STATISTICS (C) UNIT 2 TEST PAPER 2

1. (i) Explain why it is often useful to take samples as a means of obtaining information. [2]

   (ii) It is proposed that a particular road should have speed bumps built, to slow traffic down. To assess this idea, all cars travelling along the road between 10 and 11 am one day are stopped, and the drivers interviewed. Comment on this way of sampling the traffic, and suggest a more suitable way of obtaining a sample. [2]

2. An insurance company conducts its business by using a Call Centre. The average number of calls per minute is 3·5. In the first minute after a TV advertisement is shown, the number of calls received is 7.

   (i) Stating your hypotheses carefully, and working at the 5% significance level, test whether the advertisement has had an effect. [5]

   (ii) Find the number of calls that would be required in the first minute for the null hypothesis to be rejected at the 0·1% significance level. [2]

   (iii) Explain what is meant by a Type I error in this situation, and find its probability. [2]

3. The leaf lengths on a certain type of oak tree are normally distributed, with standard deviation 1·6 cm. A botanist takes a sample of 5 leaves from one such tree. She had originally thought that the mean length of leaf for this type of tree was 8·7 cm.

   (i) At the 1% significance level, and using a two-tailed test, find the rejection region for her hypothesis. [4]

   (ii) If the mean length is in fact 6·7 cm, find the probability of making a Type II error. [5]

4. On average, 35% of the candidates in a certain subject get an A or B grade in their exam. In one year, 100 students take the exam.

   (i) Use a suitable approximation to find the probability that less than a quarter of the total get A or B grades. [6]

   In the following year, 26 out of 100 candidates get A or B grades.

   (ii) Test, at the 2% significance level, whether or not this is consistent with the previous average of 35%. [4]

5. A greengrocer sells apples from a barrel in his shop. He claims that no more than 5% of the apples are of poor quality. When he takes 10 apples out for a customer, 2 of them are bad.
   (i) Stating your hypotheses clearly, test his claim at the 1% significance level. [5]
   (ii) State an assumption that has been made about the selection of the apples. [1]
   (iii) When five other customers also buy 10 apples each, the numbers of bad apples they get are 1, 3, 1, 2 and 1 respectively. By combining all six customers' results, and using a suitable approximation, test at the 1% significance level whether the combined results provide evidence that the proportion of bad apples in the barrel is greater than 5%. [5]
   (iv) Comment briefly on your results from parts (i) and (iii). [1]

6. Some children are asked to mark the centre of a scale 10 cm long. The position they choose is indicated by the variable X, where $0 < X < 10$. Initially, X is modelled as a random variable with a continuous uniform distribution.
   (i) Find the mean and the standard deviation of X. [2]
   (ii) Find $P(4 < X < 6)$. [1]
   It is suggested that a better model would be the distribution with probability density function

$$f(x) = cx(10 - x), \quad 0 < x < 10,$$
$$f(x) = 0 \qquad \text{otherwise.}$$

   (iii) Write down the mean of X. [1]
   (iv) Find c, and hence find the standard deviation of X in this model. [6]
   (v) Find $P(4 < X < 6)$. [2]
   It is then proposed that an even better model for X would be a Normal distribution with the mean and standard deviation found in parts (iii) and (iv).
   (vi) Use these results to find $P(4 < X < 6)$ in the third model. [2]
   (vii) Compare your answer with (v). Which model do you think is most appropriate? [2]

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

1. (i) Quicker / cheaper to carry out; do not destroy whole population B2
   (ii) It does not give a random sample - e.g. misses rush-hour traffic B1
   Take a sample of, say, 5 cars every hour from 7 am to 9 pm B1 4

