Paper Reference(s)

6683

Edexcel GCE

Statistics S1

Advanced Subsidiary

Thursday 9 June 2005 – Morning

Time: 1 hour 30 minutes

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

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.
The total mark for this paper is 75.

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. The scatter diagrams below were drawn by a student.

Diagram *A*

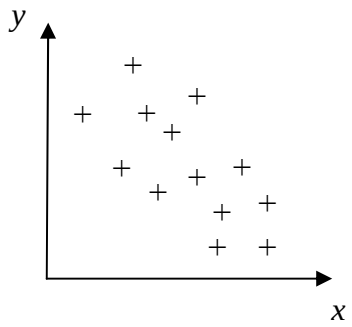


Diagram *B*

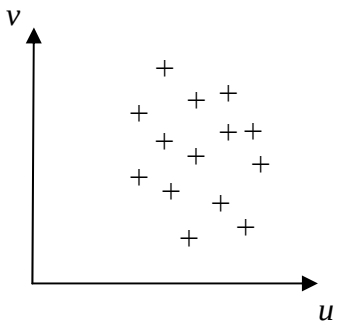
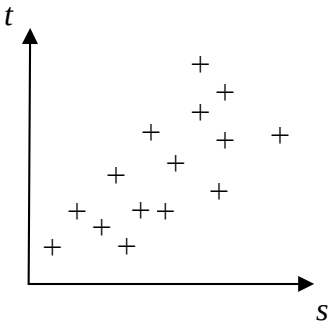


Diagram *C*



The student calculated the value of the product moment correlation coefficient for each of the sets of data.

The values were

0.68 –0.79 0.08

Write down, with a reason, which value corresponds to which scatter diagram.

(6)

---

2. The following table summarises the distances, to the nearest km, that 134 examiners travelled to attend a meeting in London.

Distance (km)	Number of examiners
41–45	4
46–50	19
51–60	53
61–70	37
71–90	15
91–150	6

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

(1)

(*b*) Calculate the frequency densities needed to draw a histogram for these data.

(DO NOT DRAW THE HISTOGRAM)

(2)

(*c*) Use interpolation to estimate the median Q_2, the lower quartile Q_1, and the upper quartile Q_3 of these data.

The mid-point of each class is represented by x and the corresponding frequency by f. Calculations then give the following values

$$\sum fx = 8379.5 \quad \text{and} \quad \sum fx^2 = 557489.75$$

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

(4)

One coefficient of skewness is given by

$$\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}.$$

(*e*) Evaluate this coefficient and comment on the skewness of these data.

(4)

(*f*) Give another justification of your comment in part (*e*).

(1)

---

3. A long distance lorry driver recorded the distance travelled, m miles, and the amount of fuel used, f litres, each day. Summarised below are data from the driver's records for a random sample of 8 days.

The data are coded such that $x = m - 250$ and $y = f - 100$.

$$\sum x = 130 \qquad \sum y = 48 \qquad \sum xy = 8880 \qquad S_{xx} = 20\,487.5$$

(a) Find the equation of the regression line of y on x in the form $y = a + bx$.

(6)

(b) Hence find the equation of the regression line of f on m.

(3)

(c) Predict the amount of fuel used on a journey of 235 miles.

(1)

---

4. Aeroplanes fly from City A to City B. Over a long period of time the number of minutes delay in take-off from City A was recorded. The minimum delay was 5 minutes and the maximum delay was 63 minutes. A quarter of all delays were at most 12 minutes, half were at most 17 minutes and 75% were at most 28 minutes. Only one of the delays was longer than 45 minutes.

An outlier is an observation that falls either 1.5 × (interquartile range) above the upper quartile or 1.5 × (interquartile range) below the lower quartile.

(a) On graph paper, draw a box plot to represent these data.

(7)

(b) Comment on the distribution of delays. Justify your answer.

(2)

(c) Suggest how the distribution might be interpreted by a passenger who frequently flies from City A to City B.

(1)

---

5. The random variable X has probability function

$$P(X = x) = \begin{cases} kx, & x = 1, 2, 3, \\ k(x+1), & x = 4, 5, \end{cases}$$

where k is a constant.

(a) Find the value of k.

(2)

(b) Find the exact value of $E(X)$.

(2)

(c) Show that, to 3 significant figures, $\text{Var}(X) = 1.47$.

(4)

(d) Find, to 1 decimal place, $\text{Var}(4 - 3X)$.

(2)

---

6. A scientist found that the time taken, M minutes, to carry out an experiment can be modelled by a normal random variable with mean 155 minutes and standard deviation 3.5 minutes.

Find

(a) $P(M > 160)$,

(3)

(b) $P(150 \leq M \leq 157)$,

(4)

(c) the value of m, to 1 decimal place, such that $P(M \leq m) = 0.30$.

(4)

---

7. In a school there are 148 students in Years 12 and 13 studying Science, Humanities or Arts subjects. Of these students, 89 wear glasses and the others do not. There are 30 Science students of whom 18 wear glasses. The corresponding figures for the Humanities students are 68 and 44 respectively.

A student is chosen at random.

Find the probability that this student

(*a*) is studying Arts subjects,

(4)

(*b*) does not wear glasses, given that the student is studying Arts subjects.

(2)

Amongst the Science students, 80% are right-handed. Corresponding percentages for Humanities and Arts students are 75% and 70% respectively.

A student is again chosen at random.

(*c*) Find the probability that this student is right-handed.

(3)

(*d*) Given that this student is right-handed, find the probability that the student is studying Science subjects.

(3)

TOTAL FOR PAPER:75 MARKS

END