STATISTICS (C) UNIT 2 — TEST PAPER 9

1. A video rental shop needs to find out whether or not videos have been rewound when they are returned; it will do this by taking a sample of returned videos
   (i) State one advantage and one disadvantage of taking a sample. [2]
   (ii) Criticise the sampling method of looking at just one particular shelf of videos. [2]

2. The random variable X is modelled by a binomial distribution $B(n, p)$, with $n = 20$ and p unknown. It is suspected that $p = 0{\cdot}3$.
   (i) Find the critical region for the test of $H_0 : p = 0{\cdot}3$ against $H_1 : p \neq 0{\cdot}3$, at the 5% significance level. [3]
   (ii) Find the critical region if, instead, the alternative hypothesis is $H_1 : p < 0{\cdot}3$. [3]

3. A random variable X has the distribution $B(80, 0{\cdot}375)$.
   (i) Write down the mean and variance of X. [2]
   (ii) Use the Normal approximation to the binomial distribution to estimate $P(X > 40)$. [4]

4. The length of time taken on a daily journey is normally distributed, with mean 13·6 minutes and standard deviation 4·1 minutes
   (i) Find the probability that the journey takes longer than 20 minutes. [2]
   (ii) I wish to state a journey time which I know I will achieve in 90% of cases. What time should I give? [3]
   (iii) Find the probability that the mean time in 5 days' journeys is less than 10 minutes. [4]

5. A traffic analyst is interested in the number of heavy lorries passing a certain junction. He counts the number, x, of lorries passing in each of 100 five-minute intervals, and gets the following results : $\sum x = 285$, $\sum x^2 = 1117$.
   (i) Calculate unbiased estimates of the mean and the variance of X. [3]
   (ii) Give a reason for thinking that X can be modelled by a Poisson distribution. [1]

   An environmental group claims that the distribution of X is indeed Poisson, but that the mean is higher and is in fact 4·2. It is also known that the presence of 10 or more lorries in any five-minute interval will cause severe congestion.
   (iii) Assuming the environmentalists are right, find the probability of 10 or more lorries passing in a randomly chosen five-minute interval. [3]
   (iv) If, in the very first interval, only one lorry is counted, decide whether this is evidence against the environmentalists' claim, at the 5% significance level. [3]

6. The random variable X has a continuous uniform distribution on the interval $a \leq X \leq 3a$.

   (i) Without assuming any standard results for this type of distribution, prove that μ, the mean value of X, is equal to $2a$ and derive an expression for σ^2, the variance of X, in terms of a. [6]

   (ii) Find the probability that $|X - \mu| < \sigma$ and compare this with the same probability when x is modelled by a Normal distribution with the same mean and variance. [6]

7. Two people are playing darts. Peg hits points randomly on the circular board, whose radius is a. The distance of the point that she hits from the centre O of the board is modelled by the continuous random variable R.

   (i) Show that the probability that $R \leq r$ is given by

$$\begin{aligned} &\mathrm{P}(R \leq r) = 0 && r < 0, \\ &\mathrm{P}(R \leq r) = \frac{r^2}{a^2} && 0 \leq r \leq a, \\ &\mathrm{P}(R \leq r) = 1 && r > a, \end{aligned}$$

   and hence show that the probability density function for R is given by

$$\begin{aligned} &\mathrm{f}(r) = \frac{2r}{a^2} && 0 \leq r \leq a, \\ &\mathrm{f}(r) = 0 && \text{otherwise.} \end{aligned}$$

[6]

   (ii) Find the mean distance from O of the points that she hits. [2]

Bob, a more experienced player, aims for O, and the points he hits have a distance X from O whose probability density function is

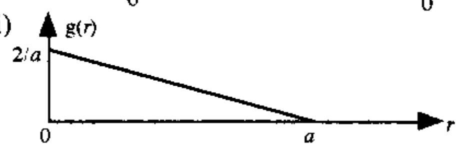
$$\begin{aligned} &\mathrm{g}(x) = \frac{2}{a} - \frac{2x}{a^2} && 0 < x < a, \\ &\mathrm{g}(x) = 0 && \text{otherwise.} \end{aligned}$$

By sketching a graph of $\mathrm{g}(x)$,

   (iii) explain why this function shows that the Bob is aiming for O. [3]

   (iv) Prove that $\mathrm{g}(x)$ is indeed a probability density function [2]

