STATISTICS (A) UNIT 1 **TEST PAPER 9**

1. Briefly describe what is meant by

   (a) a statistical model, **(2 marks)**

   (b) a refinement of a model. **(2 marks)**

2. The random variable X has the discrete uniform distribution and takes the values $\{1, \ldots, n\}$. The standard deviation of of X is $2\sqrt{6}$. Find

   (a) the mean of X, **(3 marks)**

   (b) $P(3 \leq X < \frac{1}{2}n)$. **(3 marks)**

3. The rainfall at a weather station was recorded every day of the twentieth century. One year is selected at random from the records and the total rainfall, in cm, in January of that year is denoted by R. Assuming that R can be modelled by a normal distribution with standard deviation 12·6, and given that $P(R > 100) = 0{\cdot}0764$,

   (a) find the mean of R, **(4 marks)**

   (b) calculate $P(75 < R < 80)$. **(5 marks)**

4. The length of time, in minutes, that visitors queued for a tourist attraction is given by the following table, where, for example, '20 - ' means from 20 up to but not including 30 minutes.

Queuing time (mins)	0 -	10 -	15 -	20 -	30 -	40 - 60
Number of visitors	15	24	x	13	10	y

   (a) State the upper class boundary of the first class. **(1 mark)**

   A histogram is drawn to represent this data. The total area under the histogram is 36 cm^2. The '10 - ' bar has width 1 cm and height 9·6 cm. The '15 - ' bar is ten times as high as the '40 - 60' bar.

   (b) Find the values of x and y. **(7 marks)**

   (c) On graph paper, construct the histogram accurately. **(5 marks)**

5. The discrete random variable X takes only the values 4, 5, 6, 7, 8 and 9. The probabilities of these values are given in the table:

x	4	5	6	7	8	9
$P(X = x)$	p	0.1	q	q	0.3	0.2

   It is known that $E(X) = 6{\cdot}7$. Find

   (a) the values of p and q, **(7 marks)**

   (b) the value of a for which $E(2X + a) = 0$, **(3 marks)**

   (c) $Var(X)$. **(3 marks)**

6. The marks out of 75 obtained by a group of ten students in their first and second Statistics modules were as follows:

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Module 1 (x)	54	33	42	71	60	27	39	46	59	64
Module 2 (y)	50	22	44	58	42	19	35	46	55	60

(a) Find $\sum x$ and $\sum y$. **(2 marks)**

Given that $\sum x^2 = 26\,353$ and $\sum xy = 22\,991$,

(b) obtain the equation of the regression line of y on x. **(5 marks)**

(c) Estimate the Module 2 result of a student whose mark in Module 1 was (i) 65, (ii) 5. Explain why one of these estimates is less reliable than the other. **(4 marks)**

The equation of the regression line of x on y is $x = 0{\cdot}921y + 9{\cdot}81$.

(d) Deduce the product moment correlation coefficient between x and y, and briefly interpret its value. **(4 marks)**

7. Among the families with two children in a large city, the probability that the elder child is a boy is $\frac{5}{12}$ and the probability that the younger child is a boy is $\frac{9}{16}$. The probability that the younger child is a girl, given that the elder child is a girl, is $\frac{1}{4}$.

One of the families is chosen at random. Using a tree diagram, or otherwise,

(a) show that the probability that both children are boys is $\frac{1}{8}$. **(5 marks)**

Find the probability that

(b) one child is a boy and the other is a girl, **(3 marks)**

(c) one child is a boy given that the other is a girl. **(3 marks)**

If three of the families are chosen at random,

(d) find the probability that exactly two of the families have two boys. **(3 marks)**

(e) State an assumption that you have made in answering part (d). **(1 mark)**

