Centre No.						Paper Reference							Surname	Initial(s)
Candidate No.						6	6	9	1	/	0	1	Signature	

Paper Reference(s)

6691/01

Edexcel GCE

Statistics S3

Advanced/Advanced Subsidiary

Wednesday 20 May 2015 – Morning

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
Total	

Materials required for examination
Mathematical Formulae (Pink)

Items included with question papers
Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 6 questions in this question paper. The total mark for this paper is 75.
There are 20 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

This publication may be reproduced only in accordance with Pearson Education Ltd copyright policy.
©2015 Pearson Education Ltd.

Printer's Log. No.
P44850A
W850/R6691/57570 5/1/1/

P 4 4 8 5 0 A 0 1 2 0

Turn over

PEARSON

Leave blank

1. A mobile library has 160 books for children on its records. The librarian believes that books with fewer pages are borrowed more often. He takes a random sample of 10 books for children.

(a) Explain how the librarian should select this random sample.

(2)

The librarian ranked the 10 books according to how often they had been borrowed, with 1 for the book borrowed the most and 10 for the book borrowed the least. He also recorded the number of pages in each book. The results are in the table below.

Book	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Borrowing rank	1	2	3	4	5	6	7	8	9	10
Number of pages	50	212	115	80	301	90	356	283	152	317

(b) Calculate Spearman's rank correlation coefficient for these data.

(4)

(c) Test the librarian's belief using a 5% level of significance. State your hypotheses clearly.

(3)

Leave blank

Question 1 continued

Q1

(Total 9 marks)

Leave blank

2. A researcher believes that the mean weight loss of those people using a slimming plan as part of a group is more than 1.5 kg a year greater than the mean weight loss of those using the plan on their own. The mean weight loss of a random sample of 80 people using the plan as part of a group is 8.7 kg with a standard deviation of 2.1 kg. The mean weight loss of a random sample of 65 people using the plan on their own is 6.6 kg with a standard deviation of 1.4 kg.

(a) Stating your hypotheses clearly, test the researcher's claim. Use a 1% level of significance.

(8)

(b) For the test in part (a), state whether or not it is necessary to assume that the weight loss of a person using this plan has a normal distribution. Give a reason for your answer.

(2)

Leave blank

Question 2 continued

Q2

(Total 10 marks)

Leave blank

3. A nursery has 16 staff and 40 children on its records. In preparation for an outing the manager needs an estimate of the mean weight of the people on its records and decides to take a stratified sample of size 14.

(a) Describe how this stratified sample should be taken.

(3)

The weights, x kg, of each of the 14 people selected are summarised as

$$\sum x = 437 \text{ and } \sum x^2 = 26983$$

(b) Find unbiased estimates of the mean and the variance of the weights of all the people on the nursery's records.

(4)

(c) Estimate the standard error of the mean.

(2)

The estimates of the standard error of the mean for the staff and for the children are 5.11 and 1.10 respectively.

(d) Comment on these values with reference to your answer to part (c) and give a reason for any differences.

(2)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Q3

(Total 11 marks)

Leave blank

4. The weights of bags of rice, X kg, have a normal distribution with unknown mean μ kg and known standard deviation σ kg. A random sample of 100 bags of rice gave a 90% confidence interval for μ of (0.4633, 0.5127).

(a) Without carrying out any further calculations, use this confidence interval to test whether or not $\mu = 0.5$

State your hypotheses clearly and write down the significance level you have used.

(3)

A second random sample, of 150 of these bags of rice, had a mean weight of 0.479 kg.

(b) Calculate a 95% confidence interval for μ based on this second sample.

(6)

Leave blank

Question 4 continued

Q4

(Total 9 marks)

Leave blank

5. (i) The volume, B ml, in a bottle of *Burxton's* water has a normal distribution $B \sim N(325, 6^2)$ and the volume, H ml, in a bottle of *Hargate's* water has a normal distribution $H \sim N(330, 4^2)$.
Rebecca buys 5 bottles of *Burxton's* water and one bottle of *Hargate's* water.

Find the probability that the total volume in the 5 bottles of *Burxton's* water is more than 5 times the volume in the bottle of *Hargate's* water.