STATISTICS 2 (C) TEST PAPER 9 : ANSWERS AND MARK SCHEME

1.	(i) Quicker to use a sample, but it may be inaccurate	B1 B1	
	(ii) One particular sort, e.g. horror, may be unrepresentative	B2	4
2.	(i) From tables, extreme 2·5% tails are given by $X \le 1$ and $X \ge 11$,	M1 A1	
	so this is the critical region	A1	
	(ii) The bottom 5% tail is given by the region $\{0, 1, 2\}$	M1 M1 A1	6
3.	(i) Mean $= 80 \times 0{\cdot}375 = 30$, variance $= 80 \times 0{\cdot}375 \times 0{\cdot}625 = 18{\cdot}75$	B1 B1	
	(ii) $X \sim B(80, 0{\cdot}375) \approx N(30, 18{\cdot}75)$	B1	
	$P(X > 40) = P(X > 40{\cdot}5) = P(Z > 10{\cdot}5/4{\cdot}33) = P(Z > 2{\cdot}42)$	M1 A1	
	$= 1 - 0{\cdot}9922 = 0{\cdot}0078$	A1	6
4.	(i) $P(X > 20) = P(Z > 6{\cdot}4/4{\cdot}1) = P(Z > 1{\cdot}561) = 1 - 0{\cdot}9407 = 0{\cdot}0593$	M1 A1	
	(ii) 90th percentile given by $z = 1{\cdot}282$,	B1	
	so $T = 13{\cdot}6 + 1{\cdot}282 \times 4{\cdot}1 = 18{\cdot}9$ minutes	M1 A1	
	(iii) $\overline{X}$ is distributed $N(13{\cdot}6, 4{\cdot}1^2/5)$, so $P(\overline{X} < 10)$	B1 B1	
	$= P(Z < -3{\cdot}6 / (4{\cdot}1/\sqrt{5})) = P(Z < -1{\cdot}963) = 0{\cdot}0248$	M1 A1	9
5.	(i) Mean $= 285/100 = 2{\cdot}85$	B1	
	Variance $= 1117/99 - (100/99) \times 2{\cdot}85^2 = 3{\cdot}08$	M1 A1	
	(ii) Mean $\approx$ Variance	B1	
	(iii) $H_0 : \lambda = 4{\cdot}2$	B1	
	Under H_0, $P(X > 9) = 1 - 0{\cdot}9889 = 0{\cdot}0111$	M1 A1	
	(iv) Under H_0, $P(X < 2) = 0{\cdot}0780 > 5\%$ so do not reject H_0	M1 A1 A1	10
6.	(i) $f(x) = \frac{1}{2a}$, $a \le x \le 3a$. $E(X) = \int_a^{3a} \frac{x}{2a}\,dx = \left[\frac{x^2}{4a}\right]_a^{3a} = \frac{8a^2}{4a} = 2a$	B1 M1 A1	
	$E(X^2) = \int_a^{3a} \frac{x^2}{2a}\,dx = \left[\frac{x^3}{6a}\right]_a^{3a} = \frac{13a^2}{3}$ Var$(X) = \frac{a^2}{3}$	M1 A1 A1	
	(ii) $P(\lvert X - \mu \rvert < \sigma) = P(\lvert X - 2a \rvert < \frac{a}{\sqrt{3}}) = \frac{1}{2a} \times 2\frac{a}{\sqrt{3}} = 0{\cdot}577$	M1 A1 A1	
	Normal : $P(\lvert X - \mu \rvert < \sigma) = P(\lvert Z \rvert < 1) = 2(0{\cdot}3413) = 0{\cdot}683$	M1 A1 A1	12
7.	(i) Must land on board, so $P(R \le r) = 0$ $(r < 0)$, $P(R \le r) = 1$ $(r > a)$	B1 B1	
	For $0 \le r \le a$, $P(R \le r) =$ ratio of areas $= \frac{\pi r^2}{\pi a^2} = \frac{r^2}{a^2}$	M1 A1	
	Differentiate to get $f(r) = \frac{2r}{a^2}$ $(0 \le r \le a)$; $f(r) = 0$ otherwise	M1 A1	
	(ii) Mean $= \int_0^a 2r^2/a^2\,dr = [2r^3/3a^2]_0^a = 2a/3$	M1 A1	
	(iii) 	B2	
	From graph, higher probability is nearer to O, so he is aiming at O	B1	
	(iv) Area under graph $= 0{\cdot}5 \times 2/a \times a = 1$; in addition, $g(r)$ is always	B1	
	positive, so it is a pdf	B1	13