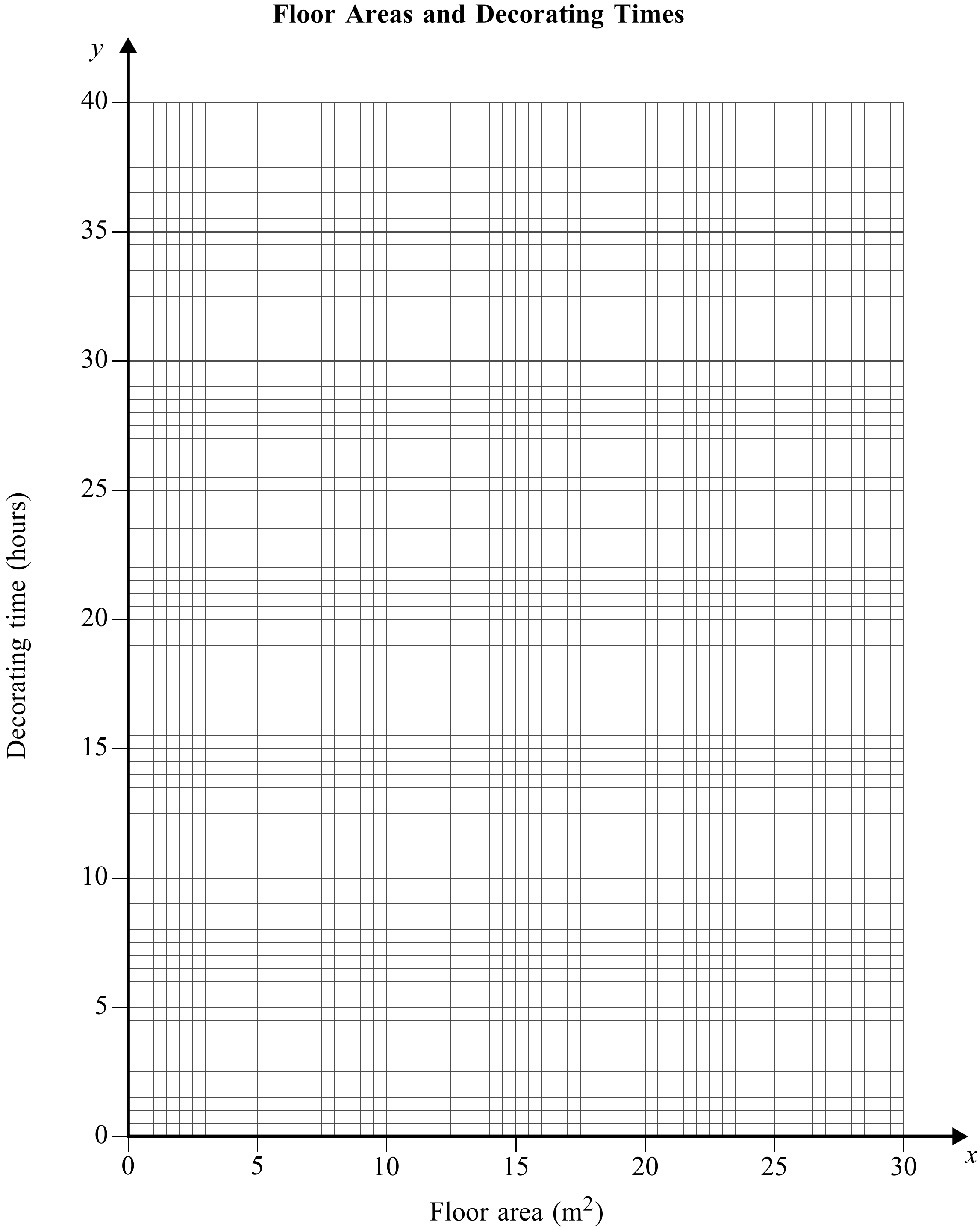
Fasten this insert securely to your answer book.

Turn over for Figure 1

Turn over ▶

P97941/Jan08/MS/SS1B 6/6/6/

Figure 1 (for use in Question 4)



General Certificate of Education
June 2008
Advanced Subsidiary Examination

MATHEMATICS **MS/SS1B**
Unit Statistics 1B

STATISTICS
Unit Statistics 1B

Wednesday 21 May 2008 1.30 pm to 3.00 pm

For this paper you must have:
- an 8-page answer book
- the blue AQA booklet of formulae and statistical tables
- an insert for use in Question 3 (enclosed).

You may use a graphics calculator.

Time allowed: 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Write the information required on the front of your answer book. The *Examining Body* for this paper is AQA. The *Paper Reference* is MS/SS1B.
- Answer **all** questions.
- Show all necessary working; otherwise marks for method may be lost.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.
- Fill in the boxes at the top of the insert.

Information
- The maximum mark for this paper is 75.
- The marks for questions are shown in brackets.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.

Answer **all** questions.

1 The table shows the times taken, y minutes, for a wood glue to dry at different air temperatures, x °C.

x	10	12	15	18	20	22	25	28	30
y	42.9	40.6	38.5	35.4	33.0	30.7	28.0	25.3	22.6

(a) Calculate the equation of the least squares regression line $y = a + bx$. *(4 marks)*

(b) Estimate the time taken for the glue to dry when the air temperature is 21 °C. *(2 marks)*

2 A basket in a stationery store contains a total of 400 marker and highlighter pens. Of the marker pens, some are permanent and the rest are non-permanent. The colours and types of pen are shown in the table.

	Colour			
Type	**Black**	**Blue**	**Red**	**Green**
Permanent marker	44	66	32	18
Non-permanent marker	36	53	21	10
Highlighter	0	41	37	42

A pen is selected at random from the basket. Calculate the probability that it is:

(a) a blue pen; *(1 mark)*

(b) a marker pen; *(2 marks)*

(c) a blue pen or a marker pen; *(2 marks)*

(d) a green pen, given that it is a highlighter pen; *(2 marks)*

(e) a non-permanent marker pen, given that it is a red pen. *(2 marks)*

3 [**Figure 1**, printed on the insert, is provided for use in this question.]

The table shows, for each of a sample of 12 handmade decorative ceramic plaques, the length, x millimetres, and the width, y millimetres.

Plaque	x	y
A	232	109
B	235	112
C	236	114
D	234	118
E	230	117
F	230	113
G	246	121
H	240	125
I	244	128
J	241	122
K	246	126
L	245	123

(a) Calculate the value of the product moment correlation coefficient between x and y. *(3 marks)*

(b) Interpret your value in the context of this question. *(2 marks)*

(c) On **Figure 1**, complete the scatter diagram for these data. *(3 marks)*

(d) In fact, the 6 plaques A, B, ..., F are from a different source to the 6 plaques G, H, ..., L.

With reference to your scatter diagram, **but without further calculations**, estimate the value of the product moment correlation coefficient between x and y for **each** source of plaque. *(2 marks)*

Turn over ▶

4 The runs scored by a cricketer in 11 innings during the 2006 season were as follows.

47 63 0 28 40 51 a 77 0 13 35

The exact value of a was unknown but it was greater than 100.

(a) Calculate the median and the interquartile range of these 11 values. *(4 marks)*

(b) Give a reason why, for these 11 values:

(i) the mode is **not** an appropriate measure of average;

(ii) the range is **not** an appropriate measure of spread. *(2 marks)*

5 When a particular make of tennis ball is dropped from a vertical distance of 250 cm on to concrete, the height, X centimetres, to which it first bounces may be assumed to be normally distributed with a mean of 140 and a standard deviation of 2.5 .

(a) Determine:

(i) $P(X < 145)$; *(3 marks)*

(ii) $P(138 < X < 142)$. *(4 marks)*

(b) Determine, to one decimal place, the maximum height exceeded by 85% of first bounces. *(4 marks)*

(c) Determine the probability that, for a random sample of 4 first bounces, the mean height is greater than 139 cm. *(4 marks)*

6 For the adult population of the UK, 35 per cent of men and 29 per cent of women do not wear glasses or contact lenses.

(a) Determine the probability that, in a random sample of 40 men:

(i) at most 15 do not wear glasses or contact lenses; *(3 marks)*

(ii) more than 10 but fewer than 20 do not wear glasses or contact lenses. *(3 marks)*

(b) Calculate the probability that, in a random sample of 10 women, exactly 3 do not wear glasses or contact lenses. *(3 marks)*

(c) (i) Calculate the mean and the variance for the number who **do** wear glasses or contact lenses in a random sample of 20 women. *(3 marks)*

(ii) The numbers wearing glasses or contact lenses in 10 groups, each of 20 women, had a mean of 16.5 and a variance of 2.50 .

Comment on the claim that these 10 groups were **not** random samples. *(3 marks)*

7 Vernon, a service engineer, is expected to carry out a boiler service in one hour.

One hour is subtracted from each of his actual times, and the resulting differences, x minutes, for a random sample of 100 boiler services are summarised in the table.

Difference	**Frequency**
$-6 \leqslant x < -4$	4
$-4 \leqslant x < -2$	9
$-2 \leqslant x < 0$	13
$0 \leqslant x < 2$	27
$2 \leqslant x < 4$	21
$4 \leqslant x < 6$	15
$6 \leqslant x < 8$	7
$8 \leqslant x \leqslant 10$	4
Total	**100**

(a) (i) Calculate estimates of the mean and the standard deviation of these differences. *(4 marks)*

(ii) Hence deduce, in minutes, estimates of the mean and the standard deviation of Vernon's actual service times for this sample. *(3 marks)*

(b) (i) Construct an approximate 98% confidence interval for the mean time taken by Vernon to carry out a boiler service. *(4 marks)*

(ii) Give a reason why this confidence interval is approximate rather than exact. *(1 mark)*

(c) Vernon claims that, more often than not, a boiler service takes more than an hour and that, on average, a boiler service takes much longer than an hour.

Comment, with a justification, on **each** of these claims. *(2 marks)*

END OF QUESTIONS

Surname		Other Names	
Centre Number		Candidate Number	
Candidate Signature			

General Certificate of Education
June 2008
Advanced Subsidiary Examination

MATHEMATICS **MS/SS1B**
Unit Statistics 1B

STATISTICS
Unit Statistics 1B

Insert

Insert for use in **Question 3**.

Fill in the boxes at the top of this page.

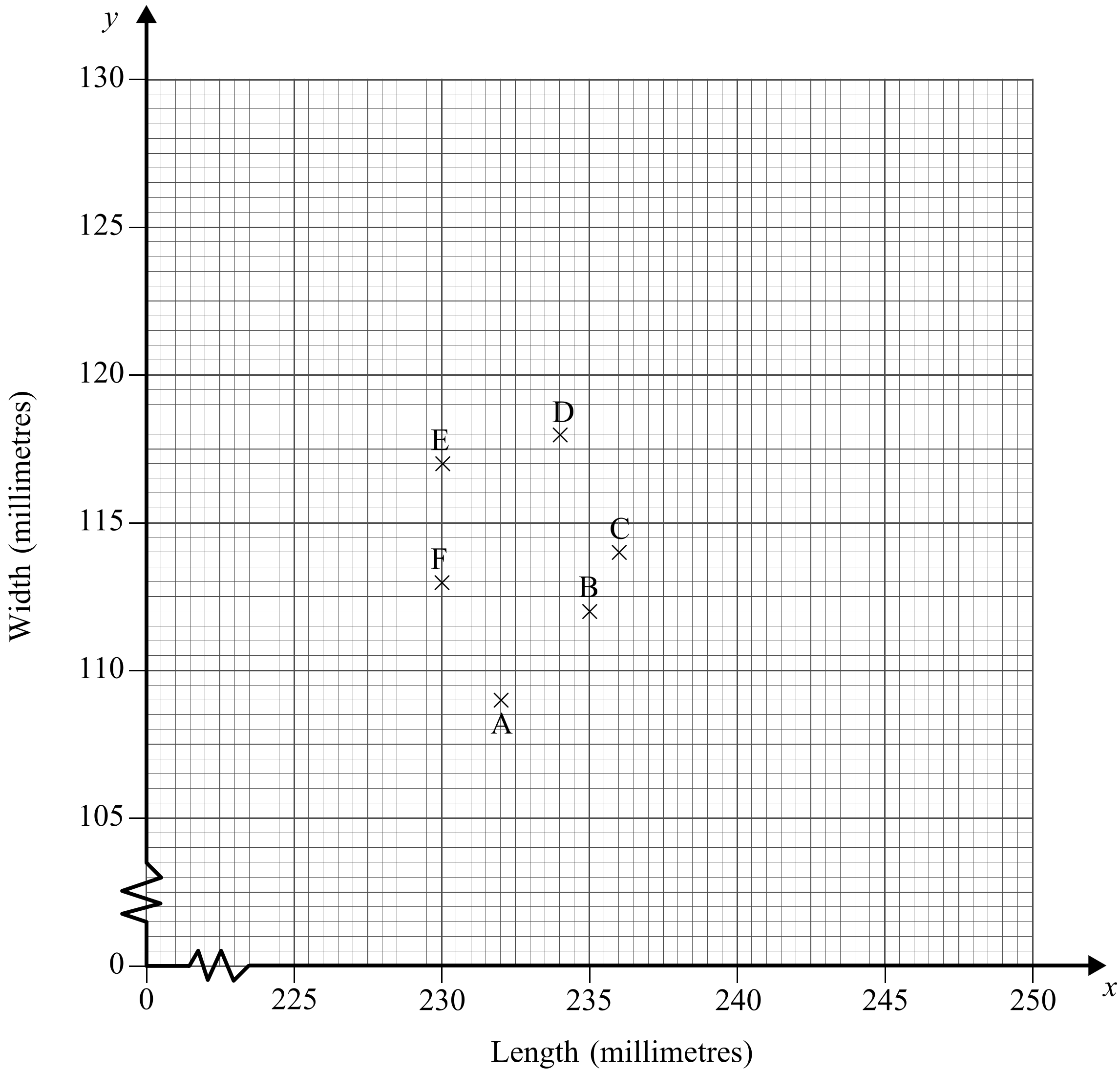
Fasten this insert securely to your answer book.

Turn over for Figure 1

Turn over ▶

Figure 1 (for use in Question 3)

Decorative Plaques



General Certificate of Education
January 2009
Advanced Subsidiary Examination

MATHEMATICS **MS/SS1B**
Unit Statistics 1B

STATISTICS
Unit Statistics 1B

Friday 9 January 2009 9.00 am to 10.30 am

For this paper you must have:
- an 8-page answer book
- the blue AQA booklet of formulae and statistical tables
- an insert for use in Question 6 (enclosed).

You may use a graphics calculator.

Time allowed: 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Write the information required on the front of your answer book. The *Examining Body* for this paper is AQA. The *Paper Reference* is MS/SS1B.
- Answer **all** questions.
- Show all necessary working; otherwise marks for method may be lost.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.
- Fill in the boxes at the top of the insert.

Information
- The maximum mark for this paper is 75.
- The marks for questions are shown in brackets.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.

Answer **all** questions.

1 Ms N Parker always reads the columns of announcements in her local weekly newspaper. During each week of 2008, she notes the number of births announced. Her results are summarised in the table.

Number of births	1	2	3	4	5	6	7	8
Number of weeks	1	2	9	13	7	13	6	1

(a) Calculate the mean, median and modes of these data. *(5 marks)*

(b) State, with a reason, which of the three measures of average in part (a) you consider to be the **least** appropriate for summarising the number of births. *(2 marks)*

2 A greengrocer sells bunches of 9 carrots at his Saturday market stall. Tom and Geri are two Statistics students who work on the stall. Each selects a bunch of carrots at random.

(a) At home, Tom measures the length, x centimetres, and the maximum diameter, y centimetres, of each carrot in his selected bunch with the following results.

x	16.2	13.1	10.4	12.1	14.6	9.7	11.8	13.6	17.3
y	4.2	3.9	4.7	3.3	3.7	2.4	3.1	3.5	2.7

(i) Calculate the value of the product moment correlation coefficient. *(3 marks)*

(ii) Interpret your value in context. *(2 marks)*

(b) At her home, Geri measures the length, in centimetres, and the **weight**, in grams, of each carrot in her selected bunch and then obtains a value of -0.986 for the product moment correlation coefficient.

Comment, with a reason, on the likely validity of Geri's value. *(2 marks)*

3 UPVC facia board is supplied in lengths labelled as 5 metres. The actual length, X metres, of a board may be modelled by a normal distribution with a mean of 5.08 and a standard deviation of 0.05.

