S3 Jan 2003

1 In a random sample of 70 students taken from a large Sixth Form College it was found that 36 of the students owned mobile phones. Calculate an approximate 95% confidence interval for the proportion of all the students in the College who owned mobile phones. [4]

2 The delegates applying to attend a large conference were asked to state on their application forms whether they required vegetarian meals and whether they required a car parking space. A random sample of 80 of the delegates' application forms showed that 46 required parking spaces of whom 4 were vegetarians, and 34 did not require a parking space of whom 10 were vegetarians.

Using a 5% significance level, test whether, for the delegates applying to attend the conference, requiring a parking space is associated with being a vegetarian. [8]

3 The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{2} & 0 \leqslant x < 1, \\ \frac{1}{2x} & 1 \leqslant x \leqslant e, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the (cumulative) distribution function of X. [4]

(ii) Show that the upper quartile of X is $\sqrt{e}$. [2]

(iii) Find $E(X)$. [2]

4 When an air conditioner in a car is switched on it increases the fuel consumption. A motoring organisation claims that air conditioners reduce the number of miles per gallon by, on average, no more than 2. An environmental pressure group believe that the effect is greater than this. The pressure group conducted a test using a random sample of 8 cars fitted with air conditioners. The number of miles per gallon achieved by each car was measured, under identical test conditions, first with the air conditioner switched off and then with the air conditioner switched on. The results are given below.

Car	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Air conditioner off	38.5	42.5	33.1	41.2	53.9	58.8	39.4	46.4
Air conditioner on	35.4	40.5	30.8	40.1	48.3	52.1	36.6	41.8

(i) Carry out an appropriate t-test, at the 10% significance level, and hence show that there is sufficient evidence to believe that the environmental pressure group is correct. State any assumption necessary to justify the use of this test. [8]

(ii) The motoring organisation insists that their claim is correct and that the sample is unrepresentative. Using your tables, what can you say about the probability of a random sample of size 8 resulting in a test statistic as extreme as that in part (i) if the null hypothesis is in fact true? [2]

5 A survey is conducted of customers using a Post Office. Starting when the Post Office opens in the morning, a count is made of the number of customers up to and including the first person to collect their old age pension. This is repeated on each of a total of 40 different days and the results are summarised below.

Number of customers	1	2	3	4	5	6	7	8	9	10	$\geqslant 11$
Frequency	23	5	3	3	2	1	0	1	0	2	0

It is thought that this distribution may be modelled by a geometric distribution with parameter p, where p is the probability that a customer is collecting their old age pension.

(i) Calculate the mean and hence show that the value of p, estimated from the data, is 0.408, correct to 3 decimal places. [2]

It is intended to test, at the 5% significance level, the goodness of fit of the model to the data.

(ii) Using this geometric model, complete the table of expected frequencies appropriately. [3]

Number of customers	1	2	
Expected frequency	16.33	9.66	

(iii) Carry out the test. [5]

6 A commercial potato grower grows a particular variety of potato using a well-known fertiliser to promote growth. A new fertiliser has become available and the grower decides to conduct a test in which he grows, in adjacent plots, 100 of the seed potatoes using the usual fertiliser and 100 using the new fertiliser. The yields per seed potato, x kg and y kg, can be summarised as follows.

Usual fertiliser: $\Sigma x = 102.2.$

New fertiliser: $\Sigma y = 111.4.$

(i) Assuming that the population standard deviation is 0.210 kg for each population, carry out a hypothesis test, at the 10% significance level, and show that the hypothesis that seed potatoes grown with the new fertiliser have a higher mean yield than those grown with the usual fertiliser would be accepted. [6]

(ii) Calculate a 95% confidence interval for the increase in mean yield. [3]

(iii) Explain what this confidence interval tells the potato grower. [1]

[Question 7 is printed overleaf.]

7 A junior secretary handles the emails that arrive for two company directors, [illegible] emails for Thelma may be assumed to arrive at random and at an average rate of 3.4 per day. Louise's emails may also be assumed to arrive at random, independently of Thelma's, at an average rate of 5.1 per day.

