Paper Reference(s)

6683

Edexcel GCE

Statistics S1

(New Syllabus)

Advanced/Advanced Subsidiary

Friday 19 January 2001 − Afternoon

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Answer Book (AB16)	Nil
Graph Paper (GP02)	
Mathematical Formulae	

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 S1), the paper reference (6683), 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 6 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 may gain no credit.

N6994 83018/R6683 3/3/3/3/3/3700

1. The students in a class were each asked to write down how many CDs they owned. The student with the least number of CDs had 14 and all but one of the others owned 60 or fewer. The remaining student owned 65. The quartiles for the class were 30, 34 and 42 respectively.

Outliers are defined to be any values outside the limits of $1.5(Q_3 - Q_1)$ below the lower quartile or above the upper quartile.

On graph paper draw a box plot to represent these data, indicating clearly any outliers. **(7 marks)**

---

2. The random variable X is normally distributed with mean 177.0 and standard deviation 6.4.

(*a*) Find $P(166 < X < 185)$. **(4 marks)**

It is suggested that X might be a suitable random variable to model the height, in cm, of adult males.

(*b*) Give two reasons why this is a sensible suggestion. **(2 marks)**

(*c*) Explain briefly why mathematical models can help to improve our understanding of real-world problems. **(2 marks)**

---

3. A fair six-sided die is rolled. The random variable Y represents the score on the uppermost, face.

(*a*) Write down the probability function of Y. **(2 marks)**

(*b*) State the name of the distribution of Y. **(1 mark)**

Find the value of

(*c*) $E(6Y + 2)$, **(4 marks)**

(*d*) $\text{Var}(4Y - 2)$. **(5 marks)**

---

4. The employees of a company are classified as management, administration or production. The following table shows the number employed in each category and whether or not they live close to the company or some distance away.

	Live close	Live some distance away
Management	6	14
Administration	25	10
Production	45	25

An employee is chosen at random.

Find the probability that this employee

(*a*) is an administrator, **(2 marks)**

(*b*) lives close to the company, given that the employee is a manager. **(2 marks)**

Of the managers, 90% are married, as are 60% of the administrators and 80% of the production employees.

(*c*) Construct a tree diagram containing all the probabilities. **(3 marks)**

(*d*) Find the probability that an employee chosen at random is married. **(3 marks)**

An employee is selected at random and found to be married.

(*e*) Find the probability that this employee is in production. **(3 marks)**

5. The following grouped frequency distribution summarises the number of minutes, to the nearest minute, that a random sample of 200 motorists were delayed by roadworks on a stretch of motorway.

Delay (mins)	Number of motorists
4—6	15
7—8	28
9	49
10	53
11—12	30
13—15	15
16—20	10

(*a*) Using graph paper represent these data by a histogram. **(4 marks)**

(*b*) Give a reason to justify the use of a histogram to represent these data. **(1 mark)**

(*c*) Use interpolation to estimate the median of this distribution. **(2 marks)**

(*d*) Calculate an estimate of the mean and an estimate of the standard deviation of these data. **(6 marks)**

One coefficient of skewness is given by

$$\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}.$$

(*e*) Evaluate this coefficient for the above data. **(2 marks)**

(*f*) Explain why the normal distribution may not be suitable to model the number of minutes that motorists are delayed by these roadworks. **(2 marks)**

6. A local authority is investigating the cost of reconditioning its incinerators. Data from 10 randomly chosen incinerators were collected. The variables monitored were the operating time x (in thousands of hours) since last reconditioning and the reconditioning cost y (in £1000). None of the incinerators had been used for more than 3000 hours since last reconditioning.

The data are summarised below,

$$\Sigma x = 25.0,\ \Sigma x^2 = 65.68,\ \Sigma y = 50.0,\ \Sigma y^2 = 260.48,\ \Sigma xy = 130.64.$$

(a) Find S_{xx}, S_{xy}, S_{yy}. **(3 marks)**

(b) Calculate the product moment correlation coefficient between x and y. **(3 marks)**

(c) Explain why this value might support the fitting of a linear regression model of the form $y = a + bx$. **(1 mark)**

(d) Find the values of a and b. **(4 marks)**

(e) Give an interpretation of a. **(1 mark)**

(f) Estimate

(i) the reconditioning cost for an operating time of 2400 hours,

(ii) the financial effect of an increase of 1500 hours in operating time. **(4 marks)**

(g) Suggest why the authority might be cautious about making a prediction of the reconditioning cost of an incinerator which had been operating for 4500 hours since its last reconditioning. **(2 marks)**

END