(a) Determine:

(i) $\mathrm{P}(X < 5)$; *(3 marks)*

(ii) $\mathrm{P}(5 < X < 5.10)$. *(2 marks)*

(b) Determine the probability that the mean length of a random sample of 4 boards:

(i) exceeds 5.05 metres; *(4 marks)*

(ii) is exactly 5 metres. *(1 mark)*

(c) Assuming that the value of the standard deviation remains unchanged, determine the mean length necessary to ensure that only 1 per cent of boards have lengths less than 5 metres. *(4 marks)*

4 Gary and his neighbour Larry work at the same place.

On any day when Gary travels to work, he uses one of three options: his car only, a bus only or both his car and a bus. The probability that he uses his car, either on its own or with a bus, is 0.6. The probability that he uses both his car and a bus is 0.25.

(a) Calculate the probability that, on any particular day when Gary travels to work, he:

(i) does not use his car; *(1 mark)*

(ii) uses his car only; *(2 marks)*

(iii) uses a bus. *(3 marks)*

(b) On any day, the probability that Larry travels to work with Gary is 0.9 when Gary uses his car only, is 0.7 when Gary uses both his car and a bus, and is 0.3 when Gary uses a bus only.

(i) Calculate the probability that, on any particular day when Gary travels to work, Larry travels with him. *(4 marks)*

(ii) Assuming that option choices are independent from day to day, calculate, to three decimal places, the probability that, during any particular week (5 days) when Gary travels to work every day, Larry never travels with him. *(2 marks)*

Turn over ▶

5 The times taken by new recruits to complete an assault course may be modelled by a normal distribution with a standard deviation of 8 minutes.

A group of 30 new recruits takes a total time of 1620 minutes to complete the course.

(a) Calculate the mean time taken by these 30 new recruits. *(1 mark)*

(b) Assuming that the 30 recruits may be considered to be a random sample, construct a 98% confidence interval for the mean time taken by new recruits to complete the course. *(4 marks)*

(c) Construct an interval within which approximately 98% of the times taken by individual new recruits to complete the course will lie. *(2 marks)*

(d) State where, if at all, in this question you made use of the Central Limit Theorem. *(1 mark)*

6 [**Figure 1**, printed on the insert, is provided for use in this question.]

For a random sample of 10 patients who underwent hip-replacement operations, records were kept of their ages, x years, and of the number of days, y, following their operations before they were able to walk unaided safely.

Patient	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
x	55	51	62	66	72	59	78	55	62	70
y	34	33	39	49	48	43	51	41	46	51

(a) On **Figure 1**, complete the scatter diagram for these data. *(2 marks)*

(b) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(c) Draw your regression line on **Figure 1**. *(2 marks)*

(d) In fact, patients H, I and J were males and the other 7 patients were females.

(i) Calculate the mean of the residuals for the 3 male patients. *(4 marks)*

(ii) Hence estimate, for a male patient aged 65 years, the number of days following his hip-replacement operation before he is able to walk unaided safely. *(3 marks)*

7 The proportion of passengers who use senior citizen bus passes to travel into a particular town on 'Park & Ride' buses between 9.30 am and 11.30 am on weekdays is 0.45.

It is proposed that, when there are n passengers on a bus, a suitable model for the number of passengers using senior citizen bus passes is the distribution $\mathrm{B}(n,\ 0.45)$.

(a) Assuming that this model applies to the 10.30 am weekday 'Park & Ride' bus into the town:

(i) calculate the probability that, when there are **16** passengers, exactly 3 of them are using senior citizen bus passes; *(3 marks)*

(ii) determine the probability that, when there are **25** passengers, fewer than 10 of them are using senior citizen bus passes; *(2 marks)*

(iii) determine the probability that, when there are **40** passengers, at least 15 but at most 20 of them are using senior citizen bus passes; *(3 marks)*

(iv) calculate the mean and the variance for the number of passengers using senior citizen bus passes when there are **50** passengers. *(2 marks)*

(b) (i) Give a reason why the proposed model may not be suitable. *(1 mark)*

(ii) Give a **different** reason why the proposed model would not be suitable for the number of passengers using senior citizen bus passes to travel into the town on the **7.15 am** weekday 'Park & Ride' bus. *(1 mark)*

END OF QUESTIONS

Centre Number						Candidate Number				
Surname										
Other Names										
Candidate Signature										

For Examiner's Use	
Examiner's Initials	
Question	Mark
1	
2	
3	
4	
5	
6	
7	
TOTAL	

General Certificate of Education
Advanced Subsidiary Examination
June 2009

Mathematics
Unit Statistics 1B

MS/SS1B

Statistics
Unit Statistics 1B

Specimen paper for examinations in June 2010 onwards

For this paper you must have:
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed
- 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- Write the question part reference (eg (a), (b)(i) etc) in the left-hand margin.
- You must answer the questions in the space provided. Do not write outside the box around each page.
- Show all necessary working; otherwise marks for method may be lost.
- Do all rough work in this book. Cross through any work that you do not want to be marked.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.

Answer **all** questions in the spaces provided.

1 A large bookcase contains two types of book: hardback and paperback. The number of books of each type in each of four subject categories is shown in the table.

		Subject category				
		Crime	**Romance**	**Science fiction**	**Thriller**	**Total**
Type	**Hardback**	8	16	18	18	60
	Paperback	16	40	14	30	100
	Total	24	56	32	48	160

(a) A book is selected at random from the bookcase. Calculate the probability that the book is:

(i) a paperback; *(1 mark)*

(ii) not science fiction; *(2 marks)*

(iii) science fiction or a hardback; *(2 marks)*

(iv) a thriller, given that it is a paperback. *(2 marks)*

(b) Three books are selected at random, without replacement, from the bookcase.

Calculate, to three decimal places, the probability that one is crime, one is romance and one is science fiction. *(4 marks)*

QUESTION PART REFERENCE

2 Hermione, who is studying reptiles, measures the length, x cm, and the weight, y grams, of a sample of 11 adult snakes of the same type. Her results are shown in the table.

Snake	A	B	C	D	E	F	G	H	I	J	K
x	46	39	54	79	47	58	73	35	43	51	36
y	55	48	58	88	61	55	82	51	50	66	57

(a) Calculate the value of the product moment correlation coefficient, r, between x and y. *(3 marks)*

(b) Interpret your value in context. *(2 marks)*

(c) Complete the scatter diagram, opposite, for these data. *(2 marks)*

(d) Subsequently it is found that, of the 11 adult snakes, 9 are male and 2 are female.

(i) Given that female adult snakes are generally larger than male adult snakes, identify the 2 snakes which are most likely to be female. *(1 mark)*

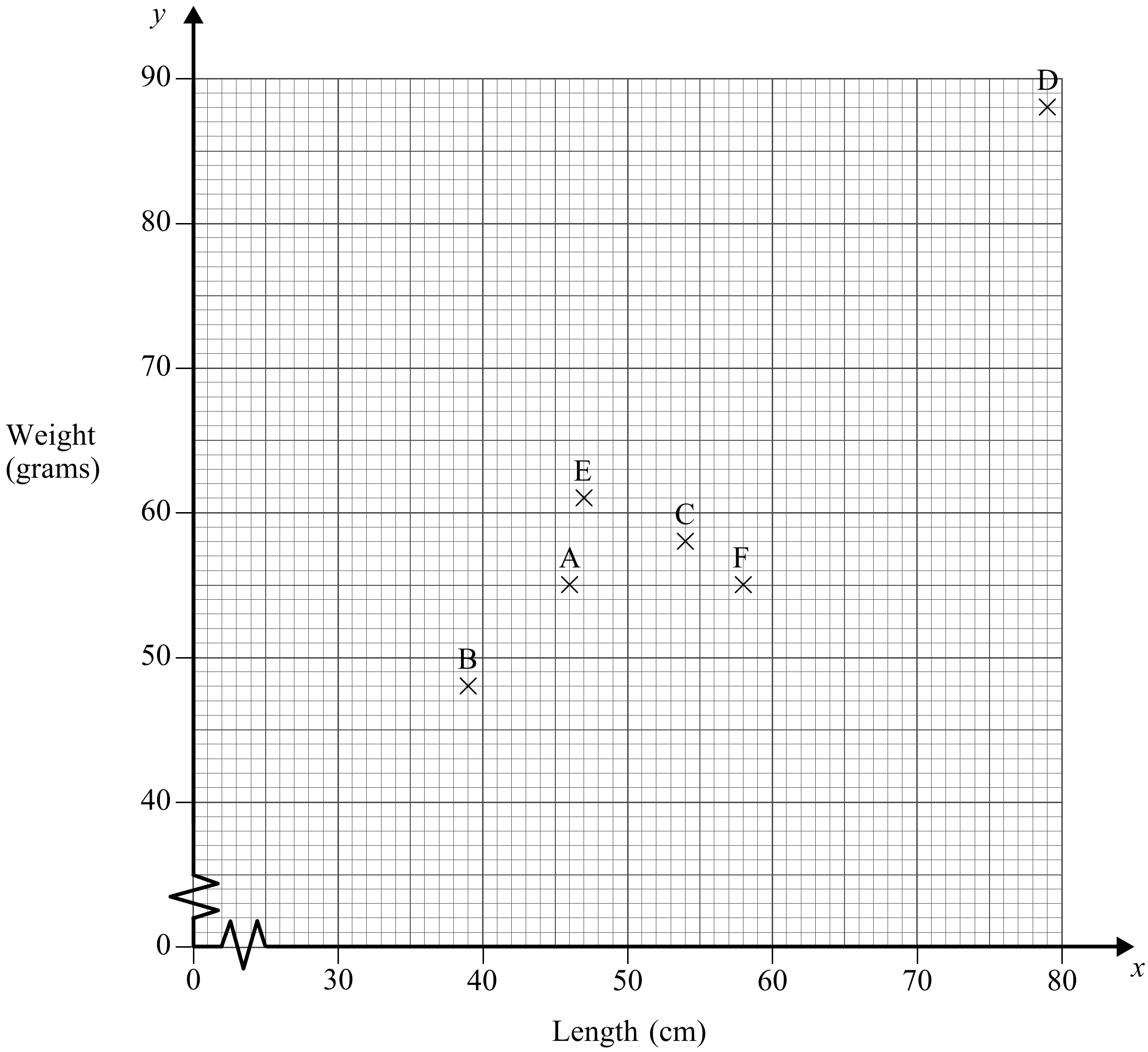
(ii) Hence, **without further calculation**, estimate the value of r for the 9 male snakes and revise, as necessary, your interpretation in part **(b)**. *(2 marks)*

QUESTION PART REFERENCE

Do not write outside the box

QUESTION PART REFERENCE

(c)



3 The weight, X grams, of talcum powder in a tin may be modelled by a normal distribution with mean 253 and standard deviation σ.

(a) Given that $\sigma = 5$, determine:

(i) $\mathrm{P}(X < 250)$; *(3 marks)*

(ii) $\mathrm{P}(245 < X < 250)$; *(2 marks)*

(iii) $\mathrm{P}(X = 245)$. *(1 mark)*

(b) Assuming that the value of the mean remains unchanged, determine the value of σ necessary to ensure that 98% of tins contain more than 245 grams of talcum powder. *(4 marks)*

QUESTION PART REFERENCE

4 As part of an investigation, a chlorine block is immersed in a large tank of water held at a constant temperature. The block slowly dissolves, and its weight, y grams, is noted x days after immersion. The results are shown in the table.

***x* days**	5	10	15	20	30	40	50	60	75
***y* grams**	47	44	42	38	35	27	23	16	9

(a) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(b) Hence estimate, to the nearest gram, the initial weight of the block. *(1 mark)*

(c) A company which markets the chlorine blocks claims that a block will usually dissolve completely after about 13 weeks. Comment, with justification, on this claim. *(3 marks)*

QUESTION PART REFERENCE

5 A survey of all the households on an estate is undertaken to provide information on the number of children per household.

The results, for the 99 households with children, are shown in the table.

Number of children (x)	1	2	3	4	5	6	7
Number of households (f)	14	35	25	13	9	2	1

(a) For these 99 households, calculate values for:

(i) the median and the interquartile range; *(3 marks)*

(ii) the mean and the standard deviation. *(3 marks)*

(b) In fact, 163 households were surveyed, of which 64 contained no children.

(i) For all 163 households, calculate the value for the mean number of children per household. *(2 marks)*

(ii) State whether the value for the standard deviation, when calculated for all 163 households, will be smaller than, the same as, or greater than that calculated in part **(a)(ii)**. *(1 mark)*

(iii) It is claimed that, for all 163 households on the estate, the number of children per household may be modelled approximately by a normal distribution.

Comment, with justification, on this claim. Your comment should refer to a fact other than the discrete nature of the data. *(2 marks)*

QUESTION PART REFERENCE

6 (a) The time taken, in minutes, by *Domesat* to install a domestic satellite system may be modelled by a normal distribution with unknown mean, μ, and standard deviation 15.

The times taken, in minutes, for a random sample of 10 installations are as follows.

47 39 25 51 47 36 63 41 78 43

Construct a 98% confidence interval for μ. *(5 marks)*

(b) The time taken, Y minutes, by *Teleair* to erect a TV aerial and then connect it to a TV is known to have a mean of 108 and a standard deviation of 28.

Estimate the probability that the mean of a random sample of 40 observations of Y is more than 120. *(4 marks)*

(c) Indicate, with a reason, where, if at all, in this question you made use of the Central Limit Theorem. *(2 marks)*

QUESTION PART REFERENCE

7 Mr Alott and Miss Fewer work in a postal sorting office.

(a) The number of letters per batch, R, sorted incorrectly by Mr Alott when sorting batches of 50 letters may be modelled by the distribution $\mathrm{B}(50, 0.15)$.

Determine:

(i) $\mathrm{P}(R < 10)$;

(ii) $\mathrm{P}(5 \leqslant R \leqslant 10)$. *(4 marks)*

(b) It is assumed that the probability that Miss Fewer sorts a letter incorrectly is 0.06, and that her sorting of a letter incorrectly is independent from letter to letter.

(i) Calculate the probability that, when sorting a batch of **22** letters, Miss Fewer sorts exactly 2 letters incorrectly. *(3 marks)*

(ii) Calculate the probability that, when sorting a batch of **35** letters, Miss Fewer sorts at least 1 letter incorrectly. *(2 marks)*

(iii) Calculate the mean and the variance for the number of letters sorted **correctly** by Miss Fewer when she sorts a batch of **120** letters. *(2 marks)*

(iv) Miss Fewer sorts a random sample of 20 batches, each containing 120 letters. The number of letters sorted **correctly** per batch has a mean of 112.8 and a variance of 56.86.

Comment on the assumptions that the probability that Miss Fewer sorts a letter incorrectly is 0.06, and that her sorting of a letter incorrectly is independent from letter to letter. *(3 marks)*

QUESTION PART REFERENCE

Surname		Other Names	
Centre Number		Candidate Number	
Candidate Signature			

General Certificate of Education
June 2009
Advanced Subsidiary Examination

MATHEMATICS **MS/SS1B**
Unit Statistics 1B

STATISTICS
Unit Statistics 1B

Insert

Insert for use in **Question 2**.

Fill in the boxes at the top of this page.

