GCE Examinations

Statistics
Module S1

Advanced Subsidiary / Advanced Level

Paper I

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 6 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

1. *(a)* (i) Name a suitable distribution for modelling the volume of liquid in bottles of wine sold as containing 75 cl.

(ii) Explain why the mean in such a model would probably be greater than 75 cl.

(2 marks)

(b) (i) Name a suitable distribution for modelling the score on a single throw of a fair four-sided die with the numbers 1, 2, 3 and 4 on its faces.

(ii) Use your suggested model to find the mean and variance of the score on a single throw of the die.

(6 marks)

---

2. The events A and B are independent and such that

$$\mathrm{P}(A) = 2\,\mathrm{P}(B) \quad \text{and} \quad \mathrm{P}(A \cap B) = \tfrac{1}{8}.$$

(a) Show that $\mathrm{P}(B) = \tfrac{1}{4}$. **(5 marks)**

(b) Find $\mathrm{P}(A \cup B)$. **(3 marks)**

(c) Find $\mathrm{P}(A \mid B')$. **(2 marks)**

---

3. A call-centre dealing with complaints collected data on how long customers had to wait before an operator was free to take their call.

The lower quartile of the data was 12.7 minutes and the interquartile range was 5.8 minutes.

(a) Find the value of the upper quartile of the data. **(1 mark)**

It is suggested that a normal distribution could be used to model the waiting time.

(b) Calculate correct to 3 significant figures the mean and variance of this normal distribution based on the values of the quartiles.

(8 marks)

The actual mean and variance of the data were 15.3 minutes and 20.1 minutes^2 respectively.

(c) Comment on the suitability of the model. **(2 marks)**

---

4. A College offers evening classes in GCSE Mathematics and English.

In order to assess which age groups were reluctant to use the classes, the College collected data on the age in completed years of those currently attending each course. The results are shown in this back-to-back stem and leaf diagram.

Totals	Mathematics	Age	English	Totals
(6)	9 9 9 8 8 7	1	9 9	(2)
(8)	8 5 3 1 1 1 0 0	2	0 1 3 5 5 8	(6)
(7)	7 6 6 4 2 2 1	3	2 3 7 9	(4)
(4)	9 7 5 4	4	0 2 6 8 9	(5)
(3)	8 6 0	5	0 3 7 7	(4)
(2)	5 2	6	2 4 4 8	(4)
(0)		7	1	(1)

Key: 1 | 3 | 2 means age 31 doing Mathematics and age 32 doing English

(a) Find the median and quartiles of the age in completed years of those attending the Mathematics classes.

(4 marks)

(b) On graph paper, draw a box plot representing the data for the Mathematics class.

(3 marks)

The median and quartiles of the age in completed years of those attending the English classes are 25, 41 and 57 years respectively.

(c) Draw a box plot representing the data for the English class using the same scale as for the data from the Mathematics class.

(3 marks)

(d) Using your box plots, compare and contrast the ages of those taking each class.

(4 marks)

Turn over

5. A netball team are in a league with three other teams from which one team will progress to the next stage of the competition. The team's coach estimates their chances of winning each of their three matches in the league to be 0.6, 0.5 and 0.3 respectively, and believes these probabilities to be independent of each other.

(a) Show that the probability of the team winning exactly two of their three matches is 0.36

(4 marks)

Let the random variable W be the number of matches that the team win in the league.

(b) Find the probability distribution of W. **(4 marks)**

(c) Find $\mathrm{E}(W)$ and $\mathrm{Var}(W)$. **(6 marks)**

(d) Comment on the coach's assumption that the probabilities of success in each of the three matches are independent.

(2 marks)

---

6. The Principal of a school believes that more students are absent on days when the temperature is lower. Over a two-week period in December she records the percentage of students who are absent, A%, and the temperature, T °C, at 9 am each morning giving these results.

T (°C)	4	$^-3$	$^-2$	$^-6$	0	3	7	$^-1$	3	2
A (%)	8.5	14.1	17.0	20.3	17.9	15.5	12.4	12.8	13.7	11.6

(a) Represent these data on a scatter diagram. **(4 marks)**

You may use

$$\Sigma T = 7, \quad \Sigma A = 143.8, \quad \Sigma T^2 = 137, \quad \Sigma A^2 = 2172.66, \quad \Sigma TA = 20.7$$

(b) Calculate the product moment correlation coefficient for these data and comment on the Principal's hypothesis.

(6 marks)

(c) Find an equation of the regression line of A on T in the form $A = p + qT$. **(4 marks)**

(d) Draw the regression line on your scatter diagram. **(2 marks)**

END