(i) Find the probability that, on one particular day, the junior secretary has to handle more than 10 emails. [3]

(ii) The senior secretary handles the emails addressed to the company website. The number of these emails arriving in a five-day working week can be modelled by a normal distribution with mean 70 and variance 36. The senior secretary claims that she often handles at least twice as many emails in a week as the junior secretary does. Using a suitable approximation, estimate the number of weeks in a working year of 48 weeks in which this would happen. [Do not use a continuity correction.] [7]

2643 STATS 3 Mark Scheme January 2003

1. $p = \frac{56}{70} = 0.8$	B1	At any stage – may be implied
Interval is $0.8 \pm 1.96\sqrt{\frac{0.8(1-0.8)}{70}}$	M1	Calculation of form $p \pm z\sqrt{\frac{p(1-p)}{n}}$
	A1	Relevant use of 1.96
Interval is $0.706 < p < 0.894$	A1 4	Correct interval
OR $56 \pm 1.96\sqrt{70.0.8.(1-0.8)}$	M1	Calculation of form $56 \pm z\sqrt{n.p.(1-p)}$
	A1	Relevant use of 1.96
$0.8 \pm 1.96\sqrt{\frac{0.8(1-0.8)}{70}}$	M1	Divide by 70
Interval is $0.706 < p < 0.894$	A1 4	

2.

	Veg	Non Veg	Total
Parking	4	42	46
No Parking	10	24	34
Total	14	66	80

(contingency table above)	B1	Display data as contingency table
Expected freqs. 8.05, 37.95, 5.95, 28.05	M1	Correct method for at least one cell
	A1	All correct
$\sum\frac{(O-E)^2}{E} = \frac{(\lvert 4-8.05\rvert-0.5)^2}{8.05} + \ldots$	M1	At least 1 correct term.
$= 1.565.. + 0.332 + 2.118 + 0.449..$	M1	Include Yates correction
$= 4.46$	A1	Correct to 1 dp.(4.4 or 4.5)
This is greater than the critical value of 3.84	M1	Comparison with correct table value
Reject null hypothesis - there is evidence of association between vegetarianism and need for car parking space.	A1 8	Conclusion, in context, following correct working (Yates not required)
OR Difference of proportions		
H_0 Proportion of parkers being Veg. = Proportion .non veg	B1	
H_1 Proportion of parkers being Veg. $\neq$ Proportion .non veg		
Combined estimate of $p = \frac{14}{80}$	B1	c.a.o.
$z = \dfrac{\frac{4}{46} - \frac{10}{34}}{\sqrt{\frac{14}{80}\cdot\frac{66}{80}\cdot\left(\frac{1}{46}+\frac{1}{34}\right)}}$	M1	Diff. of proportions – s.d based on estimate of p
	M1	Completely correct form
	A1	Correct substitutions
$= 2.41$	A1	Correct to 1 dp
Comparison with 1.96	M1	As above
Reject H_0 - accept proportions different or there is association between vegetarians and parking spaces	A1 8	As above

3. (i) $0 \le x \le 1$ $F(x) = \frac{1}{2}x$	B1	Completely correct
$1 \le x \le e$ $F(x) = \frac{1}{2} + \int_1^x \frac{1}{2x}\,dx$	M1	Use of $\int f(x)\,dx$ to find function of x.
$= \frac{1}{2} + \frac{1}{2}\ln x$	A1	c.a.o.
$x < 0 \Rightarrow F(x) = 0$ and $x > e \Rightarrow F(x) = 1$	B1 4	
ii) $\frac{1}{2} + \frac{1}{2}\ln q_3 = 0.75$	M1	Substitution in F(x) = 0.75 or equivalent.
$\ln q_3 = 0.5$ => $q_3 = e^{0.5}$ (AG)	A1 2	
(iii) $E(X) = \int_0^1 \frac{1}{2}x\,dx + \int_1^e x\,\frac{1}{2x}\,dx$	M1	Use of $E(X) = \int x.f(x)\,dx$ – both parts.
$= [\frac{1}{4}x^2]_0^1 + [\frac{1}{2}x]_1^e$		
$= \frac{1}{2}e - \frac{1}{4}$ or 1.11(3sf)	A1 2	Accept $\frac{1}{4} + \frac{1}{2}e - \frac{1}{2}$
4. (i) $H_0: \mu_d = 2$, $H_1: \mu_d > 2$	B1	Both hypotheses stated
$\bar{d} = 3.525$	B1	
s = 1.9255..	B1	allow $s^2 = 3.707...$ or biased estimate 3.244
Test statistic is $t = \dfrac{3.525 - 2}{\frac{1.9255}{\sqrt{8}}}$	M1	calculation of form $(\bar{x} - \mu)/(s/\sqrt{n})$ - omission of 2 condoned for M1
$= 2.24$	A1	Correct to 2 decimal places
This is greater than the critical value of 1.415	M1	Comparison of test statistic with 1.415
Reject H_0 - conclude that the reduction in mpg is more than 2	A1	Correct conclusion, stated in context, following correct working
Assumption is that the differences are normally distributed	B1 8	
(ii) P(T>1.895)=0.95, P(T>2.365)=0.975	B1	Use of t tables – may be implied.
Probability of sample this extreme is between 2.5% and 5%	B1 2	Any reasonable statement about probability.
SR. 2 Sample solution		
H_0, H_1 as above ; $\bar{x} = 44.255$, $\bar{y} = 40.7$	B1;B1	
Pooled sample estimate $\hat{\sigma}^2 = 59.64$	B1	
$t = \dfrac{3.525 - 2}{\hat{\sigma}\sqrt{\frac{1}{8} + \frac{1}{8}}} = 0.3949$	M1A0	
Comparison with critical t value = 1.345 Assume common variance and both popnulations normal.	M1A0 B1	Max 6/8