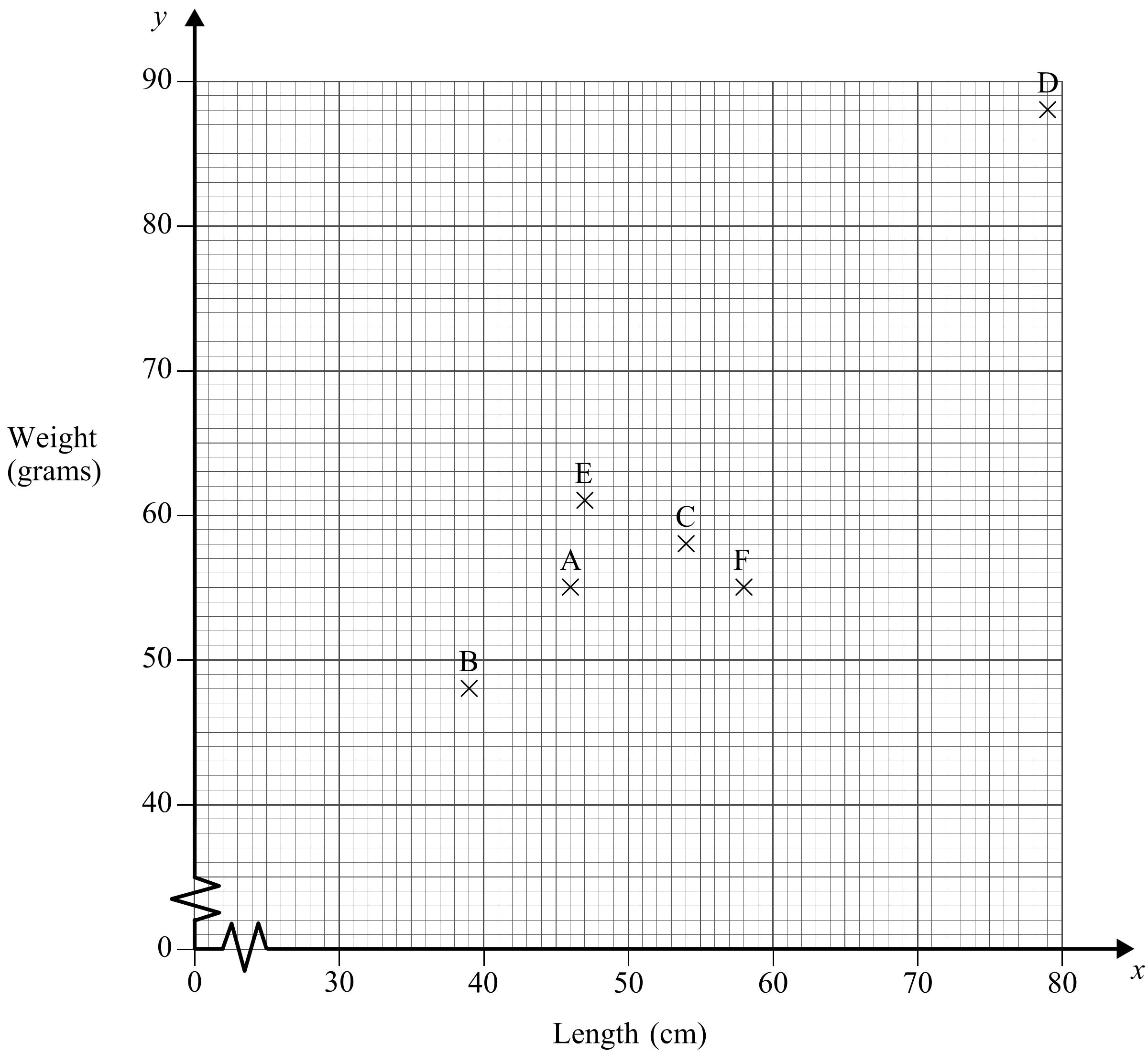
Fasten this insert securely to your answer book.

Turn over for Figure 1

Turn over ▶

Figure 1 (for use in Question 2)

Lengths and Weights of Snakes



General Certificate of Education
Advanced Subsidiary Examination
January 2010

Mathematics MS/SS1B

Unit Statistics 1B

Statistics

Unit Statistics 1B

Wednesday 13 January 2010 1.30 pm to 3.00 pm

For this paper you must have:
- an 8-page answer book
- the blue AQA booklet of formulae and statistical tables
- an insert for use in Question 7 (enclosed).

You may use a graphics calculator.

Time allowed
- 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Write the information required on the front of your answer book. The **Examining Body** for this paper is AQA. The **Paper Reference** is MS/SS1B.
- Answer **all** questions.
- Show all necessary working; otherwise marks for method may be lost.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.
- Fill in the boxes at the top of the insert.

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.

Answer **all** questions.

1 Draught excluder for doors and windows is sold in rolls of nominal length 10 metres.

The actual length, X metres, of draught excluder on a roll may be modelled by a normal distribution with mean 10.2 and standard deviation 0.15 .

(a) Determine:

(i) $P(X < 10.5)$; *(3 marks)*

(ii) $P(10.0 < X < 10.5)$. *(3 marks)*

(b) A customer randomly selects six 10-metre rolls of the draught excluder.

Calculate the probability that all six rolls selected contain more than 10 metres of draught excluder. *(3 marks)*

2 Lizzie, the receptionist at a dental practice, was asked to keep a weekly record of the number of patients who failed to turn up for an appointment. Her records for the first 15 weeks were as follows.

20 26 32 a 37 14 27 34 15 18 b 25 37 29 25

Unfortunately, Lizzie forgot to record the actual values for two of the 15 weeks, so she recorded them as a and b. However, she did remember that $a < 10$ and that $b > 40$.

(a) Calculate the median and the interquartile range of these 15 values. *(4 marks)*

(b) Give a reason why, for these data:

(i) the mode is **not** an appropriate measure of average;

(ii) the standard deviation **cannot** be used as a measure of spread. *(2 marks)*

(c) Subsequent investigations revealed that the missing values were 8 and 43.

Calculate the mean and the standard deviation of the 15 values. *(2 marks)*

3 The table shows, for each of a random sample of 7 weeks, the number of customers, x, who purchased fuel from a filling station, together with the total volume, y litres, of fuel purchased by these customers.

x	230	184	165	147	241	174	210
y	4551	3410	3252	3756	3787	4024	4254

(a) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(b) Estimate the volume of fuel sold during a week in which 200 customers purchase fuel. *(2 marks)*

(c) Comment on the likely reliability of your estimate in part (b), given that, for the regression line calculated in part (a), the values of the 7 residuals lie between approximately -415 litres and $+430$ litres. *(2 marks)*

4 Each school-day morning, three students, Rita, Said and Ting, travel independently from their homes to the same school by one of three methods: walk, cycle or bus. The table shows the probabilities of their independent daily choices.

	Walk	**Cycle**	**Bus**
Rita	0.65	0.10	0.25
Said	0.40	0.45	0.15
Ting	0.25	0.55	0.20

(a) Calculate the probability that, on any given school-day morning:

(i) all 3 students walk to school; *(2 marks)*

(ii) only Rita travels by bus to school; *(2 marks)*

(iii) at least 2 of the 3 students cycle to school. *(4 marks)*

(b) Ursula, a friend of Rita, never travels to school by bus. The probability that:

Ursula walks to school when Rita walks to school is 0.9;
Ursula cycles to school when Rita cycles to school is 0.7.

Calculate the probability that, on any given school-day morning, Rita and Ursula travel to school by:

(i) the same method; *(3 marks)*

(ii) different methods. *(1 mark)*

Turn over ▶

5 In a random sample of 12 bags of flour, the weight, in grams, of flour in each bag was recorded as follows.

1011 995 1018 1022 1014 1005 1017 1015 993 1018 992 1020

(a) It may be assumed that the weight of flour in a bag is normally distributed with a standard deviation of 10.5 grams.

(i) Construct a 98% confidence interval for the mean weight, μ grams, of flour in a bag, giving the limits to four significant figures. *(5 marks)*

(ii) State why, in constructing your confidence interval, use of the Central Limit Theorem was **not** necessary. *(1 mark)*

(iii) If the distribution of the weight of flour in a bag was unknown, indicate a minimum number of weights that you would consider necessary for a confidence interval for μ to be valid. *(1 mark)*

(b) The statement '1 kg' is printed on each bag.

Comment on this statement using **both** the confidence interval that you constructed in part (a)(i) and the weights of the given sample of 12 bags. *(3 marks)*

(c) Given that $\mu = 1000$, state the probability that a 98% confidence interval for μ will **not** contain 1000. *(1 mark)*

6 During the winter, the probability that Barry's cat, Sylvester, chooses to stay outside all night is 0.35, and the cat's choice is independent from night to night.

(a) Determine the probability that, during a period of 2 weeks (14 nights) in winter, Sylvester chooses to stay outside:

(i) on at most 7 nights; *(2 marks)*

(ii) on at least 11 nights; *(2 marks)*

(iii) on more than 5 nights but fewer than 10 nights. *(3 marks)*

(b) Calculate the probability that, during a period of **3 weeks** in winter, Sylvester chooses to stay outside on exactly 4 nights. *(3 marks)*

(c) Barry claims that, during the summer, the number of nights per week, S, on which Sylvester chooses to stay outside can be modelled by a binomial distribution with $n = 7$ and $p = \frac{5}{7}$.

(i) Assuming that Barry's claim is correct, find the mean and the variance of S. *(2 marks)*

(ii) For a period of 13 weeks during the summer, the number of nights per week on which Sylvester chose to stay outside had a mean of 5 and a variance of 1.5.

Comment on Barry's claim. *(2 marks)*

Turn over for the next question

7 [**Figure 1**, printed on the insert, is provided for use in this question.]

Harold considers himself to be an expert in assessing the auction value of antiques. He regularly visits car boot sales to buy items that he then sells at his local auction rooms.

Harold's father, Albert, who is not convinced of his son's expertise, collects the following data from a random sample of 12 items bought by Harold.

Item	**Purchase price (£x)**	**Auction price (£y)**
A	20	30
B	35	45
C	18	25
D	50	50
E	45	38
F	55	45
G	43	50
H	81	90
I	90	85
J	30	190
K	57	65
L	112	25

(a) Calculate the value of the product moment correlation coefficient between x and y. *(3 marks)*

(b) Interpret your value in the context of this question. *(2 marks)*

(c) (i) On **Figure 1**, complete the scatter diagram for these data. *(3 marks)*

(ii) Comment on what this reveals. *(2 marks)*

(d) When items J and L are omitted from the data, it is found that

$$S_{xx} = 4854.4 \qquad S_{yy} = 4216.1 \qquad S_{xy} = 4268.8$$

(i) Calculate the value of the product moment correlation coefficient between x and y for the remaining 10 items. *(2 marks)*

(ii) Hence revise as necessary your interpretation in part (b). *(1 mark)*

END OF QUESTIONS

Centre Number						Candidate Number				
Surname										
Other Names										
Candidate Signature										

For Examiner's Use	
Examiner's Initials	
Question	Mark
1	
2	
3	
4	
5	
6	
7	
TOTAL	

General Certificate of Education
Advanced Subsidiary Examination
June 2010

Mathematics MS/SS1B

Unit Statistics 1B

Statistics

Unit Statistics 1B

Thursday 27 May 2010 9.00 am to 10.30 am

For this paper you must have:
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed
- 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- Write the question part reference (eg (a), (b)(i) etc) in the left-hand margin.
- You must answer the questions in the spaces provided. Do not write outside the box around each page.
- Show all necessary working; otherwise marks for method may be lost.
- Do all rough work in this book. Cross through any work that you do not want to be marked.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.

Answer **all** questions in the spaces provided.

1 The weight, x kg, and the engine power, y bhp, of each car in a random sample of 10 hatchback cars are shown in the table.

x	1196	1062	1335	1429	1012	1355	1145	1417	1275	1284
y	123	88	150	158	69	120	94	143	107	128

(a) Calculate the value of the product moment correlation coefficient between x and y. *(3 marks)*

(b) Interpret your value in the context of the question. *(2 marks)*

QUESTION PART REFERENCE

2 Before leaving for a tour of the UK during the summer of 2008, Eduardo was told that the UK price of a 1.5-litre bottle of spring water was about 50p.

Whilst on his tour, Eduardo noted the prices, x pence, which he paid for 1.5-litre bottles of spring water from 12 retail outlets.

He then subtracted 50p from each price and his resulting differences, in pence, were

$$-18 \quad -11 \quad 1 \quad 15 \quad 7 \quad -1 \quad 17 \quad -16 \quad 18 \quad -3 \quad 0 \quad 9$$

(a) (i) Calculate the mean and the standard deviation of these differences. *(2 marks)*

(ii) Hence calculate the mean and the standard deviation of the prices, x pence, paid by Eduardo. *(2 marks)*

(b) Based on an exchange rate of €1.22 to £1, calculate, in euros, the mean and the standard deviation of the prices paid by Eduardo. *(3 marks)*

QUESTION PART REFERENCE

3 Each day, Margot completes the crossword in her local morning newspaper. Her completion times, X minutes, can be modelled by a normal random variable with a mean of 65 and a standard deviation of 20.

(a) Determine:

(i) $\mathrm{P}(X < 90)$;

(ii) $\mathrm{P}(X > 60)$. *(5 marks)*

(b) Given that Margot's completion times are independent from day to day, determine the probability that, during a particular period of 6 days:

(i) she completes one of the six crosswords in exactly 60 minutes; *(1 mark)*

(ii) she completes each crossword in less than 60 minutes; *(3 marks)*

(iii) her mean completion time is less than 60 minutes. *(4 marks)*

QUESTION PART REFERENCE

4 In a certain country, 15 per cent of the male population is left-handed.

(a) Determine the probability that, in a random sample of 50 males from this country:

(i) at most 10 are left-handed; *(2 marks)*

(ii) at least 5 are left-handed; *(2 marks)*

(iii) more than 6 but fewer than 12 are left-handed. *(3 marks)*

(b) In the same country, 11 per cent of the female population is left-handed.

Calculate the probability that, in a random sample of 35 females from this country, exactly 4 are left-handed. *(3 marks)*

(c) A sample of 2000 people is selected at random from the population of the country. The proportion of males in the sample is 52 per cent.

How many people in the sample would you expect to be left-handed? *(4 marks)*

QUESTION PART REFERENCE

5 Hugh owns a small farm.

(a) He has two sows, Josie and Rosie, which he feeds at a trough in their field at 8.00 am each day.

The probability that Josie is waiting at the trough at 8.00 am on any given day is 0.90. The probability that Rosie is waiting at the trough at 8.00 am on any given day is 0.70 when Josie is waiting at the trough, but is only 0.20 when Josie is not waiting at the trough.

Calculate the probability that, at 8.00 am on a given day:

(i) both sows are waiting at the trough; *(2 marks)*

(ii) neither sow is waiting at the trough; *(2 marks)*

(iii) at least one sow is waiting at the trough. *(1 mark)*

(b) Hugh also has two cows, Daisy and Maisy. Each day at 4.00 pm, he collects them from the gate to their field and takes them to be milked.

The probability, $\mathrm{P}(D)$, that Daisy is waiting at the gate at 4.00 pm on any given day is 0.75.
The probability, $\mathrm{P}(M)$, that Maisy is waiting at the gate at 4.00 pm on any given day is 0.60.
The probability that both Daisy and Maisy are waiting at the gate at 4.00 pm on any given day is 0.40.

(i) In the table of probabilities, D' and M' denote the events 'not D' and 'not M' respectively.

	M	M'	**Total**
D	0.40		0.75
D'			
Total	0.60		1.00

Complete the copy of this table which is printed on page 13. *(2 marks)*

(ii) Hence, or otherwise, find the probability that, at 4.00 pm on a given day:

(A) neither cow is waiting at the gate; *(1 mark)*

(B) only Daisy is waiting at the gate; *(1 mark)*

(C) exactly one cow is waiting at the gate. *(2 marks)*

Do not write outside the box

QUESTION PART REFERENCE

(b)(i)

	M	M'	Total
D	0.40		0.75
D'			
Total	0.60		1.00

Turn over ▶

6 During a study of reaction times, each of a random sample of 12 people, aged between 40 and 80 years, was asked to react as quickly as possible to a stimulus displayed on a computer screen.

Their ages, x years, and reaction times, y milliseconds, are shown in the table.

Person	Age (x years)	Reaction time (y ms)
A	41	520
B	54	750
C	66	650
D	72	920
E	71	280
F	57	620
G	60	740
H	47	950
I	77	970
J	65	780
K	51	550
L	59	730

(a) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(b) (i) Draw your regression line on the scatter diagram on page 16. *(2 marks)*

(ii) Comment on what this reveals. *(2 marks)*

(c) It was later discovered that the reaction times for persons E and H had been recorded incorrectly. The values should have been 820 and 590 respectively.

