STATISTICS (C) UNIT 1 TEST PAPER 6

1. Describe what is meant by
(i) a measure of central tendency, (ii) a measure of variation,
giving the names for two examples of each. [4]

2. The discrete random variable X has probability distribution

$$f(x) = \begin{cases} k(x+1) & x = 1, 2, \ldots, 19, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that $k = {}^{1}/_{209}$. [3]
(ii) Find $P(3 < X \leq 6)$. [3]

3. Mike and Norma play two games of tennis every Saturday. The probability that Mike wins the first game is p. If Mike wins the first game then the probability that he wins the second is 0·8. If Norma wins the first game then the probability that she wins the second is 0·9.
(i) Assuming that there are no drawn games, find in terms of p
(a) m, the probability that Mike wins the second game, [2]
(b) n, the probability that Norma wins the second game. [2]
(ii) Given that $m = n$, show that $p = {}^{4}/_{7}$. [2]
(iii) Using this value of p, calculate the conditional probability that Mike wins the first game, given that he wins the second game. [2]

4. In a large college, the probability that more than 10% of the students are absent on any given day is ${}^{1}/_{5}$. The number of days in a 20-day working month on which more than 10% are absent is denoted by X.
State an assumption that must be made if X is to be modelled by a binomial distribution. [1]
Find the probability that, in one 20-day month, more than 10% of the students are absent on
(i) at most five days, [1]
(ii) more than five days but less than ten days. [2]
There are eight 20-day working months in the college year. A month in which there are more than five days with over 10% of the students absent is deemed to have an unacceptable level of absenteeism.
Find the probability that there is an unacceptable level of absenteeism in more than five of the eight months. [4]

5. The frequency distribution for the lengths of 108 fish in an aquarium is given by the following table.

Length (cm)	5 - 10	10 - 20	20 - 25	25 - 30	30 - 40	40 - 60	60 - 90
Frequency	8	16	20	18	20	14	12

The lengths of the fish range from 5 cm to 90 cm. The median length is 27·8 cm, the upper quartile is 39·5 cm and the interquartile range is 18·75 cm.

(i) On graph paper, draw a box and whisker plot of the data, showing your scale. [4]

(ii) Calculate an estimate of the mean length of the fish. [3]

(iii) If the data were represented by a histogram, what would be the ratio of the heights of the shortest and highest bars? [3]

6. The discrete random variable X has the following probability distribution :

x	0	1	2	3	4	5
$P(X = x)$	0·11	0·17	0·2	0·13	p	p^2

(i) Find the value of p. [4]

(ii) Find (a) $P(0 < X \le 2)$, (b) $P(X \ge 3)$. [3]

(iii) Find the mean and the variance of X. [5]

7. The following marks out of 50 were given by two judges to the contestants in a talent contest:

Contestant	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Judge 1 (x)	45	30	43	27	48	25	29	32
Judge 2 (y)	44	27	48	31	40	28	30	26

Given that $\Sigma x = 279$, $\Sigma y = 274$, $\Sigma x^2 = 10\,297$, $\Sigma y^2 = 9\,890$ and $\Sigma xy = 10\,013$,

(i) calculate the product-moment correlation coefficient between the two judges' marks. [2]

(ii) Find an equation of the regression line of x on y. [3]

Contestant I was awarded 20 marks by Judge 2.

(iii) Estimate the mark that this contestant would have received from Judge 1. [2]

(iv) Calculate Spearman's coefficient of rank correlation between the two judges' marks. [4]

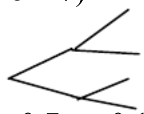
(v) It is sometimes claimed that the coefficient of rank correlation provides a good estimate of the product moment correlation coefficient. Comment on this claim in the light of your answers to parts (i) and (iv). [1]

STATISTICS 1 (C) TEST PAPER 6 : ANSWERS AND MARK SCHEME

1. (i) An 'average' value, e.g. mean, median — B1 B1
   (ii) A measure of spread, e.g. standard deviation, interquartile range — B1 B1 4