2. (i) $X \sim \text{Po}(\lambda)$ $\quad H0 : \lambda = 3{\cdot}5$, $\quad H1 : \lambda > 3{\cdot}5$. B1 B1
   Under H0, $P(X \geq 7) = 1 - 0{\cdot}9347 > 5\%$, so do not reject H0 M1 A1 A1
   (ii) Need $P(X \geq n) < 0{\cdot}001$ $\qquad P(X \leq n - 1) > 0{\cdot}999$ M1
   $P(X \leq 10) = 0.999$ so $P(X \geq = 0.001$ $\quad n=11$ A1
   (iii) Type I error is when the null hypothesis is mistakenly rejected -
   in this case, assume the advert has had an effect, when it hasn't B1
   The probability = the significance level, 0·1% or 0·001 B1 9

3.	(i) Two-tailed, 1% significance level gives critical value of z as 2·576	B1	
	so rejection region is outside $8{\cdot}7 \pm 2{\cdot}576 \times 1{\cdot}6 / \sqrt{5}$	M1	
	i.e. reject if $X < 6{\cdot}86$ or if $X > 10{\cdot}5$	A1 A1	
	(ii) If mean = 6·7, $P(6{\cdot}857 < X < 10{\cdot}543) = P(0{\cdot}219 < Z < 5{\cdot}370) = 0{\cdot}413$	M1 A1 A1	
	This is P(Type II error), i.e. prob. of accepting H0 when it is false B2	9	
4.	(i) $X \sim B(100, 0{\cdot}35)$ $X \sim N(35, 22{\cdot}75)$ $P(X < 25) = P(X < 24{\cdot}5)$	M1 A1 M1	
	$= P(Z < -10{\cdot}5/4{\cdot}77) = P(Z < -2{\cdot}201) = 1 - 0{\cdot}9861 = 0{\cdot}0139$	M1 A1 A1	
	(ii) $P(X < 27) = P(X < 26{\cdot}5) = P(Z < -8{\cdot}5/4{\cdot}77) = P(Z < -1{\cdot}782)$	M1 A1	
	$= 1 - 0{\cdot}9682 = 0{\cdot}0372$	A1	
	At the 2% significance level, this is acceptable, and it is consistent		
	with the usual pass rate of 35%	B1	10
5.	(i) $X \sim B(10, p)$; H0 : $p = 0{\cdot}05$, H1 : $p > 0{\cdot}05$	B1 B1	
	Under H0, $P(X \geq 2) = 1 - 0{\cdot}9139 = 0{\cdot}0861 > 1\%$, so accept H0	M1 A1 A1	
	(ii) Assumed that apples are selected randomly	B1	
	(iii) Now have B(60, 0·05), assuming H0. This is approx. Po(3)	M1 A1	
	$P(X \geq 10) = 1 - 0{\cdot}9989 = 0{\cdot}0011 < 1\%$, so reject H0	M1 A1 A1	
	(iv) More data gives greater evidence and can be more decisive	B1	12
6.	(i) Mean = 5 by symmetry Standard dev. = $\sqrt{(100/12)} = 2{\cdot}89$	B1 B1	
	(ii) $P(4 < X < 6) = 2/10 = 0{\cdot}2$	B1	
	(iii) Mean = 5 (iv) Need $c\left[5x^2 - \frac{x^3}{3}\right]_0^{10} = 1$ so $c = {}^{3}/_{500}$	B1; M1 A1	
	Variance $= \int_0^{10} (10x^3 - x^4)\,dx - 5^2$	M1 A1 A1	
	$= 30 - 25 = 5$, so s.d. $= \sqrt{5} = 2{\cdot}24$	A1	
	(v) $P(4 < X < 6) = \int_4^6 (10x - x^2)\,dx = 0{\cdot}296$	M1 A1	
	(vi) $P(4 < X < 6) = P(-1/\sqrt{5} < Z < 1/\sqrt{5}) = P(-0{\cdot}44 < Z < 0{\cdot}44)$	M1	
	$= 2(0{\cdot}1725) = 0{\cdot}345$	A1	
	(vii) Similar; slightly more concentration near the mean for the Normal	B1	
	model. People are aiming at the middle, so this is probably better	B1	16