After making these corrections, computations gave

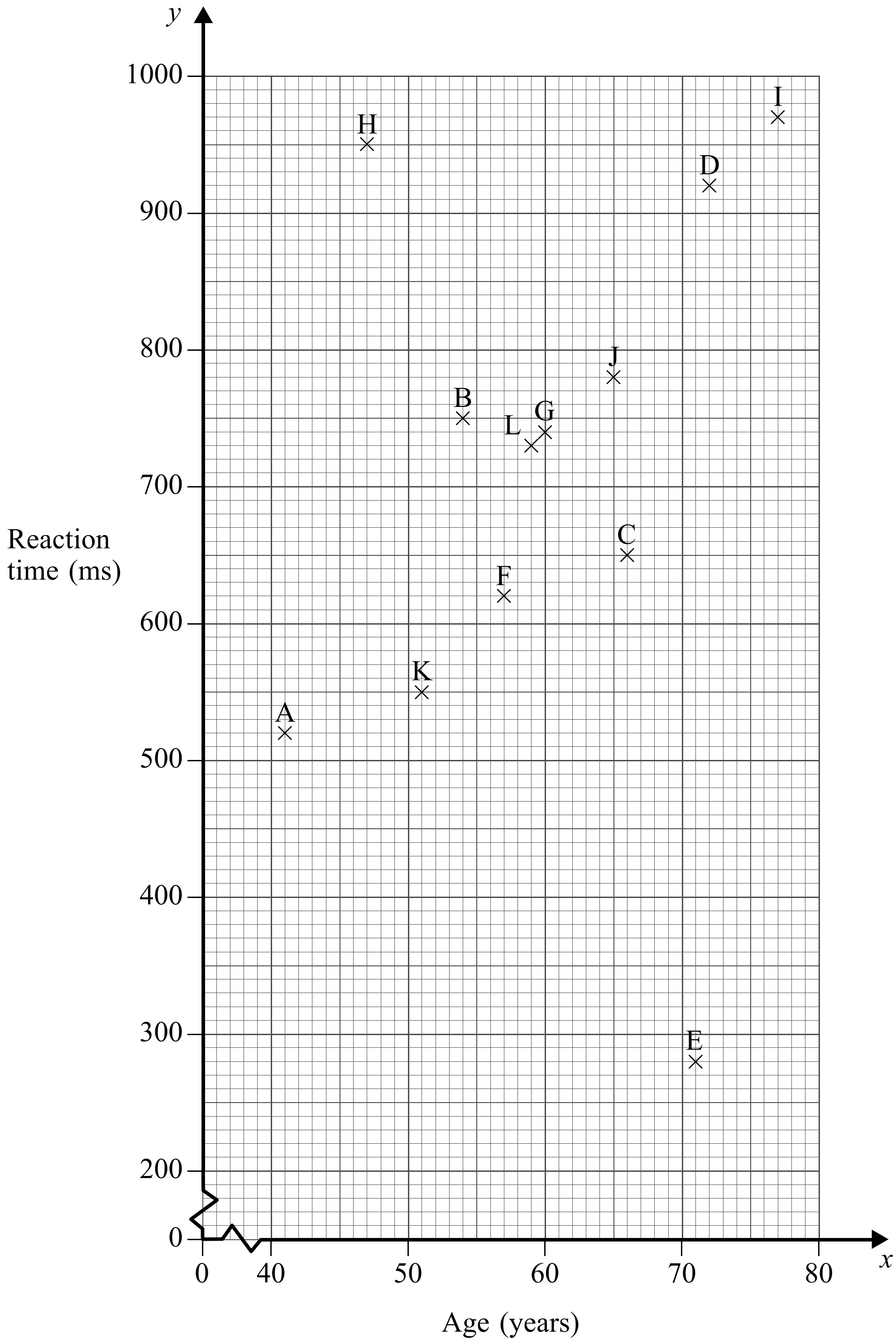
$$S_{xx} = 1272 \qquad S_{xy} = 14\,760 \qquad \bar{x} = 60 \qquad \bar{y} = 720$$

(i) Using the symbol $\odot$, plot the correct values for persons E and H on the scatter diagram on page 16. *(1 mark)*

(ii) Recalculate the equation of the least squares regression line of y on x, and draw this regression line on the scatter diagram on page 16. *(3 marks)*

(iii) Hence revise as necessary your comments in part **(b)(ii)**. *(2 marks)*

Reaction Times



7 An ambulance control centre responds to emergency calls in a rural area. The response time, T minutes, is defined as the time between the answering of an emergency call at the centre and the arrival of an ambulance at the given location of the emergency.

Response times have an unknown mean μ_T and an unknown variance.

Anita, the centre's manager, asked Peng, a student on supervised work experience, to record and summarise the values of T obtained from a random sample of 80 emergency calls.

Peng's summarised results were

Mean, $\bar{t} = 6.31$ Variance (unbiased estimate), $s^2 = 19.3$

Only 1 of the 80 values of T exceeded 20

(a) Anita then asked Peng to determine a confidence interval for μ_T. Peng replied that, from his summarised results, T was **not** normally distributed and so a valid confidence interval for μ_T could **not** be constructed.

(i) Explain, using the value of $\bar{t} - 2s$, why Peng's conclusion that T was not normally distributed was likely to be **correct**. *(2 marks)*

(ii) Explain why Peng's conclusion that a valid confidence interval for μ_T could not be constructed was **incorrect**. *(2 marks)*

(b) Construct a 98% confidence interval for μ_T. *(4 marks)*

(c) Anita had two targets for T. These were that $\mu_T < 8$ and that $\mathrm{P}(T \leqslant 20) > 95\%$.

Indicate, with justification, whether **each** of these two targets was likely to have been met. *(3 marks)*

QUESTION PART REFERENCE

Centre Number						Candidate Number				
Surname										
Other Names										
Candidate Signature										

For Examiner's Use	
Examiner's Initials	
Question	Mark
1	
2	
3	
4	
5	
6	
TOTAL	

General Certificate of Education
Advanced Subsidiary Examination
January 2011

Mathematics MS/SS1B

Unit Statistics 1B

Statistics

Unit Statistics 1B

Friday 14 January 2011 1.30 pm to 3.00 pm

For this paper you must have:
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed
- 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- Write the question part reference (eg (a), (b)(i) etc) in the left-hand margin.
- You must answer the questions in the spaces provided. Do not write outside the box around each page.
- Show all necessary working; otherwise marks for method may be lost.
- Do all rough work in this book. Cross through any work that you do not want to be marked.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

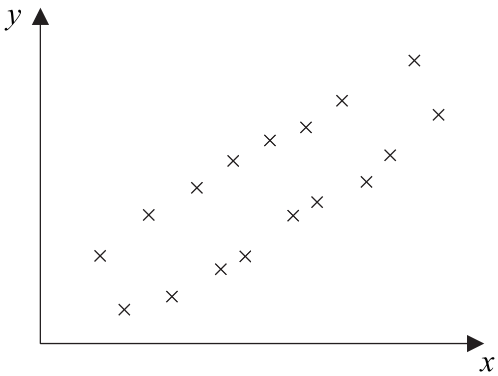
Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.

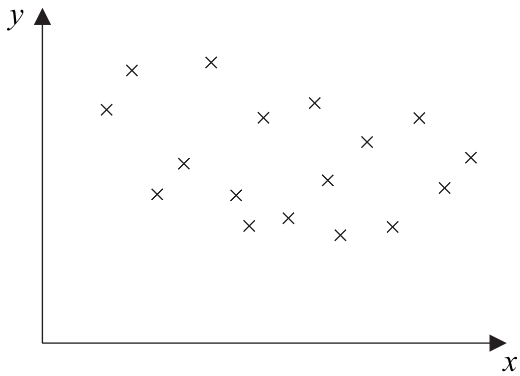
Answer **all** questions in the spaces provided.

1 (a) Estimate, **without undertaking any calculations**, the value of the product moment correlation coefficient between the variables x and y for each of the two scatter diagrams.

(i)



(ii)



(2 marks)

(b) The table gives the circumference, x centimetres, and the weight, y grams, of each of 12 new cricket balls.

x	22.5	22.7	22.6	22.4	22.5	22.8	22.6	22.7	22.8	22.4	22.9	22.6
y	160.3	159.4	157.8	158.0	157.3	159.8	158.3	159.6	161.3	156.4	162.5	161.2

(i) Calculate the value of the product moment correlation coefficient between x and y. *(3 marks)*

(ii) Assuming that the 12 balls may be considered to be a random sample, interpret your value in context. *(2 marks)*

QUESTION PART REFERENCE

2 The number of MPs in the House of Commons was 645 at the beginning of August 2009. The genders of these MPs and the political parties to which they belonged are shown in the table.

		Political Party				
		Labour	**Conservative**	**Liberal Democrat**	**Other**	**Total**
Gender	**Male**	255	175	54	35	519
	Female	94	18	9	5	126
	Total	349	193	63	40	645

(a) One MP was selected at random for an interview. Calculate, to three decimal places, the probability that the MP was:

(i) a male Conservative; *(1 mark)*

(ii) a male; *(1 mark)*

(iii) a Liberal Democrat; *(1 mark)*

(iv) Labour, given that the MP was female; *(2 marks)*

(v) male, given that the MP was **not** Labour. *(3 marks)*

(b) Two **female** MPs were selected at random for an enquiry. Calculate, to three decimal places, the probability that both MPs were Labour. *(2 marks)*

(c) Three MPs were selected at random for a committee. Calculate, to three decimal places, the probability that exactly one MP was Labour and exactly one MP was Conservative. *(4 marks)*

QUESTION PART REFERENCE

3 The volume, X litres, of orange juice in a 1-litre carton may be modelled by a normal distribution with unknown mean μ.

The volumes, x litres, recorded to the nearest 0.01 litre, in a random sample of 100 cartons are shown in the table.

Volume (x litres)	**Number of cartons (f)**
0.95 – 0.97	2
0.98 – 1.00	7
1.01 – 1.03	15
1.04 – 1.06	32
1.07 – 1.09	22
1.10 – 1.12	14
1.13 – 1.15	7
1.16 – 1.18	1
Total	**100**

(a) For the group '0.98 – 1.00':

(i) show that it has a mid-point of 0.99 litres;

(ii) state the minimum and the maximum values of x that could be included in this group. *(2 marks)*

(b) Calculate, to three decimal places, estimates of the mean and the standard deviation of these 100 volumes. *(3 marks)*

(c) (i) Construct an approximate 99% confidence interval for μ. *(4 marks)*

(ii) State why use of the Central Limit Theorem was **not** required when calculating this confidence interval. *(1 mark)*

(iii) Give a reason why the confidence interval is approximate rather than exact. *(1 mark)*

(d) Give a reason in support of the claim that:

(i) $\mu > 1$;

(ii) $\mathrm{P}(0.94 < X < 1.16)$ is approximately 1. *(2 marks)*

4 Clay pigeon shooting is the sport of shooting at special flying clay targets with a shotgun.

(a) Rhys, a novice, uses a single-barrelled shotgun. The probability that he hits a target is 0.45, and may be assumed to be independent from target to target.

Determine the probability that, in a series of shots at 15 targets, he hits:

(i) at most 5 targets; *(1 mark)*

(ii) more than 10 targets; *(2 marks)*

(iii) exactly 6 targets; *(2 marks)*

(iv) at least 5 but at most 10 targets. *(3 marks)*

(b) Sasha, an expert, uses a double-barrelled shotgun. She shoots at each target with the gun's first barrel and, only if she misses, does she then shoot at the target with the gun's second barrel.

The probability that she hits a target with a shot using her gun's first barrel is 0.85. The conditional probability that she hits a target with a shot using her gun's second barrel, given that she has missed the target with a shot using her gun's first barrel, is 0.80. Assume that Sasha's shooting is independent from target to target.

(i) Show that the probability that Sasha hits a target is 0.97. *(2 marks)*

(ii) Determine the probability that, in a series of shots at 50 targets, Sasha hits at least 48 targets. *(3 marks)*

(iii) In a series of shots at 80 targets, calculate the mean number of times that Sasha shoots at targets with her gun's second barrel. *(2 marks)*

QUESTION PART REFERENCE

Do not write outside the box

5 Craig uses his car to travel regularly from his home to the area hospital for treatment. He leaves home at x minutes after 7.30 am and then takes y minutes to arrive at the hospital's reception desk.

His results for 11 mornings are shown in the table.

x	0	5	10	15	20	25	30	35	40	45	50
y	31	42	32	58	47	56	79	68	89	95	85

(a) Explain why the time taken by Craig between leaving home and arriving at the hospital's reception desk is the response variable. *(1 mark)*

(b) Calculate the equation of the least squares regression line of y on x, writing your answer in the form $y = a + bx$. *(5 marks)*

(c) On a particular day, Craig needs to arrive at the hospital's reception desk no later than 9.00 am. He leaves home at 7.45 am.

Estimate the number of minutes **before** 9.00 am that Craig will arrive at the hospital's reception desk. Give your answer to the nearest minute. *(5 marks)*

(d) (i) Use your equation to estimate y when $x = 85$. *(1 mark)*

(ii) Give **one** statistical reason and **one** reason based on the context of this question as to why your estimate in part **(d)(i)** is unlikely to be realistic. *(2 marks)*

QUESTION PART REFERENCE

6 The volume of shampoo, V millilitres, delivered by a machine into bottles may be modelled by a normal random variable with mean μ and standard deviation σ.

(a) Given that $\mu = 412$ and $\sigma = 8$, determine:

(i) $\mathrm{P}(V < 400)$; *(3 marks)*

(ii) $\mathrm{P}(V > 420)$; *(2 marks)*

(iii) $\mathrm{P}(V = 410)$. *(1 mark)*

(b) A new quality control specification requires that the values of μ and σ are changed so that

$$\mathrm{P}(V < 400) = 0.05 \quad \text{and} \quad \mathrm{P}(V > 420) = 0.01$$

(i) Show, with the aid of a suitable sketch, or otherwise, that

$$400 - \mu = -1.6449\sigma \quad \text{and} \quad 420 - \mu = 2.3263\sigma$$ *(3 marks)*

(ii) Hence calculate values for μ and σ. *(3 marks)*

QUESTION PART REFERENCE

General Certificate of Education
Advanced Subsidiary Examination
June 2011

Mathematics MS/SS1B

Unit Statistics 1B

Statistics

Unit Statistics 1B

Friday 20 May 2011 1.30 pm to 3.00 pm

For this paper you must have:
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed
- 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- Write the question part reference (eg (a), (b)(i) etc) in the left-hand margin.
- You must answer the questions in the spaces provided. Do not write outside the box around each page.
- Show all necessary working; otherwise marks for method may be lost.
- Do all rough work in this book. Cross through any work that you do not want to be marked.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.

1 The number of matches in each of a sample of 85 boxes is summarised in the table.

Number of matches	**Number of boxes**
Less than 239	1
239–243	1
244–246	2
247	3
248	4
249	6
250	10
251	13
252	16
253	20
254	5
255–259	3
More than 259	1
Total	**85**

(a) For these data:

(i) state the modal value; *(1 mark)*

(ii) determine values for the median and the interquartile range. *(3 marks)*

(b) Given that, on investigation, the 2 extreme values in the above table are 227 and 271 :

(i) calculate the range; *(1 mark)*

(ii) calculate estimates of the mean and the standard deviation. *(4 marks)*

(c) For the numbers of matches in the 85 boxes, suggest, with a reason, the most appropriate measure of spread. *(2 marks)*

2 The diameter, D millimetres, of an American pool ball may be modelled by a normal random variable with mean 57.15 and standard deviation 0.04.

(a) Determine:

(i) $P(D < 57.2)$; *(3 marks)*

(ii) $P(57.1 < D < 57.2)$. *(2 marks)*

(b) A box contains 16 of these pool balls. Given that the balls may be regarded as a random sample, determine the probability that:

(i) all 16 balls have diameters less than 57.2 mm; *(2 marks)*

(ii) the mean diameter of the 16 balls is greater than 57.16 mm. *(4 marks)*

---

3 (a) During a particular summer holiday, Rick worked in a fish and chip shop at a seaside resort.

He suspected that the shop's takings, £y, on a weekday were dependent upon the forecast of that day's maximum temperature, x °C, in the resort, made at 6.00 pm on the previous day.

To investigate this suspicion, he recorded values of x and y for a random sample of 7 weekdays during July.

x	23	18	27	19	25	20	22
y	4290	3188	5106	3829	5057	4264	4485

(i) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(ii) Estimate the shop's takings on a weekday during July when the maximum temperature was forecast to be 24 °C. *(2 marks)*

(iii) Explain why your equation may not be suitable for estimating the shop's takings on a weekday during February. *(1 mark)*

(iv) Describe, in the context of this question, a variable other than the maximum temperature, x, that may affect y. *(1 mark)*

(b) Seren, who also worked in the fish and chip shop, investigated the possible linear relationship between the shop's takings, £z, recorded in £000s, and **each** of two other explanatory variables, v and w.

(i) She calculated correctly that the regression line of z on v had a z-intercept of -1 and a gradient of 0.15.

Draw this line, for values of v from 0 to 40, on **Figure 1** on page 4.

Turn over ▶

(ii) She also calculated correctly that the regression line of z on w had a z-intercept of 5 and a gradient of -0.40.

