GCE Examinations

Statistics
Module S3

Advanced Subsidiary / Advanced Level

Paper D

Time: 1 hour 30 minutes

Instructions and Information

Candidates may use any calculator except those with a facility for symbolic algebra and / or calculus.

Full marks may be obtained for answers to ALL questions.

Mathematical and statistical formulae and tables are available.

This paper has 7 questions.

Advice to Candidates

You must show sufficient working to make your methods clear to an examiner. Answers without working will gain no credit.

Written by Shaun Armstrong & Chris Huffer

© Solomon Press

These sheets may be copied for use solely by the purchaser's institute.

1. A charity has 240 volunteers and wishes to consult a sample of them of size 20.

(a) Explain briefly how a systematic sample can be taken using random numbers.

(3 marks)

(b) Give one advantage and one disadvantage of using systematic sampling compared with simple random sampling.

(2 marks)

---

2. A teacher gives each student in his class a list of 30 numbers. All the numbers have been generated at random by a computer from a normal distribution with a fixed mean and variance. The teacher tells the class that the variance of the distribution is 25 and asks each of them to calculate a 95% confidence interval based on their list of numbers.

The sum of the numbers given to one student is 1419.

(a) Find the confidence interval that should be obtained by this student. **(5 marks)**

Assuming that all the students calculate their confidence intervals correctly,

(b) state the proportion of the students you would expect to have a confidence interval that includes the true mean of the distribution,

(1 mark)

(c) explain why the probability of any one student's confidence interval including the true mean is not 0.95

(1 mark)

---

3. A newly promoted manager is present when an experienced manager interviews six candidates, A, B, C, D, E and F for a job. Both managers rank the candidates in order of preference, starting with the best candidate, giving the following lists:

Experienced Manager:	$B \ F \ A \ C \ E \ D$
New Manager:	$F \ C \ B \ D \ E \ A$

(a) Calculate Spearman's rank correlation coefficient for these data. **(5 marks)**

(b) Stating your hypotheses clearly, test at the 5% level of significance whether or not there is evidence of positive correlation.

(4 marks)

(c) Comment on whether the new manager needs training in the assessment of candidates at interview.

(1 mark)

© *Solomon Press*

4. A student collected data on the number of text messages, t, sent by 30 students in her year group in the previous week. Her results are summarised as follows:

$$\Sigma t = 1039, \quad \Sigma t^2 = 65\,393.$$

(a) Calculate unbiased estimates of the mean and variance of the number of text messages sent by these students per week.

(4 marks)

Another student collected similar data for 20 different students and calculated unbiased estimates of the mean and variance of 32.0 and 963.4 respectively.

(b) Calculate unbiased estimates of the mean and variance for the combined sample of 50 students.

(6 marks)

---

5. An organic farm produces eggs which it sells through a local shop. The weight of the eggs produced on the farm are normally distributed with a mean of 55 grams and a standard deviation of 3.9 grams.

(a) Find the probability that two of the farm's eggs chosen at random differ in weight by more than 4 grams.

(5 marks)

The farm sells boxes of six eggs selected at random. The weight of the boxes used are normally distributed with a mean of 28 grams and a standard deviation of 1.2 grams.

(b) Find the probability that a randomly chosen box with six eggs in weighs less than 350 grams.

(6 marks)

Turn over

© *Solomon Press*

6. A survey found that of the 320 people questioned who had passed their driving test aged under twenty-five, 104 had been involved in an accident in the two years following their test. Of the 80 people in the survey who were aged twenty-five or over when they passed their test, 16 had been involved in an accident in the following two years.

(a) Draw up a contingency table showing this information. **(2 marks)**

It is desired to test whether the proportion of drivers having accidents within two years of passing their test is different for those who were aged under twenty-five at the time of passing their test than for those aged twenty-five or over.

(b) (i) Stating your hypotheses clearly, carry out the test at the 5% level of significance.

(ii) Explain clearly why there is only one degree of freedom. **(11 marks)**

It is found that 12 people who were aged under twenty-five when they took their test and had been involved in an accident in the following two years had been omitted from the information given.

(c) Explain why you do not need to repeat the calculation to know the correct result of the test.

(2 marks)

---

7. A shoe manufacturer sees a report from another country stating that the length of adult male feet is normally distributed with a mean of 22.4 cm and a standard deviation of 2.8 cm. The manufacturer wishes to see if this model is appropriate for his customers and collects data on the length, correct to the nearest cm, of the right foot of a random sample of 200 males giving the following results:

Length (cm)	≤ 18	19 – 21	22 – 24	25 – 27	≥ 28
No. of Men	24	48	69	41	18

The expected frequencies for the ≤ 18 and 19 – 21 groups are calculated as 16.46 and 58.44 respectively, correct to 2 decimal places.

(a) Calculate expected frequencies for the other three classes. **(7 marks)**

(b) Stating your hypotheses clearly, test at the 10% level of significance whether or not this data can be modelled by the distribution $N(22.4, 2.8^2)$.

(7 marks)

The manufacturer wishes to refine the model by not assuming a mean and standard deviation.

(c) Explain briefly how the manufacturer should proceed. **(3 marks)**

END

© *Solomon Press*