5. (i) $\bar{x} = 2.45$	B1		Mean correct to 2 dp.
$p = \frac{1}{2.45} = 0.408$	B1	2	Allow 40/98
(ii) 1, 2, 3, 4, ...OR ≥ 4			
0.408, 0.241, 0.142, 0.0846, ...0.207	M1		Values up to x =4 at least required
16.33, 9.66, 5.72, 3.38, ... 8.29	M1		Values up to x =4 at least required
	A1		All values correct to 1 dp.
Combine to 4+ giving O = 9, E = 8.29	M1		Combining cells correctly to E $\geq$ 5
$\sum \frac{(O-E)^2}{E} = \frac{(23-16.3)^2}{16.3} + \frac{(5-9.7)^2}{9.7} +$	M1		At least two terms correct
$= 2.754.. + 2.277.. + 1.278.. + 0.059$ $= 6.3$ or 6.4(1dp)	A1		Correct to 1 dp
This is greater than the critical chi-square value of 5.991	M1		Compare sensible statistic with relevant value from chi-square tables – allow 7.815 for M1
There is sufficient evidence to reject H_0 so the geometric distribution is not a good fit.	A1	8	Correct conclusion from correct working, in context.
6.(i) $H_0: \mu_x = \mu_y$ $H_1: \mu_x < \mu_y$	B1		Both hypotheses
$Z = \dfrac{\frac{102.2}{100} - \frac{111.4}{100}}{\sqrt{\frac{0.210^2}{100} + \frac{0.210^2}{100}}}$	M1		Use of correct formula
	A1		Completely correct
$= -3.10.$	A1		Correct to 1 dp
This is less than the critical z value of -1.282	M1		Comparison of sensible z value with -1.282
Sufficient evidence to reject H_0 and accept that the mean yield is greater with the new fertiliser	A1	6	Correct conclusion, in context, following correct working
(ii) Interval is $1.114 - 1.022 \pm 1.96\sqrt{\frac{0.210^2}{100} + \frac{0.210^2}{100}}$	M1		Correct formula – condone 0.210 for M1A1
	A1		Relevant use of 1.96
$0.034 < \mu < 0.150$	A1	3	Correct interval
(iii) The grower can expect an extra yield of between 34 and 150 grammes per seed potato. (95 times out of 100).	B1	1	Clear , helpful interpretation from sensible interval – reference to increase required.

7. (i) $T+L \sim Po(8.5)$	B1	Summing 2 Poissons with correct parameter.
$P(T+L \leq 10) = 0.7634$	M1	Use of Poisson Tables – allow $T+L < 10$, giving 0.6530.
$P(T+L > 10) = 0.237$(3 dp)	A1 3	Correct to 3 dp.
(ii) $\sum_1^5 T + \sum_1^5 L \sim Po(42.5)$	B1	
Approximate by $N(42.5, 42.5)$	M1	Use of Normal Approximation to Poisson
$S - 2(\sum_1^5 T + \sum_1^5 L) \sim N(70-85, 36+4\,.\,42.5)$	M1*	Consideration of $S - 2(\sum_1^5 T + \sum_1^5 L)$
$\sim N(-15, 206)$	M1dep*	Normal, attempt to combine means/variances
	A1	Correct mean and variance
$P(S - 2(\sum_1^5 T + \sum_1^5 L) > 0) = P(z > \frac{15}{\sqrt{206}})$	M1	For standardisation
$= 0.1480$		
$0.1480 \times 48 \approx 7.$	A1 7	c.a.o.
S.R. $2S - (\sum_1^5 T + \sum_1^5 L)$ gets B1,M1,M1,M0,A0,M1,A0 Max 4/7		