Draw this line, for values of w from 0 to 10, on **Figure 2** below. *(3 marks)*

Figure 1

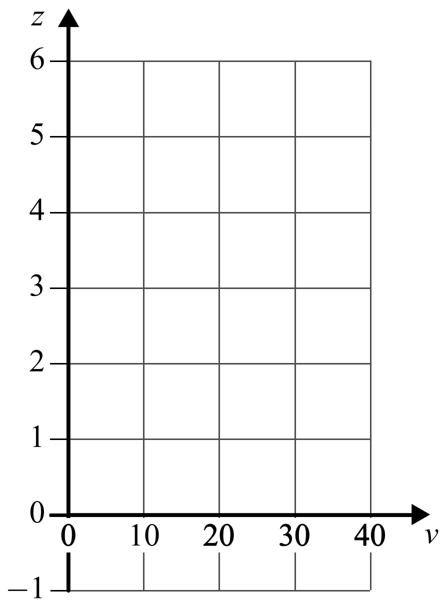
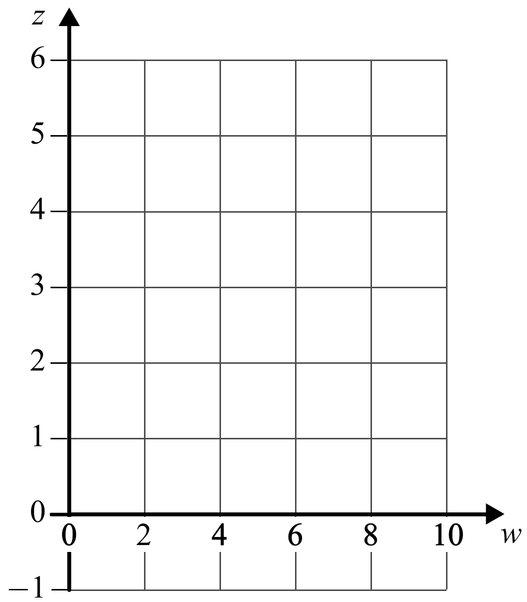


Figure 2



4 Rice that can be cooked in microwave ovens is sold in packets which the manufacturer claims contain a mean weight of more than 250 grams of rice.

The weight of rice in a packet may be modelled by a normal distribution.

A consumer organisation's researcher weighed the contents, x grams, of each of a random sample of 50 packets. Her summarised results are:

$$\bar{x} = 251.1 \quad \text{and} \quad \sum (x - \bar{x})^2 = 184.5$$

(a) Show that, correct to two decimal places, $s = 1.94$, where s^2 denotes the unbiased estimate of the population variance. *(1 mark)*

(b) (i) Construct a 96% confidence interval for the mean weight of rice in a packet, giving the limits to one decimal place. *(4 marks)*

(ii) Hence comment on the manufacturer's claim. *(2 marks)*

(c) The statement '250 grams' is printed on each packet.

Explain, with reference to the values of $\bar{x}$ and s, why the consumer organisation may consider this statement to be dubious. *(2 marks)*

---

5 (a) Emma visits her local supermarket every Thursday to do her weekly shopping.

The event that she buys orange juice is denoted by J, and the event that she buys bottled water is denoted by W. At each visit, Emma may buy neither, or one, or both of these items.

(i) Complete the table of probabilities, printed below, for these events, where J' and W' denote the events 'not J' and 'not W' respectively. *(3 marks)*

(ii) Hence, or otherwise, find the probability that, on any given Thursday, Emma buys either orange juice or bottled water but not both. *(2 marks)*

(iii) Show that:

(A) the events J and W are **not** mutually exclusive;

(B) the events J and W are **not** independent. *(3 marks)*

Turn over ▶

(b) Rhys visits the supermarket every Saturday to do his weekly shopping. Items that he may buy are milk, cheese and yogurt.

The probability, $\mathrm{P}(M)$, that he buys milk on any given Saturday is 0.85.
The probability, $\mathrm{P}(C)$, that he buys cheese on any given Saturday is 0.60.
The probability, $\mathrm{P}(Y)$, that he buys yogurt on any given Saturday is 0.55.
The events M, C and Y may be assumed to be independent.

Calculate the probability that, on any given Saturday, Rhys buys:

(i) none of the 3 items; *(2 marks)*

(ii) exactly 2 of the 3 items. *(3 marks)*

	J	J'	**Total**
W			0.65
W'	0.15		
Total		0.30	1.00

---

6 An amateur tennis club purchases tennis balls that have been used previously in professional tournaments.

The probability that each such ball fails a standard bounce test is 0.15.

The club purchases boxes each containing 10 of these tennis balls. Assume that the 10 balls in any box represent a random sample.

(a) Determine the probability that the number of balls in a box which fail the bounce test is:

(i) at most 2; *(1 mark)*

(ii) at least 2; *(2 marks)*

(iii) more than 1 but fewer than 5. *(3 marks)*

(b) Determine the probability that, in **5 boxes**, the total number of balls which fail the bounce test is:

(i) more than 5; *(2 marks)*

(ii) at least 5 but at most 10. *(3 marks)*

7 (a) Three airport management trainees, Ryan, Sunil and Tim, were each instructed to select a random sample of 12 suitcases from those waiting to be loaded onto aircraft.

Each trainee also had to measure the volume, x, and the weight, y, of each of the 12 suitcases in his sample, and then calculate the value of the product moment correlation coefficient, r, between x and y.

- Ryan obtained a value of -0.843.
- Sunil obtained a value of $+0.007$.

Explain why neither of these two values is likely to be correct. *(2 marks)*

(b) Peggy, a supervisor with many years' experience, measured the volume, x cubic feet, and the weight, y pounds, of each suitcase in a random sample of 6 suitcases, and then obtained a value of 0.612 for r.

- Ryan and Sunil each claimed that Peggy's value was different from their values because she had measured the volumes in cubic feet and the weights in pounds, whereas they had measured the volumes in cubic metres and the weights in kilograms.
- Tim claimed that Peggy's value was almost exactly half his calculated value because she had used a sample of size 6 whereas he had used one of size 12.

Explain why neither of these two claims is valid. *(2 marks)*

(c) Quentin, a manager, recorded the volumes, v, and the weights, w, of a random sample of 8 suitcases as follows.

v	28.1	19.7	46.4	23.6	31.1	17.5	35.8	13.8
w	14.9	12.1	21.1	18.0	19.8	19.2	16.2	14.7

(i) Calculate the value of r between v and w. *(3 marks)*

(ii) Interpret your value in the context of this question. *(2 marks)*

END OF QUESTIONS

General Certificate of Education
Advanced Subsidiary Examination
January 2012

Mathematics MS/SS1B

Unit Statistics 1B

Statistics

Unit Statistics 1B

Tuesday 17 January 2012 9.00 am to 10.30 am

For this paper you must have:
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed
- 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- Write the question part reference (eg (a), (b)(i) etc) in the left-hand margin.
- You must answer the questions in the spaces provided. Do not write outside the box around each page.
- Show all necessary working; otherwise marks for method may be lost.
- Do all rough work in this book. Cross through any work that you do not want to be marked.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.
- You do not necessarily need to use all the space provided.

1 Giles, a keen gardener, rents a council allotment.

During early April 2011, he planted 27 seed potatoes.

When he harvested his potato crop during the following August, he counted the number of new potatoes that he obtained from each seed potato.

He recorded his results as follows.

Number of new potatoes	$\leqslant 6$	7	8	9	10	11	$\geqslant 12$
Frequency	2	2	1	4	8	6	4

(a) Calculate values for the median and the interquartile range of these data. *(3 marks)*

(b) Advise Giles on how to record his corresponding data for 2012 so that it would then be possible to calculate the mean number of new potatoes per seed potato. *(1 mark)*

---

2 Dr Hanna has a special clinic for her older patients. She asked a medical student, Lenny, to select a random sample of 25 of her male patients, aged between 55 and 65 years, and, from their clinical records, to list their heights, weights and waist measurements.

Lenny was then asked to calculate three values of the product moment correlation coefficient based upon his collected data. His results were:

(a) 0.365 between height and waist measurement;

(b) 1.16 between height and weight;

(c) -0.583 between weight and waist measurement.

For **each** of Lenny's three calculated values, state whether the value is definitely correct, probably correct, probably incorrect or definitely incorrect. *(3 marks)*

3 During June 2011, the volume, X litres, of unleaded petrol purchased per visit at a supermarket's filling station by private-car customers could be modelled by a normal distribution with a mean of 32 and a standard deviation of 10.

(a) Determine:

(i) $\mathrm{P}(X < 40)$;

(ii) $\mathrm{P}(X > 25)$;

(iii) $\mathrm{P}(25 < X < 40)$. *(7 marks)*

(b) Given that during June 2011 unleaded petrol cost £1.34 per litre, calculate the probability that the unleaded petrol bill for a visit during June 2011 by a private-car customer exceeded £65. *(3 marks)*

(c) Give **two** reasons, in context, why the model $\mathrm{N}(32, 10^2)$ is unlikely to be valid for a visit by **any** customer purchasing fuel at this filling station during June 2011. *(2 marks)*

---

4 The records at a passport office show that, on average, 15 per cent of photographs that accompany applications for passport renewals are unusable.

Assume that exactly one photograph accompanies each application.

(a) Determine the probability that in a random sample of 40 applications:

(i) exactly 6 photographs are unusable;

(ii) at most 5 photographs are unusable;

(iii) more than 5 but fewer than 10 photographs are unusable. *(7 marks)*

(b) Calculate the mean and the standard deviation for the number of photographs that are unusable in a random sample of **32** applications. *(3 marks)*

(c) Mr Stickler processes 32 applications each day. His records for the previous 10 days show that the numbers of photographs that he deemed unusable were

8 6 10 7 9 7 8 9 6 7

By calculating the mean and the standard deviation of these values, comment, with reasons, on the suitability of the $\mathrm{B}(32, 0.15)$ model for the number of photographs deemed unusable each day by Mr Stickler. *(4 marks)*

5 An experiment was undertaken to collect information on the burning of a specific type of wood as a source of energy. At given fixed levels of the wood's moisture content, x per cent, its corresponding calorific value, y MWh/tonne, on burning was determined. The results are shown in the table.

x	5	10	15	20	25	30	35	40	45	50	55	60	65
y	5.2	4.7	4.3	4.0	3.2	2.8	2.5	2.2	1.8	1.5	1.3	1.0	0.6

(a) Explain why calorific value is the response variable. *(1 mark)*

(b) Calculate the equation of the least squares regression line of y on x, giving your answer in the form $y = a + bx$. *(5 marks)*

(c) Interpret, in context, your values for a and b. *(3 marks)*

(d) Use your equation to estimate the wood's calorific value when it has a moisture content of 27 per cent. *(2 marks)*

(e) Calculate the value of the residual for the point $(35, 2.5)$. *(2 marks)*

(f) Given that the values of the 13 residuals lie between -0.28 and $+0.23$, comment on the likely accuracy of your estimate in part **(d)**. *(1 mark)*

(g) (i) Give a **general reason** why your equation should not be used to estimate the wood's calorific value when it has a moisture content of 80 per cent. *(1 mark)*

(ii) Give a **specific reason**, based on the context of this question and with numerical support, why your equation cannot be used to estimate the wood's calorific value when it has a moisture content of 80 per cent. *(2 marks)*

---

6 Twins Alec and Eric are members of the same local cricket club and play for the club's under 18 team.

The probability that Alec is selected to play in any particular game is 0.85.
The probability that Eric is selected to play in any particular game is 0.60.
The probability that both Alec and Eric are selected to play in any particular game is 0.55.

(a) By using a table, or otherwise:

(i) show that the probability that neither twin is selected for a particular game is 0.10;

(ii) find the probability that at least one of the twins is selected for a particular game;

(iii) find the probability that exactly one of the twins is selected for a particular game. *(5 marks)*

(b) The probability that the twins' younger brother, Cedric, is selected for a particular game is:

0.30 given that both of the twins have been selected;
0.75 given that exactly one of the twins has been selected;
0.40 given that neither of the twins has been selected.

Calculate the probability that, for a particular game:

(i) all three brothers are selected;

(ii) at least two of the three brothers are selected. *(6 marks)*

---

7 A random sample of 50 full-time university employees was selected as part of a higher education salary survey.

The annual salary in thousands of pounds, x, of each employee was recorded, with the following summarised results.

$$\sum x = 2290.0 \quad \text{and} \quad \sum (x - \bar{x})^2 = 28\,225.50$$

Also recorded was the fact that 6 of the 50 salaries exceeded £60 000.

(a) (i) Calculate values for $\bar{x}$ and s, where s^2 denotes the unbiased estimate of σ^2. *(3 marks)*

(ii) Hence show why the annual salary, X, of a full-time university employee is unlikely to be normally distributed. Give numerical support for your answer. *(2 marks)*

(b) (i) Indicate why the mean annual salary, $\overline{X}$, of a random sample of 50 full-time university employees may be assumed to be normally distributed. *(2 marks)*

(ii) Hence construct a 99% confidence interval for the mean annual salary of full-time university employees. *(4 marks)*

(c) It is claimed that the annual salaries of full-time university employees have an average which exceeds £55 000 and that more than 25% of such salaries exceed £60 000.

Comment on **each** of these two claims. *(3 marks)*

General Certificate of Education
Advanced Subsidiary Examination
June 2012

Mathematics MS/SS1B

Unit Statistics 1B

Statistics

Unit Statistics 1B

Friday 18 May 2012 9.00 am to 10.30 am

For this paper you must have:
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed
- 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- Write the question part reference (eg (a), (b)(i) etc) in the left-hand margin.
- You must answer each question in the space provided for that question. If you require extra space, use an AQA supplementary answer book; do **not** use the space provided for a different question.
- Do not write outside the box around each page.
- Show all necessary working; otherwise marks for method may be lost.
- Do all rough work in this book. Cross through any work that you do not want to be marked.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.
- You do not necessarily need to use all the space provided.

1 A production line in a rolling mill produces lengths of steel.

A random sample of 20 lengths of steel from the production line was selected. The minimum width, x centimetres, and the minimum thickness, y millimetres, of each selected length was recorded.

The following summarised information was then calculated from these records.

$$S_{xx} = 2.030 \qquad S_{yy} = 1.498 \qquad S_{xy} = -0.410$$

(a) Calculate the value of the product moment correlation coefficient between x and y. *(2 marks)*

(b) Interpret your value in the context of the question. *(2 marks)*

---

2 Katy works as a clerical assistant for a small company. Each morning, she collects the company's post from a secure box in the nearby Royal Mail sorting office.

Katy's supervisor asks her to keep a daily record of the number of letters that she collects.

Her records for a period of 175 days are summarised in the table.

Daily number of letters (x)	**Number of days (f)**
0–9	5
10–19	16
20	23
21	27
22	31
23	34
24	16
25–29	10
30–34	5
35–39	3
40–49	4
50 or more	1
Total	175

(a) For these data:

(i) state the modal value; *(1 mark)*

(ii) determine values for the median and the interquartile range. *(3 marks)*

(b) The most letters that Katy collected on any of the 175 days was 54. Calculate estimates of the mean and the standard deviation of the daily number of letters collected by Katy. *(4 marks)*

(c) During the same period, a total of 280 letters was also delivered to the company by private courier firms.

Calculate an estimate of the mean daily number of **all** letters received by the company during the 175 days. *(2 marks)*

---

3 The table shows the maximum weight, y_A grams, of *Salt A* that will dissolve in 100 grams of water at various temperatures, x °C.

x	10	15	20	25	30	35	40	45	50	60	70	80
y_A	20	35	48	57	77	92	101	111	121	137	159	182

(a) Calculate the equation of the least squares regression line of y_A on x. *(4 marks)*

(b) The data in the above table are plotted on the scatter diagram on page 4.