(5)

(ii) Two independent random samples X_1, X_2, X_3, X_4, X_5 and Y_1, Y_2, Y_3, Y_4, Y_5 are each taken from a normal population with mean μ and standard deviation σ.

(a) Find the distribution of the random variable $D = Y_1 - \overline{X}$

(3)

(b) Hence show that $P(Y_1 > \overline{X} + \sigma) = 0.181$ correct to 3 decimal places.

(2)

Ankit believes that $P(U_1 > \overline{U} + \sigma) = 0.181$ correct to 3 decimal places, for **any** random sample U_1, U_2, U_3, U_4, U_5 taken from a normal population with mean μ and standard deviation σ.

(c) Explain briefly why the result from part (b) should not be used to confirm Ankit's belief.

(1)

(d) Find, correct to 3 decimal places, the actual value of $P(U_1 > \overline{U} + \sigma)$.

(6)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Q5

(Total 17 marks)

Leave blank

6.

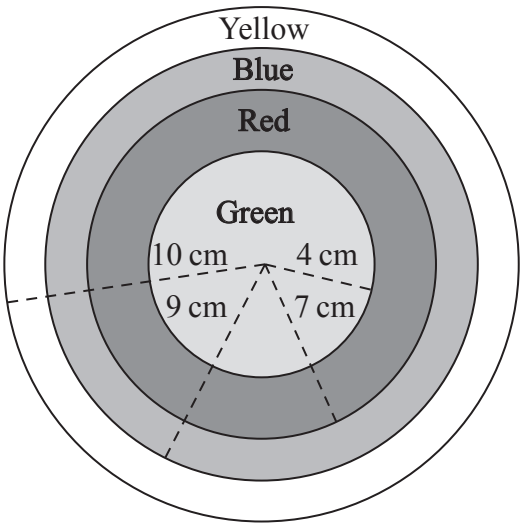


Figure 1

The sketch in Figure 1 represents a target which consists of 4 regions formed from 4 concentric circles of radii 4 cm, 7 cm, 9 cm and 10 cm. The regions are coloured as labelled in Figure 1.
A random sample of 100 children each choose a point on the target and their results are summarised in the table below.

Colour of region	Green	Red	Blue	Yellow
Frequency	22	39	25	14

Caitland is trying to model the distribution of the points chosen by the children. She defines the random variable D to be the distance, in cm, of a point from the centre of the target and assumes $D \sim \text{U}[0, 10]$.

(a) Stating your hypotheses clearly and using a 1% level of significance, test whether or not U[0, 10] is a suitable model for these data.

(9)

Question 6 continues on page 18

Leave blank

Question 6 continued

Question 6 continues on page 18

Leave blank

Question 6 continued

Henry claims that the points are randomly distributed over the target and the probability of a point being in any particular region is proportional to the area of that region.
He calculates expected frequencies and obtains the following table.

Colour of region	Green	Red	Blue	Yellow
Expected frequency	16	33	r	s

(b) Find the value of r and the value of s.

(3)

Henry obtained a test statistic of 6.188 and no groups were pooled.

(c) State what conclusion Henry should make about his claim.

(2)

Phoebe believes that the children chose the region of the target according to colour. She believes that boys and girls would favour different colours and splits the original data by gender to obtain the following table.

Observed frequencies

Colour of region	Green	Red	Blue	Yellow	**Total**
Boys	10	12	10	3	35
Girls	12	27	15	11	65

(d) State suitable hypotheses to test Phoebe's belief.

(1)

Phoebe calculated the following expected frequencies to carry out a suitable test.

Expected frequencies

Colour of region	Green	Red	Blue	Yellow
Boys	7.7	13.65	8.75	4.9
Girls	14.3	25.35	16.25	9.1

(e) Show how the value of 25.35 was obtained.

(1)

Phoebe carried out the test using 2 degrees of freedom and a 10% level of significance. She obtained a test statistic of 1.411

(f) Explain clearly why Phoebe used 2 degrees of freedom.

(1)

(g) Stating your critical value clearly, determine whether or not these data support Phoebe's belief.

(2)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 19 marks)

TOTAL FOR PAPER: 75 MARKS

END