

6685/01

Edexcel GCE

Statistics

Unit S3 Mock paper

Advanced Subsidiary / Advanced

Time: 1 hour 30 minutes

Materials required for the examination

Answer Book (AB04)
Graph Paper (GP02)
Mathematical Formulae

Items included with these question papers

Nil

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

Instructions to Candidates

In the boxes on the Answer Book provided, 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 names 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 including Statistical Formulae and Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has 5 questions. Pages 6, 7 and 8 are blank.

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 will gain no credit.

© 2001 Edexcel Foundation

This publication may only be reproduced in accordance with Edexcel copyright policy.

Edexcel Foundation is a Registered charity.

Edexcel
Success through qualifications

1. A random sample X_1, X_2, \dots, X_{10} is taken from a normal population with mean 100 and standard deviation 14.

(a) Write down the distribution of \bar{X} , the mean of this sample. **(2 marks)**

(b) Find $P(|\bar{X} - 100| > 5)$. **(3 marks)**

2. A random sample of the invoices, for books purchased by the customers of a large bookshop, was classified by book cover (hardback, paperback) and type of book (novel, textbook, general interest). As part of the analysis of these invoices, an approximate χ^2 statistic was calculated and found to be 11.09.

Assuming that there was no need to amalgamate any of the classifications, carry out an appropriate test to determine whether or not there was any association between book cover and type of book. State your hypotheses clearly and use a 5% level of significance.

(6 marks)

3. As part of a research project into the role played by cholesterol in the development of heart disease a random sample of 100 patients was put on a special fish-based diet. A different random sample of 80 patients was kept on a standard high-protein low-fat diet. After several weeks their blood cholesterol was measured and the results summarised in the table below.

Group	Sample size	Mean drop in cholesterol (mg/dl)	Standard deviation
Special diet	100	75	22
Standard diet	80	64	31

(a) Stating your hypotheses clearly and using a 5% level of significance, test whether or not the special diet is more effective in reducing blood cholesterol levels than the standard diet.

(9 marks)

(b) Explain briefly any assumptions you made in order to carry out this test. **(2 marks)**

4. Breakdowns on a certain stretch of motorway were recorded each day for 80 consecutive days. The results are summarised in the table below.

Number of breakdowns	0	1	2	> 2
Frequency	38	32	10	0

It is suggested that the number of breakdowns per day can be modelled by a Poisson distribution.

Using a 5% level of significance, test whether or not the Poisson distribution is a suitable model for these data. State your hypotheses clearly. **(13 marks)**

5. The random variable R is defined as $R = X + 4Y$ where $X \sim N(8, 2^2)$, $Y \sim N(14, 3^2)$ and X and Y are independent.

Find

- (a) $E(R)$, **(2 marks)**
 (b) $\text{Var}(R)$, **(3 marks)**
 (c) $P(R < 41)$ **(3 marks)**

The random variables Y_1 , Y_2 and Y_3 are independent and each has the same distribution as Y . The random variable S is defined as

$$S = \sum_{i=1}^3 Y_i - \frac{1}{2} X.$$

- (d) Find $\text{Var}(S)$. **(4 marks)**
-

6. As part of her statistics project, Deepa decided to estimate the amount of time A-level students at her school spend on private study each week. She took a random sample of students from those studying Arts subjects, Science subjects and a mixture of Arts and Science subjects. Each student kept a record of the time they spent on private study during the third week of term.

(a) Write down the name of the sampling method used by Deepa. **(1 mark)**

(b) Give a reason for using this method and give one advantage this method has over simple random sampling. **(2 marks)**

The results Deepa obtained are summarised in the table below.

Type of student	Sample size	Mean number of hours
Arts	12	12.6
Science	12	14.1
Mixture	8	10.2

(c) Show that an estimate of the mean time spent on private study by A level students at Deepa's school, based on these 32 students is 12.56, to 2 decimal places. **(3 marks)**

The standard deviation of the time spent on private study by students at the school was 2.48 hours.

(d) Assuming that the number of hours spent on private study is normally distributed, find a 95% confidence interval for the mean time spent on private study by A level students at Deepa's school. **(4 marks)**

A member of staff at the school suggested that A level students should spend on average 12 hours each week on private study.

(e) Comment on this suggestion in the light of your interval. **(2 marks)**

7. For one of the activities at a gymnastics competition, 8 gymnasts were awarded marks out of 10 for each of artistic performance and technical ability. The results were as follows.

Gymnast	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>
Technical ability	8.5	8.6	9.5	7.5	6.8	9.1	9.4	9.2
Artistic performance	6.2	7.5	8.2	6.7	6.0	7.2	8.0	9.1

The value of the product moment correlation coefficient for these data is 0.774.

- (a) Stating your hypotheses clearly and using a 1% level of significance, interpret this value. **(5 marks)**
- (b) Calculate the value of the rank correlation coefficient for these data. **(6 marks)**
- (c) Stating your hypotheses clearly and using a 1% level of significance, interpret this coefficient. **(3 marks)**
- (d) Explain why the rank correlation coefficient might be the better one to use with these data. **(2 marks)**

END