2. (i) $k(2 + 3 + \ldots + 20) = 1$ $209k = 1$ $k = {}^{1}/_{209}$ — M1 A1 A1
   (ii) $P(3 < X \le 6) = {}^{1}/_{209}\,(5 + 6 + 7) = {}^{18}/_{209}$ — M1 A1 A1 6

3. (i) Dsraw tree diagram



   (a) $m = 0{\cdot}8p + 0{\cdot}1(1 - p) = 0{\cdot}7p + 0{\cdot}1$ — M1 A1
   (b) $n = 0{\cdot}2p + 0{\cdot}9(1 - p) = 0{\cdot}9 - 0{\cdot}7p$ — M1 A1
   (ii) $0{\cdot}7p + 0{\cdot}1 = 0{\cdot}9 - 0{\cdot}7p$ $1{\cdot}4p = 0{\cdot}8$ $p = {}^{4}/_{7}$ — M1 A1
   (iii) $$P(\text{Mike wins 1st / wins 2nd}) = \frac{P(\text{Mike wins 1st AND wins 2nd})}{P(\text{wins 2nd})} = \frac{\frac{4}{7}\times 0.8}{0.5} = 0{\cdot}914$$
   $P(\text{wins }2^{nd}) = {}^{4}/_{7}\times 0.8 + (1 - {}^{4}/_{7})\times 0.1 = 0.5$ — M1 A1 8

4. Assuming independence, $X \sim B(20, 0{\cdot}2)$ — B1
   (i) $P(X \le 5) = 0{\cdot}804$ — B1
   (ii) $P(5 < X < 10) = P(X \le 9) - P(X \le 5) = 0{\cdot}997 - 0{\cdot}804 = 0{\cdot}193$ — M1 A1
   $B(8, 0{\cdot}1958)$: $P(X > 5) =$ — B1
   $0{\cdot}1958^8 + 8(0{\cdot}1958)^7(0{\cdot}8042) + 28(0{\cdot}1958)^6(0{\cdot}8042)^2 = 0{\cdot}00109$ — M1 A1 A1 8

5. (i) Box plot drawn; scale shown — B4
   (ii) Mean $=[8(7{\cdot}5) + 16(15) + 20(22{\cdot}5) + 18(27{\cdot}5) + 20(35)$ — M1 A1
   $+ 14(50) + 12(75)] \div 108 = 3545 \div 108 = 32{\cdot}8$ cm — A1
   (iii) Freq. densities 1·6, 1·6, 4, 3·6, 2, 0·7, 0·4 Ratio 1 : 10 — B1 M1 A1 10

6. (i) $p^2 + p = 0{\cdot}39$ $(p + 1{\cdot}3)(p - 0{\cdot}3) = 0$ $p = 0{\cdot}3$ — M1 A1 M1 A1
   (ii) (a) $P(0 < X \le 2) = 0{\cdot}17 + 0{\cdot}2 = 0{\cdot}37$ — B1
   (b) $P(X \ge 3) = 0{\cdot}13 + 0{\cdot}3 + 0{\cdot}09 = 0{\cdot}52$ — M1 A1
   (iii) $E(X) = 0{\cdot}17 + 0{\cdot}4 + 0{\cdot}39 + 1{\cdot}2 + 0{\cdot}45 = 2{\cdot}61$ — M1 A1
   $E(X^2) = 9{\cdot}19$ $Var(X) = 9{\cdot}19 - 2{\cdot}61^2 = 2{\cdot}38$ — M1 A1 A1 12

7. (i) $Sxx = 566{\cdot}875$, $Syy = 505{\cdot}5$, $Sxy = 457{\cdot}25$ $r = 0{\cdot}854$ — M1 A1
   (ii) $x - {}^{279}/_{8} = {}^{45725}/_{5055}\,(y - {}^{274}/_{8})$ $x = 0{\cdot}905y + 3{\cdot}89$ — M1 A1 A1
   (iii) Approx. 22 — M1 A1
   (iv)

	7	4	6	2	8	1	3	5	
	7	2	8	5	6	3	4	1	B1
d	0	2	2	3	2	2	1	4	B1

   $\Sigma d^2 = 42$ $r_S = 1 - (6 \times 42)/(8 \times 63) = 0{\cdot}5$ — M1 A1
(v) Not an accurate predictor this case — B1 12