STATISTICS (C) UNIT 1 TEST PAPER 4

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 Thirty cards, marked with the **even** numbers from 2 to 60 inclusive, are shuffled and one card is withdrawn at random and then replaced.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 of times this experiment must be repeated until a number greater than 50 appears.
   What assumption must be made if the distribution of X is modelled by a geometric distribution? [1]
   Making this modelling assumption, find
   (i) the expectation of X, [1]
   (ii) $P(X > 3)$. [2]

2. Nine houses in a village changed hands in 1990 and again in 2000. The selling prices in the two years, in thousands of pounds, are given in the table :

1990 price (£1000s)	135	210	156	174	188	165	143	90	251
2000 price (£1000s)	180	242	195	206	270	184	155	173	400

   Calculate Spearman's coefficient of rank correlation between the prices in the two years. [5]

3. The random variable X has the binomial distribution B(11, 0·76). Find
   (i) $E(X)$, [1]
   (ii) $P(X = 7)$, [2]
   (iii) $P(X \geq 9)$. [3]

4. Explain briefly why, for data grouped in unequal classes, the class with the highest frequency may not be the modal class. [2]
   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 $15\,000 < x < 20\,000$), 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.
   Find the height and the width of the bar representing the modal class. [6]

5. A group of 10 students obtained the following marks out of 150 in an examination:

60, 66, 76, 80, 94, 106, 110, 112, 120, 145.

(i) Find the median mark, M. [1]

The given marks, x, are transformed using the formula $y = x - M$.

(ii) Find the mean and the variance of the values of y obtained by this transformation, and deduce the mean and the standard deviation of the original marks. [7]

(iii) State, with a reason, whether you consider that the median or the mean is a more appropriate measure of central tendency for the data. [2]

6. 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

$\Sigma x = 12{\cdot}905$, $\Sigma x^2 = 26{\cdot}8951$, $\Sigma y = 108{\cdot}2$, $\Sigma y^2 = 1781{\cdot}64$, $\Sigma xy = 183{\cdot}176$.

(i) Calculate the equation of the regression line of x on y, expressing your answer in the form $x = ay + b$. [4]

(ii) Calculate the product moment correlation coefficient between y and x and give a brief interpretation of its value. [3]

(iii) 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]

(iv) Comment on the use of the line to find values of x as y gets very small. [2]

7. The ten letters of the word ASSESSMENT are written on tiles which are arranged in order in a line. Find the number of possible distinguishable arrangements if

(i) there are no restrictions, [2]

(ii) all the vowels come together. [3]

The tiles are now placed in a bag and four tiles are drawn out at random. The variable X represents the number of these tiles which carry the letter 'S'.

(iii) Show that $P(X = 0) = {}^1/_{14}$ and $P(X = 4) = {}^1/_{210}$ [4]

Given that $P(X = 1) = {}^8/_{21}$, $P(X = 2) = {}^3/_7$ and $P(X = 3) = {}^4/_{35}$,

(iv) find the expectation and the standard deviation of X. [6]

STATISTICS 1 (C) TEST PAPER 4 : ANSWERS AND MARK SCHEME

1. Successive trials are independent B1

(i) Geo ($\frac{1}{6}$) : E(X) = 6 B1

(ii) P($X > 3$) = $(1-\frac{1}{6})^3 = {}^{125}/_{216}$ M1 A1 4

2.

	2	8	4	6	7	5	3	1	9
	3	7	5	6	8	4	1	2	9
d	1	1	1	0	1	1	2	1	0

B1 B1 B1

$\Sigma d^2 = 10$ $r_s = 1 - 60 \div (9 \times 80) = 0{\cdot}917$ M1 A1 5

3. (i) E(X) = np = 11 x 0·76 = 8·36 B1

(ii) P($X = 7$) = 11C7(0·76)7(0·24)4 = 0·160 M1 A1

(iii) P($X \geq 9$) = P($X = 11$) + P($X = 10$) + P($X = 9$) M1

$= 0{\cdot}76^{11} + 11(0{\cdot}76)^{10}(0{\cdot}24) + 55(0{\cdot}76)^9(0{\cdot}24)^2 = 0{\cdot}487$ M1 A1 6

4. Frequency density may not be greatest in that class B2

1000 families : 50 cm^2, so 400 families : 20 cm^2 M1

Width of '30 - 40' = 2·5 cm, so width of '15 - 20' = 1·25 cm M1 A1

Area of '15 - 20' = 15 cm^2, so height = 15 ÷ 1·25 = 12 cm B1 M1 A1 8

5. (i) Median = (94 + 106) ÷ 2 = 100 B1

(ii) y values : – 40, – 34, – 24, – 20, – 6, 6, 10, 12, 20, 45 B1

E(Y) = – 3·1 Var(Y) = 637·69 B1 M1 A1

E(X) = – 3·1 + 100 = 96·9 Var(X) = Var(Y) so s.d. = 25·3 M1 A1 A1

(iii) Mean better as it shows bias to lower values B1 B1 10

6. (i) Syy = 109·177, Sxy = – 16·298 x – ${}^{12.905}/_7 = {}^{16.298}/_{109177}$(y – ${}^{108.2}/_7$) M1 A1

x – 1·84357 = – 0·14928(y – 15·4571) x = – 0·149y + 4·15 M1 A1

(ii) Sxx = 3·1038 r = – 0·885 Quite good negative correlation M1 A1 B1

(iii) y = 12 gives $x \approx 2{\cdot}36$ Not necessarily accurate - M1 A1

n is small, which reduces significance of strong correlation B1

(iv) When y is close to 0, x tends to 4·15, suggesting that a 4·15 litre car would travel no km on any amount of fuel – meaningless B2 12

7. (i) 10! ÷ (4! × 2!) = 75 600

(ii) $\frac{8!}{4!} \times \frac{3!}{2!} = 5040$ put vowels as 1 group = 8 letters and mix up vowels M1 A1; M1 M1 A1

(iii) No 'S's : P(X=0) = $\frac{6}{10} \times \frac{5}{9} \times \frac{4}{8} \times \frac{3}{7} = \frac{1}{14}$ B1M1 A1

All 4 'S's : P($X = 4$) = $\frac{4}{10} \times \frac{3}{9} \times \frac{2}{8} \times \frac{1}{7} = \frac{1}{210}$ (iv) E(X) = 1·6 ie $\Sigma x \times P(x=x)$ B1; B1

E(X^2) = 3·2 Var(X) = 3·2 – 1·6^2 = 0·64 s.d. = 0·8 M1 A1 M1 A1 A1 15