Edexcel GCE

Statistics S3

Advanced Subsidiary

Thursday 9 June 2005 – Morning

Time: 1 hour 30 minutes

Materials required for examination	**Items included with question papers**
Mathematical Formulae (Lilac)	Nil
Graph Paper (ASG2)	

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S3), the paper reference (6685), your surname, other name and signature.
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.
This paper has seven questions.
The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. (*a*) State two reasons why stratified sampling might be chosen as a method of sampling when carrying out a statistical survey.

(2)

(*b*) State one advantage and one disadvantage of quota sampling.

(2)

(Total 4 marks)

---

2. A sample of size 5 is taken from a population that is normally distributed with mean 10 and standard deviation 3. Find the probability that the sample mean lies between 7 and 10.

(Total 6 marks)

---

3. A researcher carried out a survey of three treatments for a fruit tree disease. The contingency table below shows the results of a survey of a random sample of 60 diseased trees.

	No action	Remove diseased branches	Spray with chemicals
Tree died within 1 year	10	5	6
Tree survived for 1–4 years	5	9	7
Tree survived beyond 4 years	5	6	7

Test, at the 5% level of significance, whether or not there is any association between the treatment of the trees and their survival. State your hypotheses and conclusion clearly.

(Total 11 marks)

4. Over a period of time, researchers took 10 blood samples from one patient with a blood disease. For each sample, they measured the levels of serum magnesium, s mg/dl, in the blood and the corresponding level of the disease protein, d mg/dl. The results are shown in the table.

s	1.2	1.9	3.2	3.9	2.5	4.5	5.7	4.0	1.1	5.9
d	3.8	7.0	11.0	12.0	9.0	12.0	13.5	12.2	2.0	13.9

[Use $\Sigma s^2 = 141.51$, $\Sigma d^2 = 1081.74$ and $\Sigma sd = 386.32$]

(*a*) Draw a scatter diagram to represent these data.

(3)

(*b*) State what is measured by the product moment correlation coefficient.

(1)

(*c*) Calculate S_{xx}, S_{dd} and S_{sd}.

(3)

(*d*) Calculate the value of the product moment correlation coefficient r between s and d.

(2)

(*c*) Stating your hypotheses clearly, test, at the 1% significance level, whether or not the correlation coefficient is greater than zero.

(3)

(*d*) With reference to your scatter diagram, comment on your result in part (*e*).

(1)

(Total 13 marks)

---

5. The number of times per day a computer fails and has to be restarted is recorded for 200 days. The results are summarised in the table.

Number of restarts	Frequency
0	99
1	65
2	22
3	12
4	2

Test whether or not a Poisson model is suitable to represent the number of restarts per day. Use a 5% level of significance and state your hypothesis clearly.

(Total 12 marks)

---

6. A computer company repairs large numbers of PCs and wants to estimate the mean time to repair a particular fault. Five repairs are chosen at random from the company's records and the times taken, in seconds, are

$$205 \quad 310 \quad 405 \quad 195 \quad 320.$$

(*a*) Calculate unbiased estimates of the mean and the variance of the population of repair times from which this sample has been taken.

(4)

It is known from previous results that the standard deviation of the repair time for this fault is 100 seconds. The company manager wants to ensure that there is a probability of at least 0.95 that the estimate of the population mean lies within 20 seconds of its true value.

(*b*) Find the minimum sample size required.

(6)

(Total 10 marks)

---

7. A manufacturer produces two flavours of soft drink, cola and lemonade. The weights, *C* and *L*, in grams, of randomly selected cola and lemonade cans are such that $C \sim N(350, 8)$ and $L \sim N(345, 17)$.

(*a*) Find the probability that the weights of two randomly selected cans of cola will differ by more than 6 g.

(6)

One can of each flavour is selected at random.

(*b*) Find the probability that the can of cola weighs more than the can of lemonade.

(6)

Cans are delivered to shops in boxes of 24 cans. The weights of empty boxes are normally distributed with mean 100 g and standard deviation 2 g.

(*c*) Find the probability that a full box of cola cans weighs between 8.51 kg and 8.52 kg.

(6)

(*d*) State an assumption you made in your calculation in part (*c*).

(1)

(Total 19 marks)

TOTAL FOR PAPER: 75 MARKS

END