Draw your regression line on this scatter diagram. *(2 marks)*

(c) For water temperatures in the range 10 °C to 80 °C, the maximum weight, y_B grams, of *Salt B* that will dissolve in 100 grams of water is given by the equation

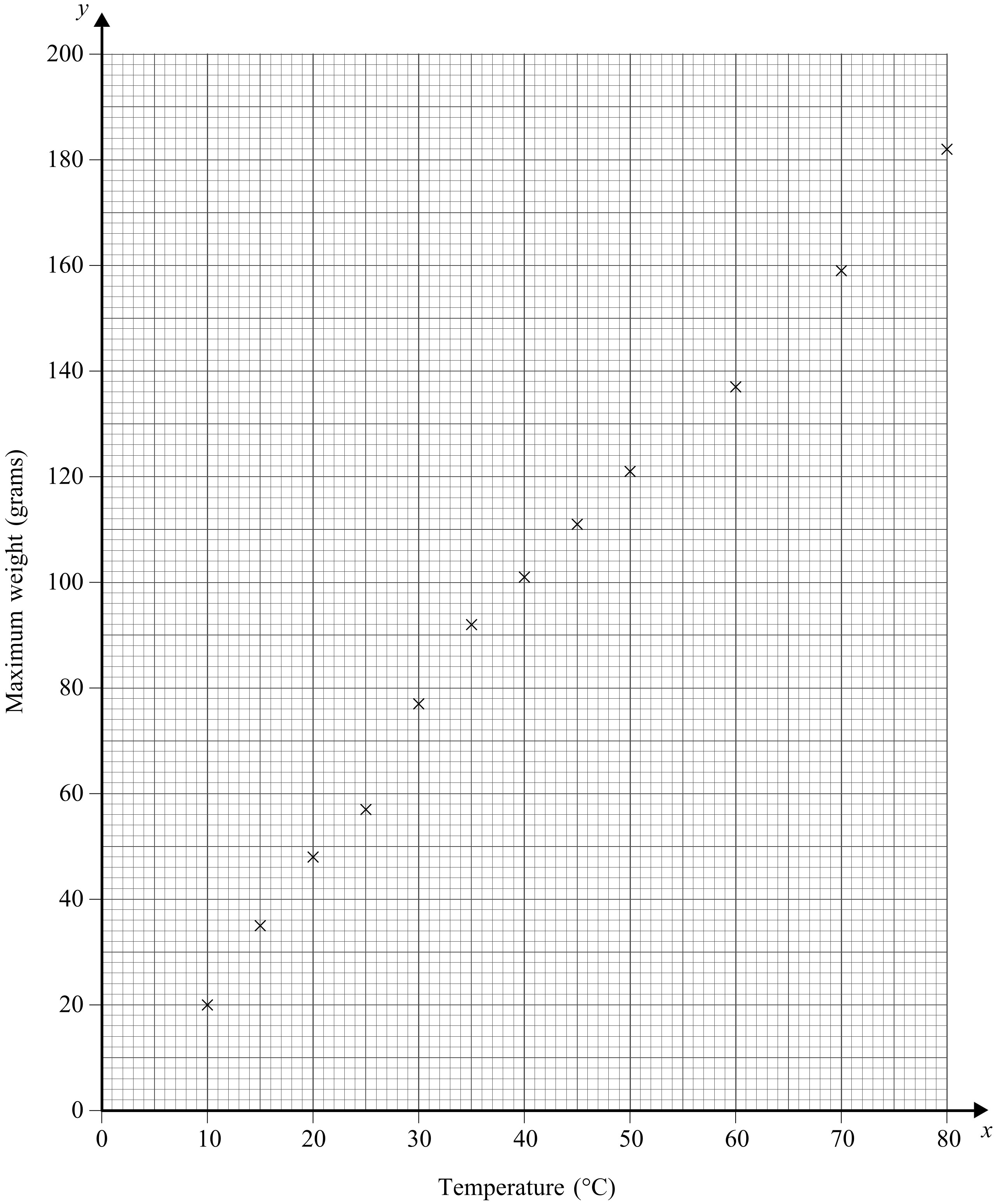
$$y_B = 60.1 + 0.255x$$

(i) Draw this line on the scatter diagram. *(2 marks)*

(ii) Estimate the water temperature at which the maximum weight of *Salt A* that will dissolve in 100 grams of water is the same as that of *Salt B*. *(1 mark)*

(iii) For *Salt A* and *Salt B*, compare the effects of water temperature on the maximum weight that will dissolve in 100 grams of water. Your answer should identify **two** distinct differences. *(2 marks)*

Temperatures and Maximum Weights



4 A survey of the 640 properties on an estate was undertaken. Part of the information collected related to the number of bedrooms and the number of toilets in each property.

This information is shown in the table.

		Number of toilets				
		1	**2**	**3**	**4 or more**	**Total**
Number of bedrooms	**1**	46	14	0	0	60
	2	24	67	23	0	114
	3	7	72	99	16	194
	4	0	19	123	48	190
	5 or more	0	0	11	71	82
	Total	77	172	256	135	640

(a) A property on the estate is selected at random.

Find, giving your answer to three decimal places, the probability that the property has:

(i) exactly 3 bedrooms; *(1 mark)*

(ii) at least 2 toilets; *(2 marks)*

(iii) exactly 3 bedrooms and at least 2 toilets; *(2 marks)*

(iv) at most 3 bedrooms, given that it has exactly 2 toilets. *(3 marks)*

(b) Use relevant answers from part **(a)** to show that the number of toilets is **not** independent of the number of bedrooms. *(2 marks)*

(c) Three properties are selected at random from those on the estate which have exactly 3 bedrooms.

Calculate the probability that one property has 2 toilets, one has 3 toilets and the other has at least 4 toilets. Give your answer to three decimal places. *(4 marks)*

5 A general store sells lawn fertiliser in 2.5 kg bags, 5 kg bags and 10 kg bags.

(a) The actual weight, W kilograms, of fertiliser in a 2.5 kg bag may be modelled by a normal random variable with mean 2.75 and standard deviation 0.15.

Determine the probability that the weight of fertiliser in a 2.5 kg bag is:

(i) less than 2.8 kg;

(ii) more than 2.5 kg. *(5 marks)*

(b) The actual weight, X kilograms, of fertiliser in a 5 kg bag may be modelled by a normal random variable with mean 5.25 and standard deviation 0.20.

(i) Show that $\mathrm{P}(5.1 < X < 5.3) = 0.372$, correct to three decimal places. *(2 marks)*

(ii) A random sample of **four** 5 kg bags is selected. Calculate the probability that none of the four bags contains between 5.1 kg and 5.3 kg of fertiliser. *(2 marks)*

(c) The actual weight, Y kilograms, of fertiliser in a 10 kg bag may be modelled by a normal random variable with mean 10.75 and standard deviation 0.50.

A random sample of **six** 10 kg bags is selected. Calculate the probability that the mean weight of fertiliser in the six bags is less than 10.5 kg. *(4 marks)*

---

6 A bin contains a very large number of paper clips of different colours. The proportion of each colour is shown in the table.

Colour	**White**	**Yellow**	**Green**	**Blue**	**Red**	**Purple**
Proportion	0.15	0.15	0.20	0.15	0.22	0.13

(a) Packets are filled from the bin. Each filled packet contains exactly 30 paper clips which may be considered to be a random sample.

Use binomial distributions to determine the probability that a filled packet contains:

(i) exactly 2 purple paper clips; *(3 marks)*

(ii) a **total** of more than 10 red or purple paper clips; *(3 marks)*

(iii) at least 5 but at most 10 green paper clips. *(3 marks)*

(b) Jumbo packets are also filled from the bin. Each filled jumbo packet contains exactly 100 paper clips.

(i) Assuming that the number of paper clips in a jumbo packet may be considered to be a random sample, calculate the mean and the variance of the number of **red** paper clips in a filled jumbo packet. *(2 marks)*

(ii) It is claimed that the proportion of red paper clips in the bin is greater than 0.22 and that jumbo packets do not contain random samples of paper clips.

An analysis of the number of red paper clips in each of a random sample of 50 filled jumbo packets resulted in a mean of 22.1 and a standard deviation of 4.17.

Comment, with numerical justification, on **each** of the two claims. *(3 marks)*

---

7 The volume of bleach in a 5-litre bottle may be modelled by a random variable with a standard deviation of 75 millilitres.

The volume, in litres, of bleach in each of a random sample of 36 such bottles was measured. The 36 measurements resulted in a **total** volume of 181.80 litres and exactly 8 bottles contained less than 5 litres.

(a) Construct a 98% confidence interval for the mean volume of bleach in a 5-litre bottle. *(5 marks)*

(b) It is claimed that the mean volume of bleach in a 5-litre bottle exceeds 5 litres and also that fewer than 10 per cent of such bottles contain less than 5 litres.

Comment, with numerical justification, on **each** of these two claims. *(3 marks)*

(c) State, with justification, whether you made use of the Central Limit Theorem in constructing the confidence interval in part **(a)**. *(1 mark)*

General Certificate of Education
Advanced Subsidiary Examination
January 2013

Mathematics

MS/SS1B

Unit Statistics 1B

Statistics

Unit Statistics 1B

Friday 18 January 2013 1.30 pm to 3.00 pm

For this paper you must have:
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed
- 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- Write the question part reference (eg (a), (b)(i) etc) in the left-hand margin.
- You must answer each question in the space provided for that question. If you require extra space, use an AQA supplementary answer book; do **not** use the space provided for a different question.
- Do not write outside the box around each page.
- Show all necessary working; otherwise marks for method may be lost.
- Do all rough work in this book. Cross through any work that you do not want to be marked.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.
- You do not necessarily need to use all the space provided.

1 Bob, a church warden, decides to investigate the lifetime of a particular manufacturer's brand of beeswax candle. Each candle is 30 cm in length.

From a box containing a large number of such candles, he selects one candle at random. He lights the candle and, after it has burned continuously for x hours, he records its length, y cm, to the nearest centimetre. His results are shown in the table.

x	5	10	15	20	25	30	35	40	45
y	27	25	21	19	16	11	9	5	2

(a) State the value that you would **expect** for a in the equation of the least squares regression line, $y = a + bx$. *(1 mark)*

(b) (i) Calculate the equation of the least squares regression line, $y = a + bx$. *(4 marks)*

(ii) Interpret the value that you obtain for b. *(2 marks)*

(iii) It is claimed by the candle manufacturer that the total length of time that such candles are likely to burn for is more than 50 hours.

Comment on this claim, giving a numerical justification for your answer. *(2 marks)*

---

2 The volume of *Everwhite* toothpaste in a pump-action dispenser may be modelled by a normal distribution with a mean of 106 ml and a standard deviation of 2.5 ml.

Determine the probability that the volume of *Everwhite* in a randomly selected dispenser is:

(a) less than 110 ml; *(3 marks)*

(b) more than 100 ml; *(2 marks)*

(c) between 104 ml and 108 ml; *(3 marks)*

(d) **not** exactly 106 ml. *(1 mark)*

3 *Stopoff* owns a chain of hotels. Guests are presented with the bills for their stays when they check out.

(a) Assume that the number of bills that contain errors may be modelled by a binomial distribution with parameters n and p, where $p = 0.30$.

Determine the probability that, in a random sample of 40 bills:

(i) at most 10 bills contain errors;

(ii) at least 15 bills contain errors;

(iii) exactly 12 bills contain errors. *(6 marks)*

(b) Calculate the mean and the variance for **each** of the distributions B(16, 0.20) and B(16, 0.125). *(3 marks)*

(c) Stan, who is a travelling salesperson, always uses *Stopoff* hotels. He holds one of its diamond customer cards and so should qualify for special customer care. However, he regularly finds errors in his bills when he checks out.

Each month, during a 12-month period, Stan stayed in *Stopoff* hotels on exactly 16 occasions. He recorded, each month, the number of occasions on which his bill contained errors. His recorded values were as follows.

2 1 4 3 1 3 0 3 1 0 5 1

(i) Calculate the mean and the variance of these 12 values. *(2 marks)*

(ii) Hence state with reasons which, if either, of the distributions B(16, 0.20) and B(16, 0.125) is likely to provide a satisfactory model for these 12 values. *(3 marks)*

4 Ashok is a work-experience student with an organisation that offers two separate professional examination papers, I and II.

For each of a random sample of 12 students, A to L, he records the mark, x per cent, achieved on Paper I, and the mark, y per cent, achieved on Paper II.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
x	34	46	53	62	67	72	60	54	70	71	82	85
y	61	66	72	78	88	81	49	60	54	44	49	36

(a) (i) Calculate the value of the product moment correlation coefficient, r, between x and y. *(3 marks)*

(ii) Interpret your value of r in the context of this question. *(2 marks)*

(b) (i) Give **two** possible advantages of plotting data on a graph before calculating the value of a product moment correlation coefficient. *(2 marks)*

(ii) Complete the plotting of Ashok's data on the scatter diagram on page 5. *(2 marks)*

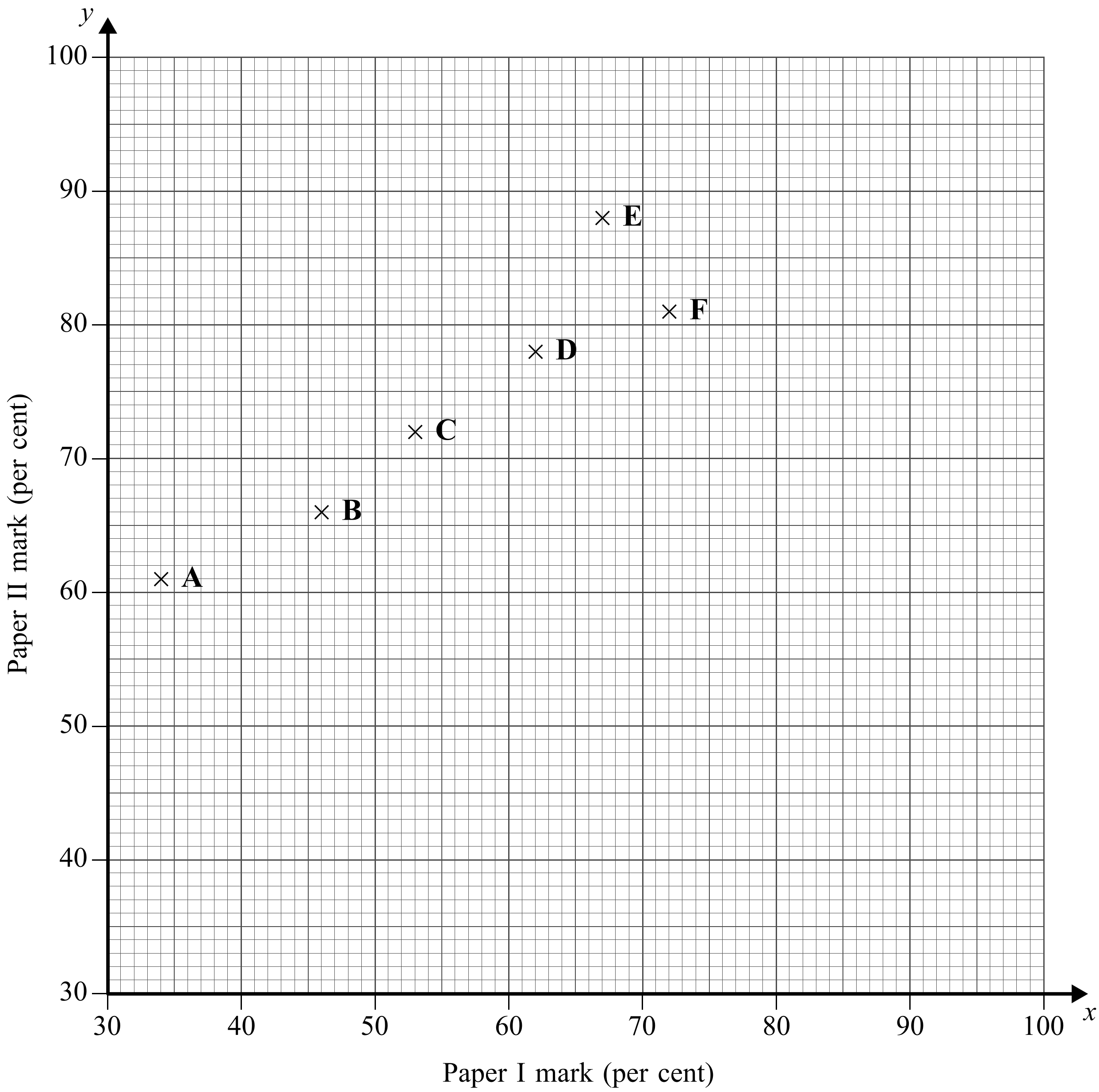
(iii) State what is now revealed by the scatter diagram. *(1 mark)*

(c) Ashok subsequently discovers that students A to F have a more scientific background than students G to L.

With reference to your scatter diagram, estimate the value of the product moment correlation coefficient for **each** of the two groups of students. You are **not** expected to calculate the two values. *(2 marks)*

	G	H	I	J	K	L
x	60	54	70	71	82	85
y	49	60	54	44	49	36

Examination Marks



5 Roger is an active retired lecturer. Each day after breakfast, he decides whether the weather for that day is going to be fine (F), dull (D) or wet (W). He then decides on only one of four activities for the day: cycling (C), gardening (G), shopping (S) or relaxing (R). His decisions from day to day may be assumed to be independent.

The table shows Roger's probabilities for each combination of weather and activity.

		Weather		
		Fine (F)	**Dull (D)**	**Wet (W)**
Activity	**Cycling (C)**	0.30	0.10	0
	Gardening (G)	0.25	0.05	0
	Shopping (S)	0	0.10	0.05
	Relaxing (R)	0	0.05	0.10

(a) Find the probability that, on a particular day, Roger decided:

(i) that it was going to be fine and that he would go cycling;

(ii) on either gardening or shopping;

(iii) to go cycling, given that he had decided that it was going to be fine;

(iv) **not** to relax, given that he had decided that it was going to be dull;

(v) that it was going to be fine, given that he did **not** go cycling. *(9 marks)*

(b) Calculate the probability that, on a particular Saturday and Sunday, Roger decided that it was going to be fine and decided on the same activity for both days. *(3 marks)*

6 (a) The length of one-metre galvanised-steel straps used in house building may be modelled by a normal distribution with a mean of 1005 mm and a standard deviation of 15 mm.

