6683/01

Edexcel GCE

Statistics

Unit S1 Mock paper

Advanced Subsidiary / Advanced

Time: 1 hour 30 minutes

<u>Materials required for the examination</u>   <u>Items included with these question papers</u>

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

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 S1), the Paper Reference (6683), 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 6 questions. There are no blank pages.

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.

Edexcel
Success through qualifications

1. The weight of coffee in glass jars labelled 100 g is normally distributed with mean 101.80 g and standard deviation 0.72 g. The weight of an empty glass jar is normally distributed with mean 260.00 g and standard deviation 5.45 g. The weight of a glass jar is independent of the weight of the coffee it contains.

    Find the probability that a randomly selected jar weighs less than 266 g and contains less than 100 g of coffee. Give your answer to 2 significant figures. **(8 marks)**

---

2. A botany student counted the number of daisies in each of 42 randomly chosen areas of 1 m by 1 m in a large field. The results are summarised in the following stem and leaf diagram.

Number of daisies								1	1	means 11
1	1	2	2	3	4	4	4		(7)	
1	5	5	6	7	8	9	9		(7)	
2	0	0	1	3	3	3	3	4	(8)	
2	5	5	6	7	9	9	9		(7)	
3	0	0	1	2	4	4			(6)	
3	6	6	7	8	8				(5)	
4	1	3							(2)	

    (*a*) Write down the modal value of these data. **(1 mark)**

    (*b*) Find the median and the quartiles of these data. **(4 marks)**

    (*c*) On graph paper and showing your scale clearly, draw a box plot to represent these data. **(4 marks)**

    (*d*) Comment on the skewness of this distribution. **(1 mark)**

    The student moved to another field and collected similar data from that field.

    (*e*) Comment on how the student might summarise both sets of raw data before drawing box plots. **(1 mark)**

---

3. Data relating to the lifetimes (to the nearest hour) of a random sample of 200 light bulbs from the production line of a manufacturer were summarised in a group frequency table. The mid-point of each group in the table was represented by x and the corresponding frequency for that group by f. The data were then coded using $y = \dfrac{(x - 755.0)}{2.5}$ and summarised as follows:

$$\Sigma fy = -467, \quad \Sigma fy^2 = 9179.$$

(*a*) Calculate estimates of the mean and the standard deviation of the lifetimes of this sample of bulbs. **(9 marks)**

An estimate of the interquartile range for these data was 27.7 hours.

(*b*) Explain, giving a reason, whether you would recommend the manufacturer to use the interquartile range or the standard deviation to represent the spread of lifetimes of the bulbs from this production line. **(2 marks)**

---

4. A customer wishes to withdraw money from a cash machine. To do this it is necessary to type a PIN number into the machine. The customer is unsure of this number. If the wrong number is typed in, the customer can try again up to a maximum of four attempts in total. Attempts to type in the correct number are independent and the probability of success at each attempt is 0.6.

(*a*) Show that the probability that the customer types in the correct number at the third attempt is 0.096. **(2 marks)**

The random variable A represents the number of attempts made to type in the correct PIN number, regardless of whether or not the attempt is successful.

(*b*) Find the probability distribution of A. **(2 marks)**

(*c*) Calculate the probability that the customer types in the correct number in four or fewer attempts. **(2 marks)**

(*d*) Calculate E(A) and Var(A). **(6 marks)**

(*e*) Find F(1 + E(A)). **(2 marks)**

---

5. A keep-fit enthusiast swims, runs or cycles each day with probabilities 0.2, 0.3 and 0.5 respectively. If he swims he then spends time in the sauna with probability 0.35. The probabilities that he spends time in the sauna after running or cycling are 0.2 and 0.45 respectively.

(*a*) Represent this information on a tree diagram. **(3 marks)**

(*b*) Find the probability that on any particular day he uses the sauna. **(3 marks)**

(*c*) Given that he uses the sauna one day, find the probability that he had been swimming. **(3 marks)**

(*d*) Given that he did not use the sauna one day, find the probability that he had been swimming. **(6 marks)**

---

6. To test the heating of tyre material, tyres are run on a test rig at chosen speeds under given conditions of load, pressure and surrounding temperature. The following table gives values of x, the test rig speed in miles per hour (mph), and the temperature, y °C, generated in the shoulder of the tyre for a particular tyre material.

x (mph)	15	20	25	30	35	40	45	50
y (°C)	53	55	63	65	78	83	91	101

(*a*) Draw a scatter diagram to represent these data. **(3 marks)**

(*b*) Give a reason to support the fitting of a regression line of the form $y = a + bx$ through these points. **(1 mark)**

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

$$\text{(You may use } \Sigma\, x^2 = 9\,500, \Sigma\, y^2 = 45\,483, \Sigma\, xy = 20\,615)$$

(*d*) Give an interpretation for each of a and b. **(2 marks)**

(*e*) Use your line to estimate the temperature at 50 mph and explain why this estimate differs from the value given in the table. **(2 marks)**

A tyre specialist wants to estimate the temperature of this tyre material at 12 mph and 85 mph.

(*f*) Explain briefly whether or not you would recommend the specialist to use this regression equation to obtain these estimates. **(4 marks)**

END