Paper Reference(s)

6683

Edexcel GCE

Statistics S1

(New Syllabus)

Advanced/Advanced Subsidiary

Wednesday 16 January 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. (*a*) Explain briefly what you understand by

(i) a statistical experiment,

(1)

(ii) an event.

(1)

(*b*) State one advantage and one disadvantage of a statistical model.

(2)

---

2. A meteorologist measured the number of hours of sunshine, to the nearest hour, each day for 100 days. The results are summarised in the table below.

Hours of sunshine	Days
1	16
2–4	32
5–6	28
7	12
8	9
9–11	2
12	1

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

(5)

(*b*) Calculate an estimate of the number of days that had between 6 and 9 hours of sunshine.

(2)

---

3. A discrete random variable X has the probability function shown in the table below.

x	0	1	2
$P(X = x)$	$\frac{1}{3}$	a	$\frac{2}{3} - a$

(*a*) Given that $E(X) = \frac{5}{6}$, find a.

(3)

(*b*) Find the exact value of Var (X).

(3)

(*c*) Find the exact value of $P(X \leq 15)$.

(1)

---

4. A contractor bids for two building projects. He estimates that the probability of winning the first project is 0.5, the probability of winning the second is 0.3 and the probability of winning both projects is 0.2.

(*a*) Find the probability that he does not win either project.

(3)

(*b*) Find the probability that he wins exactly one project.

(2)

(*c*) Given that he does not win the first project, find the probability that he wins the second.

(2)

(*d*) By calculation, determine whether or not winning the first contract and winning the second contract are independent events.

(3)

---

5. The duration of the pregnancy of a certain breed of cow is normally distributed with mean μ days and standard deviation σ days. Only 2.5% of all pregnancies are shorter than 235 days and 15% are longer than 286 days.

(*a*) Show that $\mu - 235 = 1.96\sigma$.

(2)

(*b*) Obtain a second equation in μ and σ.

(3)

(*c*) Find the value of μ and the value of σ.

(4)

(*d*) Find the values between which the middle 68.3% of pregnancies lie.

(2)

---

6. Hospital records show the number of babies born in a year. The number of babies delivered by 15 male doctors is summarised by the stem and leaf diagram below.

Babies	(4 | 5 means 45)	Totals
0		(0)
1	9	(1)
2	1 6 7 7	(4)
3	2 2 3 4 8	(5)
4	5	(1)
5	1	(1)
6	0	(1)
7		(0)
8	6 7	(2)

(*a*) Find the median and inter-quartile range of these data.

(3)

(*b*) Given that there are no outliers, draw a box plot on graph paper to represent these data. Start your scale at the origin.

(4)

(*c*) Calculate the mean and standard deviation of these data.

(5)

The records also contain the number of babies delivered by 10 female doctors.

34	30	20	15	6
32	26	19	11	4

The quartiles are 11, 19.5 and 30.

(*d*) Using the same scale as in part (*b*) and on the same graph paper draw a box plot for the data for the 10 female doctors.

(3)

(*e*) Compare and contrast the box plots for the data for male and female doctors.

(2)

7. A number of people were asked to guess the calorific content of 10 foods. The mean s of the guesses for each food and the true calorific content t are given in the table below.

Food	t	s
Packet of biscuits	170	420
1 potato	90	160
1 apple	80	110
Crisp breads	10	70
Chocolate bar	260	360
1 slice white bread	75	135
1 slice brown bread	60	115
Portion of beef curry	270	350
Portion of rice pudding	165	390
Half a pint of milk	160	200

[You may assume that $\Sigma t = 1340$, $\Sigma s = 2310$, $\Sigma ts = 396\,775$, $\Sigma t^2 = 246\,050$, $\Sigma s^2 = 694\,650$.]

(*a*) Draw a scatter diagram, indicating clearly which is the explanatory (independent) and which is the response (dependent) variable.

(3)

(*b*) Calculate, to 3 significant figures, the product moment correlation coefficient for the above data.

(7)

(*c*) State, with a reason, whether or not the value of the product moment correlation coefficient changes if all the guesses are 50 calories higher than the values in the table.

(2)

The mean of the guesses for the portion of rice pudding and for the packet of biscuits are outside the linear relation of the other eight foods.

(*d*) Find the equation of the regression line of s on t excluding the values for rice pudding and biscuits.

(3)

[You may now assume that $S_{ts} = 72\,587$, $S_{tt} = 63\,671.875$, $\bar{t} = 125.625$, $\bar{s} = 187.5$.]

(*e*) Draw the regression line on your scatter diagram.

(2)

(*f*) State, with a reason, what the effect would be on the regression line of including the values for a portion of rice pudding and a packet of biscuits.

(2)

END