The straps are supplied to house builders in packs of 12, and the straps in a pack may be assumed to be a random sample.

Determine the probability that the **mean** length of straps in a pack is less than one metre. *(4 marks)*

(b) Tania, a purchasing officer for a nationwide house builder, measures the **thickness**, x millimetres, of each of a random sample of 24 galvanised-steel straps supplied by a manufacturer. She then calculates correctly that the value of $\bar{x}$ is 4.65 mm.

(i) Assuming that the thickness, X mm, of such a strap may be modelled by the distribution $\mathrm{N}(\mu, 0.15^2)$, construct a 99% confidence interval for μ. *(4 marks)*

(ii) Hence comment on the manufacturer's specification that the mean thickness of such straps is greater than 4.5 mm. *(2 marks)*

---

7 A machine, which cuts bread dough for loaves, can be adjusted to cut dough to any specified set weight. For any set weight, μ grams, the actual weights of cut dough are known to be approximately normally distributed with a mean of μ grams and a fixed standard deviation of σ grams.

It is also known that the machine cuts dough to within 10 grams of any set weight.

(a) Estimate, with justification, a value for σ. *(2 marks)*

(b) The machine is set to cut dough to a weight of 415 grams.

As a training exercise, Sunita, the quality control manager, asked Dev, a recently employed trainee, to record the weight of each of a random sample of 15 such pieces of dough selected from the machine's output. She then asked him to calculate the mean and the standard deviation of his 15 recorded weights.

Dev subsequently reported to Sunita that, for his sample, the mean was 391 grams and the standard deviation was 95.5 grams.

Advise Sunita on whether or not **each** of Dev's values is likely to be correct. Give numerical support for your answers. *(3 marks)*

(c) Maria, an experienced quality control officer, recorded the weight, y grams, of each of a random sample of 10 pieces of dough selected from the machine's output when it was set to cut dough to a weight of 820 grams. Her summarised results were as follows.

$$\sum y = 8210.0 \quad \text{and} \quad \sum (y - \bar{y})^2 = 110.00$$

Explain, with numerical justifications, why **both** of these values are likely to be correct. *(4 marks)*

Centre Number						Candidate Number				
Surname										
Other Names										
Candidate Signature										

For Examiner's Use	
Examiner's Initials	
Question	Mark
1	
2	
3	
4	
5	
6	
TOTAL	

General Certificate of Education
Advanced Subsidiary Examination
June 2013

Mathematics — MS/SS1B

Unit Statistics 1B

Statistics

Unit Statistics 1B

Friday 17 May 2013 9.00 am to 10.30 am

For this paper you must have:
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed
- 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- Write the question part reference (eg (a), (b)(i) etc) in the left-hand margin.
- You must answer each question in the space provided for that question. If you require extra space, use an AQA supplementary answer book; do **not** use the space provided for a different question.
- Do not write outside the box around each page.
- Show all necessary working; otherwise marks for method may be lost.
- Do all rough work in this book. Cross through any work that you do not want to be marked.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.
- You do not necessarily need to use all the space provided.

Answer **all** questions.

Answer each question in the space provided for that question.

1 The average maximum monthly temperatures, u degrees Fahrenheit, and the average minimum monthly temperatures, v degrees Fahrenheit, in New York City are as follows.

	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
Maximum (u)	39	40	48	61	71	81	85	83	77	67	54	41
Minimum (v)	26	27	34	44	53	63	68	66	60	51	41	30

(a) (i) Calculate, to one decimal place, the mean and the standard deviation of the 12 values of the average **maximum** monthly temperature. *(2 marks)*

(ii) For comparative purposes with a UK city, it was necessary to convert the temperatures from degrees Fahrenheit (°F) to degrees Celsius (°C). The formula used to convert f °F to c °C is:

$$c = \frac{5}{9}(f - 32)$$

Use this formula and your answers in part **(a)(i)** to calculate, **in °C**, the mean and the standard deviation of the 12 values of the average maximum monthly temperature. *(3 marks)*

(b) The value of the product moment correlation coefficient, r_{uv}, between the above 12 values of u and v is 0.997, correct to three decimal places.

State, giving a reason, the corresponding value of r_{xy}, where x and y are the exact equivalent temperatures in °C of u and v respectively. *(2 marks)*

QUESTION PART REFERENCE

Answer space for question 1

2 The weight, X grams, of the contents of a tin of baked beans can be modelled by a normal random variable with a mean of 421 and a standard deviation of 2.5 .

(a) Find:

(i) $\mathrm{P}(X = 421)$;

(ii) $\mathrm{P}(X < 425)$;

(iii) $\mathrm{P}(418 < X < 424)$. *(6 marks)*

(b) Determine the value of x such that $\mathrm{P}(X < x) = 0.98$. *(3 marks)*

(c) The weight, Y grams, of the contents of a tin of ravioli can be modelled by a normal random variable with a mean of μ and a standard deviation of 3.0 .

Find the value of μ such that $\mathrm{P}(Y < 410) = 0.01$. *(4 marks)*

QUESTION PART REFERENCE

Answer space for question 2

3 An auction house offers items of jewellery for sale at its public auctions. Each item has a reserve price which is less than the lower price estimate which, in turn, is less than the upper price estimate. The outcome for any item is independent of the outcomes for all other items.

The auction house has found, from past records, the following probabilities for the outcomes of items of jewellery offered for sale.

Outcome	Probability
Item does not achieve its reserve price	0.15
Item achieves at least its reserve price	0.85
Item achieves at least its lower price estimate	0.50
Item achieves at least its upper price estimate	0.175

For example, the probability that an item achieves at least its lower price estimate but not its upper price estimate is 0.325.

A particular auction includes exactly 40 items of jewellery that may be assumed to be a random sample of such items.

(a) Use binomial distributions to find the probability that:

(i) at most 10 items do not achieve their reserve prices; *(1 mark)*

(ii) 25 or more items achieve at least their lower price estimates; *(2 marks)*

(iii) exactly 2 items achieve at least their upper price estimates; *(2 marks)*

(iv) more than 10 items but fewer than 15 items achieve at least their reserve prices but not their lower price estimates. *(4 marks)*

(b) How many of the 40 items of jewellery would you expect to achieve at least their reserve prices but not their upper price estimates? *(2 marks)*

QUESTION PART REFERENCE	**Answer space for question 3**

4 The girth, g metres, the length, l metres, and the weight, y kilograms, of each of a sample of 20 pigs were measured.

The data collected is summarised as follows.

$$S_{gg} = 0.1196 \quad S_{ll} = 0.0436 \quad S_{yy} = 5880 \quad S_{gy} = 24.15 \quad S_{ly} = 10.25$$

(a) Calculate the value of the product moment correlation coefficient between:

(i) girth and weight;

(ii) length and weight. *(3 marks)*

(b) Interpret, in context, **each** of the values that you obtained in part **(a)**. *(3 marks)*

(c) Weighing pigs requires expensive equipment, whereas measuring their girths and lengths simply requires a tape measure. With this in mind, the following formula is proposed to make an estimate of a pig's weight, x kilograms, from its girth and length.

$$x = 69.3 \times g^2 \times l$$

Applying this formula to the relevant data on the 20 pigs resulted in

$$S_{xx} = 5656.15 \quad S_{xy} = 5662.97$$

(i) By calculating a third value of the product moment correlation coefficient, state which of g, l or x is the most strongly correlated with y, the weight. *(2 marks)*

(ii) Estimate the weight of a pig that has a girth of 1.25 metres and a length of 1.15 metres. *(2 marks)*

(iii) Given the additional information that $\bar{x} = 115.4$ and $\bar{y} = 116.0$, calculate the equation of the least squares regression line of y on x, in the form $y = a + bx$. *(3 marks)*

(iv) Comment on the likely accuracy of the estimated weight found in part **(c)(ii)**. Your answer should make reference to the value of the product moment correlation coefficient found in part **(c)(i)** and to the values of b and a found in part **(c)(iii)**. *(4 marks)*

QUESTION PART REFERENCE

Answer space for question 4

5 Alison is a member of a tenpin bowling club which meets at a bowling alley on Wednesday and Thursday evenings.

The probability that she bowls on a Wednesday evening is 0.90. Independently, the probability that she bowls on a Thursday evening is 0.95.

(a) Calculate the probability that, during a particular week, Alison bowls on:

(i) two evenings;

(ii) exactly one evening. *(3 marks)*

(b) David, a friend of Alison, is a member of the same club.

The probability that he bowls on a Wednesday evening, given that Alison bowls on that evening, is 0.80. The probability that he bowls on a Wednesday evening, given that Alison does not bowl on that evening, is 0.15.

The probability that he bowls on a Thursday evening, given that Alison bowls on that evening, is 1. The probability that he bowls on a Thursday evening, given that Alison does not bowl on that evening, is 0.

Calculate the probability that, during a particular week:

(i) Alison and David bowl on a Wednesday evening; *(2 marks)*

(ii) Alison and David bowl on both evenings; *(2 marks)*

(iii) Alison, but not David, bowls on a Thursday evening; *(1 mark)*

(iv) neither bowls on either evening. *(3 marks)*

QUESTION PART REFERENCE

Answer space for question 5

6 The weight, X kilograms, of sand in a bag can be modelled by a normal random variable with unknown mean μ and known standard deviation 0.4.

(a) The sand in each of a random sample of 25 bags from a large batch is weighed. The **total** weight of sand in these 25 bags is found to be 497.5 kg.

(i) Construct a 98% confidence interval for the mean weight of sand in bags in the batch. *(5 marks)*

(ii) Hence comment on the claim that bags in the batch contain an average of 20 kg of sand. *(2 marks)*

(iii) State why use of the Central Limit Theorem is **not** required in answering part **(a)(i)**. *(1 mark)*

(b) The weight, Y kilograms, of cement in a bag can be modelled by a normal random variable with mean 25.25 and standard deviation 0.35.

A firm purchases 10 such bags. These bags may be considered to be a random sample.

(i) Determine the probability that the **mean** weight of cement in the 10 bags is **less** than 25 kg. *(4 marks)*

(ii) Calculate the probability that the weight of cement in **each** of the 10 bags is **more** than 25 kg. *(4 marks)*

QUESTION PART REFERENCE

Answer space for question 6

Centre Number						Candidate Number				
Surname										
Other Names										
Candidate Signature										

For Examiner's Use	
Examiner's Initials	
Question	Mark
1	
2	
3	
4	
5	
6	
7	
TOTAL	

General Certificate of Education
Advanced Subsidiary Examination
June 2014

Statistics SS1B

Unit Statistics 1B

Thursday 22 May 2014 9.00 am to 10.30 am

For this paper you must have:
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed
- 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- Write the question part reference (eg (a), (b)(i) etc) in the left-hand margin.
- You must answer each question in the space provided for that question. If you require extra space, use an AQA supplementary answer book; do **not** use the space provided for a different question.
- Do not write outside the box around each page.
- Show all necessary working; otherwise marks for method may be lost.
- Do all rough work in this book. Cross through any work that you do not want to be marked.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.
- You do not necessarily need to use all the space provided.

Answer **all** questions.

Answer each question in the space provided for that question.

1 The weights, in kilograms, of a random sample of 15 items of cabin luggage on an aeroplane were as follows.

4.6 3.8 3.9 4.5 4.9 3.6 3.7 5.2 4.0 5.1 4.1 3.3 4.7 5.0 4.8

For these data:

(a) find values for the median and the interquartile range;

[4 marks]

(b) find the value for the range;

[1 mark]

(c) state why the mode is **not** an appropriate measure of average.

[1 mark]

QUESTION PART REFERENCE

Answer space for question 1

2 (a) Tim rings the church bell in his village every Sunday morning. The time that he spends ringing the bell may be modelled by a normal distribution with mean 7.5 minutes and standard deviation 1.6 minutes.

Determine the probability that, on a particular Sunday morning, the time that Tim spends ringing the bell is:

(i) at most 10 minutes;

(ii) more than 6 minutes;

(iii) between 5 minutes and 10 minutes.

[6 marks]

(b) June rings the same church bell for weekday weddings. The time that she spends, in minutes, ringing the bell may be modelled by the distribution $\mathrm{N}(\mu, 2.4^2)$.

Given that 80 per cent of the times that she spends ringing the bell are less than 15 minutes, find the value of μ.

[4 marks]

QUESTION PART REFERENCE

Answer space for question 2

3 The table shows the body mass index (BMI), x, and the systolic blood pressure (SBP), y mmHg, for each of a random sample of 10 men, aged between 35 years and 40 years, from a particular population.

x	13	23	29	35	17	34	25	20	31	27
y	103	115	124	126	108	120	113	117	118	119

(a) Calculate the equation of the least squares regression line of y on x.

[4 marks]

(b) Use your equation to estimate the SBP of a man from this population who is aged 38 years and who has a BMI of 30.

[2 marks]

(c) State why your equation might **not** be appropriate for estimating the SBP of a man from this population:

(i) who is aged 38 years and who has a BMI of 45;

(ii) who is aged 50 years and who has a BMI of 25.

[2 marks]

(d) Find the value of the residual for the point $(20, 117)$.

[2 marks]

(e) The mean of the vertical distances of the 10 points from the regression line calculated in part **(a)** is 2.71, correct to three significant figures.

Comment on the likely accuracy of your estimate in part **(b)**.

[1 mark]

QUESTION PART REFERENCE	**Answer space for question 3**

4 Alf and Mabel are members of a bowls club and sometimes attend the club's social events.

The probability, $\mathrm{P}(A)$, that Alf attends a social event is 0.70.
The probability, $\mathrm{P}(M)$, that Mabel attends a social event is 0.55.
The probability, $\mathrm{P}(A \cap M)$, that both Alf and Mabel attend the same social event is 0.45.

(a) Find the probability that:

(i) either Alf or Mabel or both attend a particular social event;

(ii) either Alf or Mabel but **not both** attend a particular social event.

[3 marks]

(b) Give a numerical justification for the following statement.

"Events A and M are **not** independent."

[2 marks]

(c) Ben and Nora are also members of the bowls club and sometimes attend the club's social events.

The probability, $\mathrm{P}(B)$, that Ben attends a social event is 0.85.
The probability, $\mathrm{P}(N)$, that Nora attends a social event is 0.65.
The attendance of each of Ben and Nora at a social event is independent of the attendance of all other members.

Find the probability that:

(i) all four named members attend a particular social event;

[2 marks]

(ii) none of the four named members attend a particular social event.

[3 marks]

QUESTION PART REFERENCE

Answer space for question 4

5 As part of a study of charity shops in a small market town, two such shops, X and Y, were each asked to provide details of its takings on 12 randomly selected days.

The table shows, for each of the 12 days, the day's takings, £x, of charity shop X and the day's takings, £y, of charity shop Y.

Day	A	B	C	D	E	F	G	H	I	J	K	L
x	46	57	39	116	62	77	41	61	15	53	68	61
y	78	102	66	214	98	72	98	134	21	67	95	83

(a) (i) Calculate the value of the product moment correlation coefficient between x and y.

[3 marks]

(ii) Interpret your value in the context of this question.

[2 marks]

(b) Complete the scatter diagram shown on the opposite page.

[2 marks]

(c) The investigator realised subsequently that one of the 12 selected days was a particularly popular town market day and another was a day on which the weather was extremely severe.

Identify **each** of these days giving a reason for each choice.

[3 marks]

(d) Removing the two days described in part **(c)** from the data gives the following information.

$$S_{xx} = 1292.5 \qquad S_{yy} = 3850.1 \qquad S_{xy} = 407.5$$

(i) Use this information to recalculate the value of the product moment correlation coefficient between x and y.

