Paper Reference(s)

6683

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Tuesday 28 May 2002 − Afternoon

Time: 1 hour 30 minutes

Materials required for examination	**Items included with question papers**
Answer Book (AB16)	Nil
Graph Paper (ASG2)	
Mathematical Formulae (Lilac)	

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 seven 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.

1. An unbiased die has faces numbered 1 to 6 inclusive. The die is rolled and the number that appears on the uppermost face is recorded.

(*a*) State the probability of not recording a 6 in one roll of the die.

(1)

The die is thrown until a 6 is recorded.

(*b*) Find the probability that a 6 occurs for the first time on the third roll of the die.

(3)

---

2. Statistical models can be used to describe real world problems. Explain the process involved in the formulation of a statistical model.

(4)

---

3. For the events *A* and *B*,

(*a*) explain in words the meaning of the term $P(B \mid A)$,

(2)

(*b*) sketch a Venn diagram to illustrate the relationship $P(B \mid A) = 0$.

(2)

Three companies operate a bus service along a busy main road. Amber buses run 50% of the service and 2% of their buses are more than 5 minutes late. Blunder buses run 30% of the service and 10% of their buses are more than 5 minutes late. Clipper buses run the remainder of the service and only 1% of their buses run more than 5 minutes late.

Jean is waiting for a bus on the main road.

(*c*) Find the probability that the first bus to arrive is an Amber bus that is more than 5 minutes late. **(2)**

Let *A, B* and *C* denote the events that Jean catches an Amber bus, a Blunder bus and a Clipper bus respectively. Let *L* denote the event that Jean catches a bus that is more than 5 minutes late.

(*d*) Draw a Venn diagram to represent the events *A, B*, C and *L*. Calculate the probabilities associated with each region and write them in the appropriate places on the Venn diagram.

(4)

(*e*) Find the probability that Jean catches a bus that is more than 5 minutes late.

(2)

3

4. A discrete random variable X takes only positive integer values. It has a cumulative distribution function $F(x) = P(X \leq x)$ defined in the table below.

X	1	2	3	4	5	6	7	8
F(x)	0.1	0.2	0.25	0.4	0.5	0.6	0.75	1

(*a*) Determine the probability function, $P(X = x)$, of X.

(3)

(*b*) Calculate $E(X)$ and show that $Var(X) = 5.76$.

(6)

(*c*) Given that $Y = 2X + 3$, find the mean and variance of Y.

(3)

---

5. A random variable X has a normal distribution.

(*a*) Describe two features of the distribution of X.

(2)

A company produces electronic components which have life spans that are normally distributed. Only 1% of the components have a life span less than 3500 hours and 2.5% have a life span greater than 5500 hours.

(*b*) Determine the mean and standard deviation of the life spans of the components.

(6)

The company gives warranty of 4000 hours on the components.

(*c*) Find the proportion of components that the company can expect to replace under the warranty.

(4)

---

6. The labelling on bags of garden compost indicates that the bags weigh 20 kg. The weights of a random sample of 50 bags are summarised in the table below.

Weight in kg	Frequency
14.6 – 14.8	1
14.8 – 18.0	0
18.0 – 18.5	5
18.5 – 20.0	6
20.0 – 20.2	22
20.2 – 20.4	15
20.4 – 21.0	1

(*a*) On graph paper, draw a histogram of these data.

(4)

(*b*) Using the coding $y = 10$(weight in kg $-$ 14), find an estimate for the mean and standard deviation of the weight of a bag of compost.

(6)

[Use $\Sigma f y^2 = 171\,503.75$]

(*c*) Using linear interpolation, estimate the median.

(2)

The company that produces the bags of compost wants to improve the accuracy of the labelling. The company decides to put the average weight in kg on each bag.

(*d*) Write down which of these averages you would recommend the company to use. Give a reason for your answer.

(2)

_____

7. An ice cream seller believes that there is a relationship between the temperature on a summer day and the number of ice creams sold. Over a period of 10 days he records the temperature at 1p.m., t °C, and the number of ice creams sold, c, in the next hour. The data he collects is summarised in the table below.

t	c
13	24
22	55
17	35
20	45
10	20
15	30
19	39
12	19
18	36
23	54

[Use $\Sigma t^2 = 3\,025$, $\Sigma c^2 = 14\,245$, $\Sigma ct = 6\,526$.]

(*a*) Calculate the value of the product moment correlation coefficient between t and c.

(7)

(*b*) State whether or not your value supports the use of a regression equation to predict the number of ice creams sold. Give a reason for your answer. **(2)**

(*c*) Find the equation of the least squares regression line of c on t in the form $c = a + bt$. **(2)**

(*d*) Interpret the value of b. **(1)**

(*e*) Estimate the number of ice creams sold between 1 p.m. and 2 p.m. when the temperature at 1 p.m. is 16 °C.

(3)

(*f*) At 1 p.m. on a particular day, the highest temperature for 50 years was recorded. Give a reason why you should not use the regression equation to predict ice cream sales on that day.

(1)

END