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 2013

Statistics SS02

Unit Statistics 2

Wednesday 23 January 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.

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 A school Geography department keeps records of the total rainfall at the school for each quarter. The table shows data for the spring, summer, autumn and winter quarters from spring 2009 to autumn 2011. All values are given to the nearest millimetre.

Year	**2009**				**2010**				**2011**		
Quarter	**Sp**	**Su**	**Au**	**Wi**	**Sp**	**Su**	**Au**	**Wi**	**Sp**	**Su**	**Au**
Rainfall	275	383	447	315	222	323	388	255	166	244	338
Moving average		355	342	327	312	297	283	263			

The graph opposite shows the quarterly rainfall and the moving averages.

(a) (i) Calculate the value of the missing moving average. *(2 marks)*

(ii) Add the value of this moving average to the graph and draw a trend line. *(2 marks)*

(b) Using the graph:

(i) calculate the seasonal effect for the winter quarter; *(3 marks)*

(ii) forecast the rainfall for the winter quarter of 2011. *(2 marks)*

(c) Give **two** reasons why it would be inappropriate to use the method of part **(b)** to forecast the rainfall for the winter quarter of 2020. *(2 marks)*

(d) Subsequent records of the Geography department show that the actual rainfall for the winter quarter of 2011 was 259 mm, and that the values of the next three moving averages were 252 mm, 254 mm and 258 mm. Comment on:

(i) the accuracy of your forecast in part **(b)(ii)**;