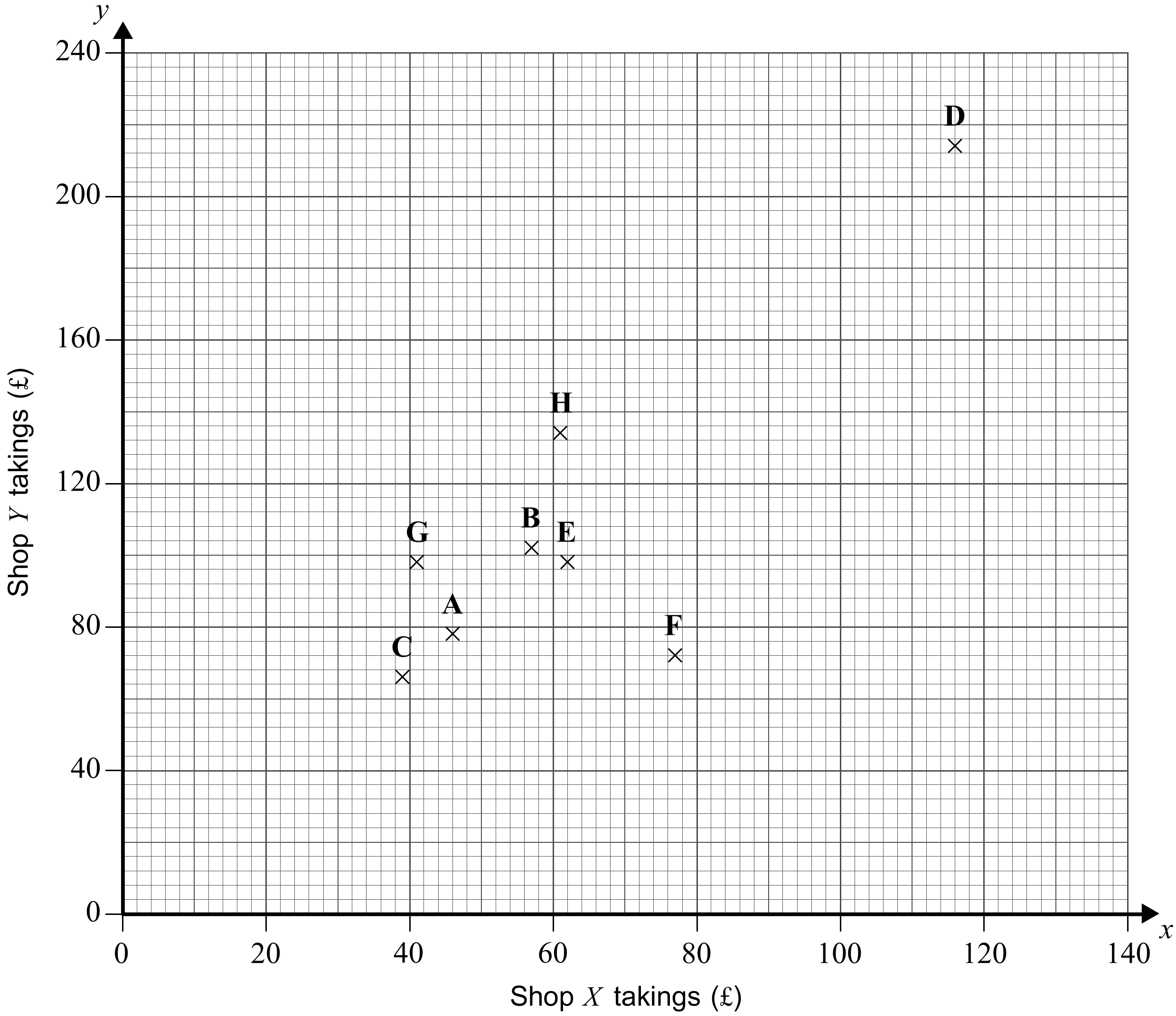
(ii) Hence revise, as necessary, your interpretation in part **(a)(ii)**.

[3 marks]

QUESTION PART REFERENCE

Answer space for question 5

Answer space for question 5



QUESTION PART REFERENCE

6 The probability that an online order from a supermarket chain has at least one item missing when delivered is 0.06.

Online orders are 'incomplete' if they contain substitute items and/or have at least one item missing when delivered. The probability that an order is incomplete is 0.15.

(a) Calculate the probability that exactly 2 out of a random sample of 26 online orders have at least one item missing when delivered.

[3 marks]

(b) Determine the probability that the number of incomplete orders in a random sample of 50 online orders is:

(i) fewer than 10;

(ii) more than 5;

(iii) more than 6 but fewer than 12.

[6 marks]

(c) Farokh, the manager of one of the supermarket's stores, examines 50 randomly selected online orders from each of a random sample of 100 of the store's customers. He records, for each of the 50 orders, the number, x, that were incomplete.

His summarised results, correct to three significant figures, for the 100 customers selected are

$$\bar{x} = 4.33 \quad \text{and} \quad s^2 = 3.94$$

Use this information to compare the performance of the store managed by Farokh with that of the supermarket chain as a whole.

[5 marks]

QUESTION PART REFERENCE

Answer space for question 6

7 For the year 2014, the table below summarises the weights, x kilograms, of a random sample of 160 women residing in a particular city who are aged between 18 years and 25 years.

Weight (x kg)	**Number of women**
35–40	4
40–45	9
45–50	12
50–55	16
55–60	24
60–65	28
65–70	24
70–75	17
75–80	12
80–85	7
85–90	4
90–95	2
95–100	1
Total	**160**

(a) Calculate estimates of the mean and the standard deviation of these 160 weights.
[4 marks]

(b) (i) Construct a 98% confidence interval for the mean weight of women residing in the city who are aged between 18 years and 25 years.
[5 marks]

(ii) Hence comment on a claim that the mean weight of women residing in the city who are aged between 18 years and 25 years has increased from that of $61.7\,\text{kg}$ in 1965.
[2 marks]

QUESTION PART REFERENCE	**Answer space for question 7**

Centre Number						Candidate Number				
Surname										
Other Names										
Candidate Signature										

For Examiner's Use	
Examiner's Initials	
Question	Mark
1	
2	
3	
4	
5	
6	
7	
TOTAL	

General Certificate of Education
Advanced Subsidiary Examination
June 2015

Statistics SS1B

Unit Statistics 1B

Wednesday 20 May 2015 9.00 am to 10.30 am

For this paper you must have:
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed
- 1 hour 30 minutes

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- Write the question part reference (eg (a), (b)(i) etc) in the left-hand margin.
- You must answer each question in the space provided for that question. If you require extra space, use an AQA supplementary answer book; do **not** use the space provided for a different question.
- Do not write outside the box around each page.
- Show all necessary working; otherwise marks for method may be lost.
- Do all rough work in this book. Cross through any work that you do not want to be marked.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.
- Unit Statistics 1B has a **written paper only**.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.
- You do not necessarily need to use all the space provided.

Answer **all** questions.

Answer each question in the space provided for that question.

1 The number of passengers getting off the 11.45 am train at a railway station on each of 35 days is summarised as follows.

Number of passengers	6	7	8	10	11	12	14	15	18
Number of days	1	1	2	9	7	4	5	3	3

For these data:

(a) find values for the mode, the median and the interquartile range;

[4 marks]

(b) calculate the value for the mean.

[2 marks]

QUESTION PART REFERENCE

Answer space for question 1

2 The length of aluminium baking foil on a roll may be modelled by a normal distribution with mean 91 metres and standard deviation 0.8 metres.

(a) Determine the probability that the length of foil on a particular roll is:

(i) less than 90 metres;

(ii) **not** exactly 90 metres;

(iii) between 91 metres and 92.5 metres.

[6 marks]

(b) The length of cling film on a roll may also be modelled by a normal distribution but with mean 153 metres and standard deviation σ metres.

It is required that 1% of rolls of cling film should have a length less than 150 metres.

Find the value of σ that is needed to satisfy this requirement.

[4 marks]

QUESTION PART REFERENCE

Answer space for question 2

3 Fourteen candidates each sat two test papers, Paper 1 and Paper 2, on the same day. The marks, out of a total of 50, achieved by the students on each paper are shown in the table.

Candidate	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**
Mark on Paper 1 (x)	36	29	33	17	42	26	45	12	25	19	26	15	28	11
Mark on Paper 2 (y)	46	18	34	24	45	21	37	15	35	17	38	11	44	21

(a) (i) Calculate the value of the product moment correlation coefficient, r, between the marks on Paper 1 and those on Paper 2.

[3 marks]

(ii) Interpret your value in the context of this question.

[2 marks]

(b) It was then discovered that seven of the candidates, Group T, had been given extra tuition in preparation for the tests, whereas the other seven candidates, Group U, had been given only the usual tuition.

The summarised data for the two groups are as follows.

Group T: $r = 0.261$ $\quad \bar{x} = 33.57$ $\quad \bar{y} = 39.86$

Group U: $S_{xx} = 279.71$ $\quad S_{yy} = 112.86$ $\quad S_{xy} = 34.57$ $\quad \bar{x} = 18.43$ $\quad \bar{y} = 18.14$

(i) For Group U, calculate the value of r.

[2 marks]

(ii) Interpret, in the context of the question, the value of r for **each** group of candidates.

[2 marks]

(iii) Comment, with justification, on the apparent effect of the extra tuition.

[2 marks]

QUESTION PART REFERENCE

Answer space for question 3

4 (a) Chris shops at his local store on his way to and from work every Friday.

The event that he buys a morning newspaper is denoted by M, and the event that he buys an evening newspaper is denoted by E.

On any one Friday, Chris may buy neither, exactly one or both of these newspapers.

(i) Complete the table of probabilities, printed on the opposite page, where M' and E' denote the events 'not M' and 'not E' respectively.

[3 marks]

(ii) Hence, or otherwise, find the probability that, on any given Friday, Chris buys exactly one newspaper.

[2 marks]

(iii) Give a numerical justification for the following statement.

'The events M and E are **not** mutually exclusive.'

[2 marks]

(b) The event that Chris buys a morning newspaper on Saturday is denoted by S, and the event that he buys a morning newspaper on the following day, Sunday, is denoted by T. The event that he buys a morning newspaper on both Saturday and Sunday is denoted by $S \cap T$.

Each combination of the events S and T is independent of any combination of the events M and E. However, the events S and T are **not** independent, with

$$\mathrm{P}(S) = 0.85, \quad \mathrm{P}(T \mid S) = 0.20 \quad \text{and} \quad \mathrm{P}(T \mid S') = 0.75$$

Find the probability that, on a particular Friday, Saturday and Sunday, Chris buys:

(i) all four newspapers;

[2 marks]

(ii) none of the four newspapers.

[2 marks]

(c) (i) State, as briefly as possible, in the context of the question, the event that is denoted by $M \cap E' \cap S \cap T'$.

[2 marks]

(ii) Calculate the value of $\mathrm{P}(M \cap E' \cap S \cap T')$.

[2 marks]

QUESTION PART REFERENCE

Answer space for question 4

(a)(i)

	M	M'	**Total**
E	0.16		0.28
E'			
Total		0.60	1.00

5 The table shows the number of customers, x, and the takings, £y, recorded to the nearest £10, at a local butcher's shop on each of 10 randomly selected weekdays.

x	86	60	65	46	71	93	56	81	75	57
y	940	790	620	530	770	1050	690	780	860	550

(a) The first 6 pairs of data values in this table are plotted on the scatter diagram shown on the opposite page.

Plot the final 4 pairs of data values on the scatter diagram.

[1 mark]

(b) (i) Calculate the equation of the least squares regression line in the form $y = a + bx$ and draw your line on the scatter diagram.

[6 marks]

(ii) Interpret your value for b in the context of the question.

[2 marks]

(iii) State why your value for a has no practical interpretation.

[1 mark]

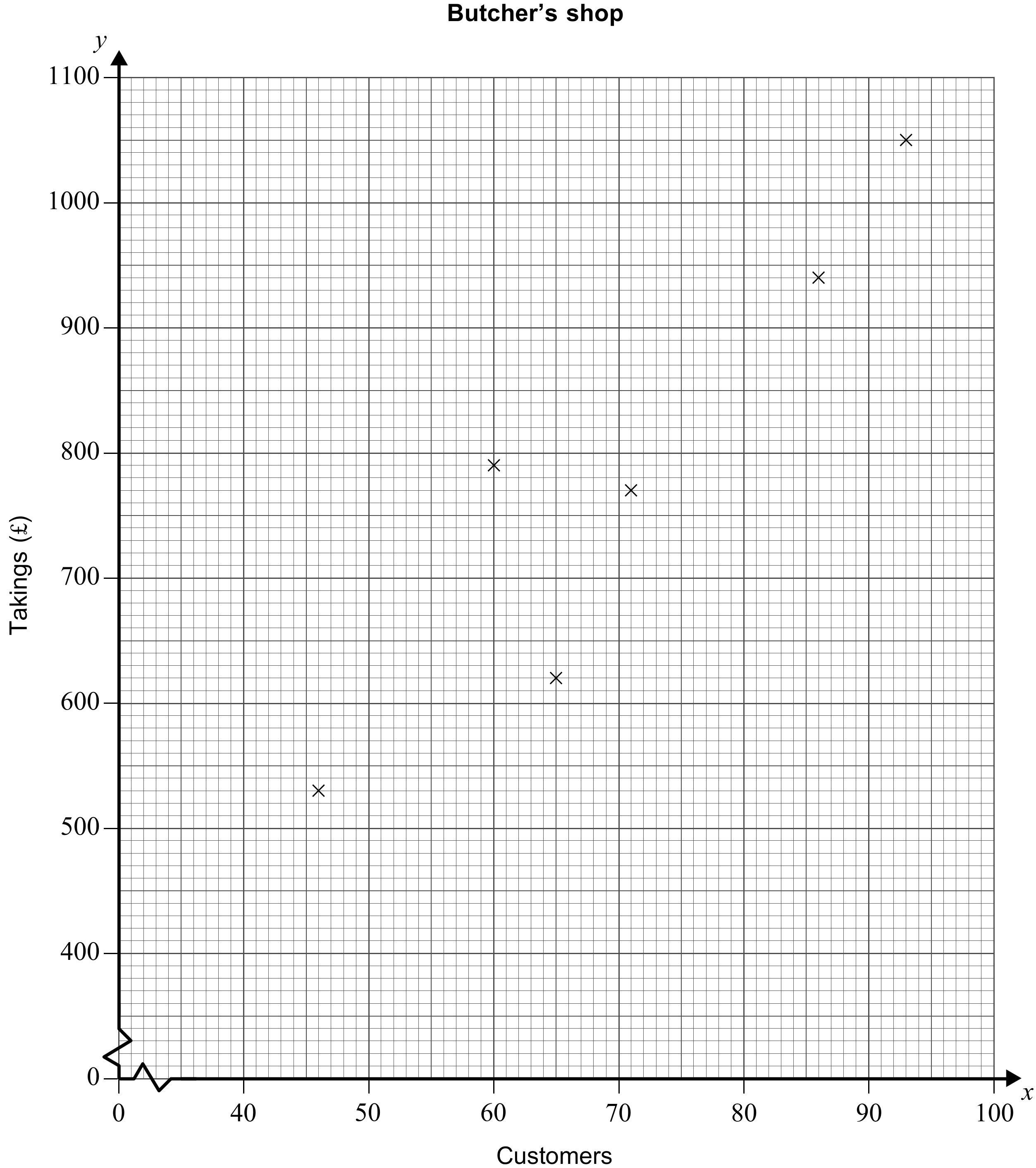
(c) Estimate, to the nearest £10, the shop's takings when the number of customers is 50.

[1 mark]

QUESTION PART REFERENCE

Answer space for question 5

Answer space for question 5



QUESTION PART REFERENCE

6 Customers at a supermarket can pay at a checkout either by cash, debit card or credit card.

(a) The probability that a customer pays by cash is 0.22.

Calculate the probability that exactly 2 customers from a random sample of 24 customers pay by cash.

[3 marks]

(b) The probability that a customer pays by debit card is 0.45.

Determine the probability that the number of customers who pay by debit card from a random sample of $\mathbf{40}$ customers is:

(i) fewer than 20;

(ii) more than 15;

(iii) at least 12 but at most 24.

[6 marks]

(c) The random variable W denotes the number of customers who pay by **credit card** from a random sample of $\mathbf{200}$ customers.

Calculate values for the mean and the variance of W.

[3 marks]

QUESTION PART REFERENCE

Answer space for question 6

7 (a) A greengrocer displays apples in trays. Each customer selects the apples he or she wishes to buy and puts them into a bag.

Records show that the weight of such bags of apples may be modelled by a normal distribution with mean 1.16 kg and standard deviation 0.43 kg.

Determine the probability that the **mean** weight of a random sample of 10 such bags of apples exceeds 1.25 kg.

[4 marks]

(b) The greengrocer also displays pears in trays. Each customer selects the pears he or she wishes to buy and puts them into a bag.

A random sample of 40 such bags of pears had a mean weight of 0.86 kg and a standard deviation of 0.65 kg.

(i) Construct a **96%** confidence interval for the mean weight of a bag of pears.

[4 marks]

(ii) Hence comment on a claim that customers wish to buy, on average, a greater weight of apples than of pears.

[2 marks]

QUESTION PART REFERENCE

Answer space for question 7