

Paper Reference(s)

6684

Edexcel GCE

Statistics S2

Advanced Level

Specimen Paper

Time: 1 hour 30 minutes

Materials required for examination papers

Answer Book (AB16)
Mathematical Formulae (Lilac)
Graph Paper (ASG2)

Items included with question

Nil

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 S2), the paper reference (6684), your surname, other name and signature.

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.

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 school held a disco for years 9, 10 and 11 which was attended by 500 pupils. The pupils were registered as they entered the disco. The disco organisers were keen to assess the success of the event. They designed a questionnaire to obtain information from those who attended.

(a) State one advantage and one disadvantage of using a sample survey rather than a census. (2)

(b) Suggest a suitable sampling frame. (1)

(c) Identify the sampling units. (1)

2. A piece of string AB has length 12 cm. A child cuts the string at a randomly chosen point P , into two pieces. The random variable X represents the length, in cm, of the piece AP .

(a) Suggest a suitable model for the distribution of X and specify it fully (2)

(b) Find the cumulative distribution function of X . (4)

(c) Write down $P(X < 4)$. (1)

3. A manufacturer of chocolates produces 3 times as many soft centred chocolates as hard centred ones.

Assuming that chocolates are randomly distributed within boxes of chocolates, find the probability that in a box containing 20 chocolates there are

(a) equal numbers of soft centred and hard centred chocolates, (3)

(b) fewer than 5 hard centred chocolates. (2)

A large box of chocolates contains 100 chocolates.

(c) Write down the expected number of hard centred chocolates in a large box. (2)

4. A company director monitored the number of errors on each page of typing done by her new secretary and obtained the following results:

No. of errors	0	1	2	3	4	5
No. of pages	37	65	60	49	27	12

- (a) Show that the mean number of errors per page in this sample of pages is 2. (2)
- (b) Find the variance of the number of errors per page in this sample. (2)
- (c) Explain how your answers to parts (a) and (b) might support the director's belief that the number of errors per page could be modelled by a Poisson distribution. (1)

Some time later the director notices that a 4-page report which the secretary has just typed contains only 3 errors. The director wishes to test whether or not this represents evidence that the number of errors per page made by the secretary is now less than 2.

- (d) Assuming a Poisson distribution and stating your hypothesis clearly, carry out this test. Use a 5% level of significance. (6)

5. In Manuel's restaurant the probability of a customer asking for a vegetarian meal is 0.30. During one particular day in a random sample of 20 customers at the restaurant 3 ordered a vegetarian meal.

- (a) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the proportion of vegetarian meals ordered that day is unusually low. (5)

Manuel's chef believes that the probability of a customer ordering a vegetarian meal is 0.10. The chef proposes to take a random sample of 100 customers to test whether or not there is evidence that the proportion of vegetarian meals ordered is different from 0.10.

- (b) Stating your hypotheses clearly, use a suitable approximation to find the critical region for this test. The probability for each tail of the region should be as close as possible to 2.5%. (6)
- (c) State the significance level of this test giving your answer to 2 significant figures. (1)

6. A biologist is studying the behaviour of sheep in a large field. The field is divided up into a number of equally sized squares and the average number of sheep per square is 2.25. The sheep are randomly spread throughout the field.

(a) Suggest a suitable model for the number of sheep in a square and give a value for any parameter or parameters required. (1)

Calculate the probability that a randomly selected sample square contains

(b) no sheep, (1)

(c) more than 2 sheep. (4)

A sheepdog has been sent into the field to round up the sheep.

(d) Explain why the model may no longer be applicable. (1)

In another field, the average number of sheep per square is 20 and the sheep are randomly scattered throughout the field.

(e) Using a suitable approximation, find the probability that a randomly selected square contains fewer than 15 sheep. (7)

7. The continuous random variable X has probability density function $f(x)$ given by

$$f(x) = \begin{cases} \frac{1}{20}x^3, & 1 \leq x \leq 3 \\ 0, & \text{otherwise} \end{cases}$$

- (a) Sketch $f(x)$ for all values of x . (3)
- (b) Calculate $E(X)$. (3)
- (c) Show that the standard deviation of X is 0.459 to 3 decimal places. (3)
- (d) Show that for $1 \leq x \leq 3$, $P(X \leq x)$ is given by $\frac{1}{80}(x^4 - 1)$ and specify fully the cumulative distribution function of X . (5)
- (e) Find the interquartile range for the random variable X . (4)

Some statisticians use the following formula to estimate the interquartile range:

$$\text{interquartile range} = \frac{4}{3} \times \text{standard deviation.}$$

- (f) Use this formula to estimate the interquartile range in this case, and comment. (2)

END