(ii) the trend. *(2 marks)*

QUESTION PART REFERENCE

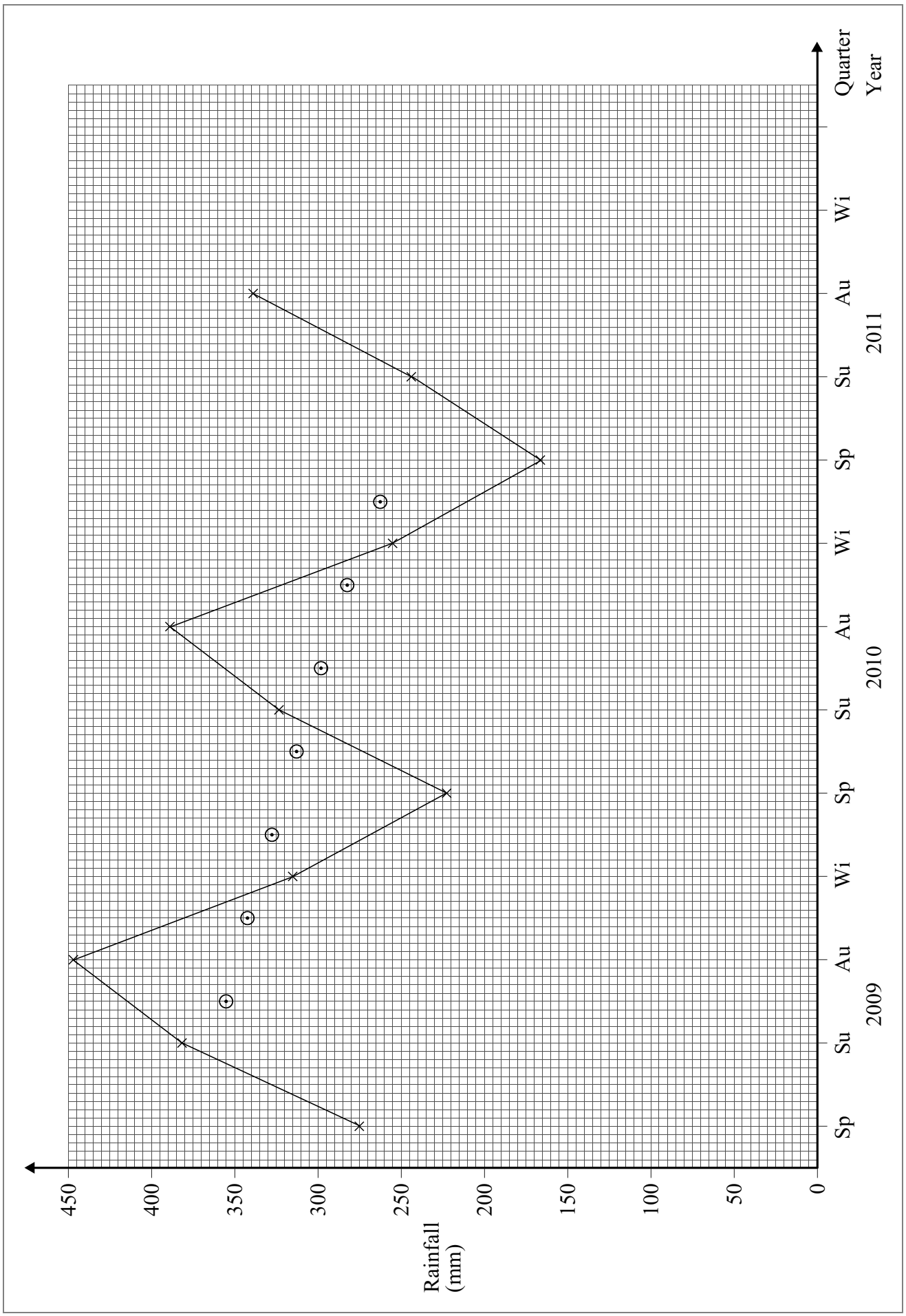
Answer space for question 1

.....

.....

.....

.....


Rainfall (mm)
0
50
100
150
200
250
300
350
400
450
Sp
Su
Au
Wi
Sp
Su
Au
Wi
Sp
Su
Au
Wi
Quarter
Year
2009
2010
2011

QUESTION PART REFERENCE

Answer space for question 1

QUESTION PART REFERENCE

Answer space for question 1

2 A machine fills paper bags with flour. Before maintenance on the machine, the weight of the flour in a bag could be modelled by a normal distribution with mean 1005 grams and standard deviation 2.1 grams. Following this maintenance, the flour in each of a random sample of 8 bags was weighed. The weights, in grams, were as follows.

1006.1 1004.9 1005.8 1007.9 1004.7 1006.3 1007.4 1007.2

(a) Carry out a test, at the 10% significance level, to decide whether the mean weight of flour in a bag filled by the machine had **changed**. Assume that the distribution of weights was still normal with standard deviation 2.1 grams. *(7 marks)*

(b) The flour in each of a random sample of 90 bags was then weighed. For this sample, the mean weight of flour in a bag was 1005.48 grams and the standard deviation was 2.41 grams.

Carry out a test, at the 2% significance level, to decide whether the mean weight of flour in a bag filled by the machine had **increased** from the value of 1005 grams. *(5 marks)*

(c) Explain why you did not have to assume that the weight of flour in a bag was normally distributed in order to carry out the test in part **(b)**. *(2 marks)*

(d) State, with a reason, which of the tests carried out in parts **(a)** and **(b)** might have resulted in a Type II error. *(2 marks)*

QUESTION PART REFERENCE

Answer space for question 2

QUESTION PART REFERENCE

Answer space for question 2

QUESTION PART REFERENCE

Answer space for question 2

QUESTION PART REFERENCE

Answer space for question 2

3 At a Roman site, coins are found at an average rate of 1 coin per $10\,m^2$. Assume that the number of coins found can be modelled by a Poisson distribution.

(a) Determine the probability that, in an area of $10\,m^2$:

(i) at most 2 coins are found; *(1 mark)*

(ii) exactly 4 coins are found. *(2 marks)*

(b) Determine the probability that more than 8 coins are found in an area of $\mathbf{100\,m^2}$. *(3 marks)*

(c) Bronze brooches are less common than coins at this site, and are found at an average rate of 1 brooch per $50\,m^2$. The number of these brooches found is independent of the number of coins found. Assume that the number of bronze brooches found can also be modelled by a Poisson distribution.

(i) Determine the probability that the **total** number of coins and bronze brooches found in an area of $100\,m^2$ is at least 15. *(3 marks)*

(ii) Sometimes, Romans buried a 'hoard' of several coins together. They did not usually bury several bronze brooches together. State, with a reason, which of

- the number of coins found or
- the number of bronze brooches found

is likely to be better modelled by a Poisson distribution. *(2 marks)*

QUESTION PART REFERENCE

Answer space for question 3

QUESTION PART REFERENCE

Answer space for question 3

4 **Table 1**, below, gives information about the sentences given to people in English courts during 2008.

(a) For females from the West Midlands, the total of the percentages for the six types of sentence is 101 . Explain how this has happened. *(1 mark)*

(b) For the North East, find the **number** of males who received an immediate custodial sentence, giving your answer to the nearest thousand. *(2 marks)*

(c) Considering the data for the East Midlands, make **two** comments about the major differences in the proportions of the types of sentence given to males as compared with those given to females. *(2 marks)*

(d) Georgina wishes to illustrate the figures for England using two comparative pie charts. She will use a circle with radius 4 cm to represent 'males'.

