

1. Thirty cards, marked with the even numbers from 2 to 60 inclusive, are shuffled and one card is withdrawn at random and then replaced. The random variable X takes the value of the number on the card each time the experiment is repeated.
- (a) What must be assumed about the cards if the distribution of X is modelled by a discrete uniform distribution? (1 mark)
- (b) Making this modelling assumption, find the expectation and the variance of X . (5 marks)

2. (a) Explain briefly why, for data grouped in unequal classes, the class with the highest frequency may not be the modal class. (2 marks)

In a histogram drawn to represent the annual incomes (in thousands of pounds) of 1000 families, the modal class was 15 - 20 (i.e. £ x , where $15\,000 \leq x < 20\,000$), with frequency 300. The highest frequency in a class was 400, for the class 30 - 40, and the bar representing this class was 8 cm high. The total area under the histogram was 50 cm^2 .

- (b) Find the height and the width of the bar representing the modal class. (7 marks)
3. The variable X represents the marks out of 150 scored by a group of students in an examination. The following ten values of X were obtained:
- 60, 66, 76, 80, 94, 106, 110, 116, 124, 140.
- (a) Write down the median, M , of the ten marks. (1 mark)
- (b) Using the coding $y = \frac{x - M}{2}$, and showing all your working clearly, find the mean and the standard deviation of the marks. (6 marks)
- (c) Find $E(3X - 5)$. (3 marks)

4. The discrete random variable X has probability function $P(X = x) = k(x + 4)$. Given that X can take any of the values $-3, -2, -1, 0, 1, 2, 3, 4$,
- (a) find the value of the constant k . (3 marks)
- (b) Find $P(X < 0)$. (2 marks)

- (c) Show that the cumulative distribution $F(x)$ is given by

$$F(x) = \lambda(x + 4)(x + 5)$$

where λ is a constant to be found. (6 marks)

STATISTICS 1 (A) TEST PAPER 4 Page 2

5. The events A and B are such that $P(A \cap B) = 0.24$, $P(A \cup B) = 0.88$ and $P(B) = 0.52$.
- (a) Find $P(A)$. (3 marks)
 - (b) Determine, with reasons, whether A and B are
 - (i) mutually exclusive,
 - (ii) independent. (4 marks)
 - (c) Find $P(B|A)$. (2 marks)
 - (d) Find $P(A' | B')$. (3 marks)

6. The times taken by a group of people to complete a task are modelled by a normal distribution with mean 8 hours and standard deviation 2 hours.

Use this model to calculate

- (a) the probability that a person chosen at random took between 5 and 9 hours to complete the task, (4 marks)
- (b) the range, symmetrical about the mean, within which 80% of the people's times lie. (5 marks)

It is found that, in fact, 80% of the people take more than 5 hours. The model is modified so that the mean is still 8 hours but the standard deviation is no longer 2 hours.

- (c) Find the standard deviation of the times in the modified model. (3 marks)

7. The following data was collected for seven cars, showing their engine size, x litres, and their fuel consumption, y km per litre, on a long journey.

Car	A	B	C	D	E	F	G
x	0.95	1.20	1.37	1.76	2.25	2.50	2.875
y	21.3	17.2	15.5	19.1	14.7	11.4	9.0

$$\sum x = 12.905, \quad \sum x^2 = 26.8951, \quad \sum y = 108.2, \quad \sum y^2 = 1781.64, \quad \sum xy = 183.176.$$

- (a) Calculate the equation of the regression line of x on y , expressing your answer in the form $x = ay + b$. (6 marks)
- (b) Calculate the product moment correlation coefficient between y and x and give a brief interpretation of its value. (4 marks)
- (c) Use the equation of the regression line to estimate the value of x when $y = 12$.
State, with a reason, how accurate you would expect this estimate to be. (3 marks)
- (d) Comment on the use of the line to find values of x as y gets very small. (2 marks)