(i) Find the angle of the sector representing 'Fine' on the pie chart for males. *(2 marks)*

(ii) Find the radius of the circle that Georgina should use for the pie chart for females. *(3 marks)*

Table 1

People aged 18 or over found guilty of offences: by sex and type of sentence, 2008

Result as a percentage of number of people sentenced

	Absolute or conditional discharge	Fine	All community penalties	Suspended sentence	Immediate custodial sentence	Otherwise dealt with	All sentenced (=100%) (numbers)
Males							
England	**7**	**66**	**11**	**4**	**9**	**3**	**918 380**
North East	13	58	14	4	7	6	54 070
North West	8	64	14	4	9	2	143 856
Yorkshire and The Humber	7	55	12	4	10	12	90 310
East Midlands	7	67	12	4	9	2	72 839
West Midlands	6	66	11	4	10	3	98 004
East	5	71	11	3	8	2	93 709
London	4	73	8	3	9	2	176 876
South East	6	68	11	4	9	2	116 758
South West	10	66	11	3	7	3	71 958
Females							
England	**7**	**77**	**7**	**2**	**3**	**3**	**257 462**
North East	9	75	8	2	2	5	19 375
North West	7	78	9	2	3	1	45 977
Yorkshire and The Humber	8	61	9	2	3	17	23 983
East Midlands	7	79	8	2	2	1	20 309
West Midlands	7	79	7	3	4	1	27 327
East	6	79	8	2	3	1	24 335
London	5	83	5	2	3	1	45 349
South East	8	76	9	2	3	2	31 021
South West	10	77	7	2	2	2	19 786

QUESTION PART REFERENCE

Answer space for questions 4(a) to 4(d)

Question 4 continues on page 15

QUESTION PART REFERENCE

Answer space for questions 4(a) to 4(d)

4 (e) In **Table 2**, below, the populations of the regions of England are shown, together with a value of 'Persons sentenced per thousand of population' calculated from the data in **Table 1**.

(i) The 'Persons sentenced per thousand of population' value for England is missing. Calculate this missing value. *(2 marks)*

(ii) State, with a reason, whether you consider that the data support the opinion that there is much more crime in London than elsewhere in England. *(1 mark)*

Table 2

	Population (thousands)	Persons sentenced per thousand of population
England	**51809.7**	
North East	2584.3	28.4
North West	6897.9	27.5
Yorkshire and The Humber	5258.1	21.7
East Midlands	4451.2	20.9
West Midlands	5431.1	23.1
East	5766.6	20.5
London	7753.6	28.7
South East	8435.7	17.5
South West	5231.2	17.5

QUESTION PART REFERENCE

Answer space for question 4(e)

5 Alex studies five different subjects at school each weekday, Monday to Friday. The number of pieces of homework, X, which Alex is given each day follows the distribution shown in the table.

x	0	1	2	3	4	5
$\mathbf{P}(X = x)$	0.00	0.03	0.12	0.34	0.33	0.18

(a) Show that the mean of X is 3.51, and calculate the variance of X. *(4 marks)*

(b) Find the probability that, on a particular day, Alex is given:

(i) more than 3 pieces of homework; *(1 mark)*

(ii) at least the modal number of pieces of homework; *(2 marks)*

(iii) fewer than the median number of pieces of homework. *(2 marks)*

(c) (i) David, a friend of Alex, suggested that the data in the table could be modelled by a Poisson distribution. Give one reason, apart from a comparison of the values of the mean and the variance found in part **(a)**, why a Poisson distribution would not be a suitable model in this context. *(1 mark)*

(ii) Siobhan, another friend of Alex, suggested that the data in the table could be modelled by the binomial distribution $\mathrm{B}(5, 0.7)$. Explain whether this suggestion is supported by your answer to part **(a)**. *(3 marks)*

QUESTION PART REFERENCE

Answer space for question 5

QUESTION PART REFERENCE

Answer space for question 5

6 The managers of a fast-food chain with several hundred outlets nationwide have hired a small research company based in Manchester to carry out some customer-satisfaction interviews. They categorise their customers into four distinct groups: pensioners, other adults, teenagers and children. They believe that these four groups are in roughly equal proportions. They wish to know the opinions of a sample of 50 pensioners, 50 other adults, 50 teenagers and 50 children.

(a) Name this type of sample. *(1 mark)*

(b) Give **three** reasons why simple random sampling from all the outlets' customers would be impractical. *(3 marks)*

(c) The research company suggests that quota sampling of customers from outlets in Manchester should be used.

(i) Give an advantage to the research company of using this quota sampling.

(ii) Give a reason why the managers might think that this quota sampling is not suitable. *(2 marks)*

(d) Describe how the research company might aim to obtain a representative sample by the use of cluster sampling. *(3 marks)*

QUESTION PART REFERENCE

Answer space for question 6

QUESTION PART REFERENCE

Answer space for question 6

QUESTION PART REFERENCE

Answer space for question 6

END OF QUESTIONS

ACKNOWLEDGEMENT OF COPYRIGHT-HOLDERS AND PUBLISHERS

Question 4 contains public sector information licenced under the Open Government Licence v1.0.