OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS

4734

Probability & Statistics 3

Candidates answer on the answer booklet.

OCR supplied materials:
- 8 page answer booklet (sent with general stationery)
- List of Formulae (MF1)

Other materials required:
- Scientific or graphical calculator

Wednesday 19 January 2011
Afternoon

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet. Please write clearly and in capital letters.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- Give non-exact numerical answers correct to 3 significant figures unless a different degree of accuracy is specified in the question or is clearly appropriate.
- You are permitted to use a scientific or graphical calculator in this paper.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- **You are reminded of the need for clear presentation in your answers.**
- The total number of marks for this paper is **72**.
- This document consists of **4** pages. Any blank pages are indicated.

1 A random variable has a normal distribution with unknown mean μ and known standard deviation 0.19. In order to estimate μ a random sample of five observations of the random variable was taken. The values were as follows.

5.44 4.93 5.12 5.36 5.40

Using these five values, calculate,

(i) an estimate of μ, **[1]**

(ii) a 95% confidence interval for μ. **[4]**

2 In a Year 8 internal examination in a large school the Geography marks, G, and Mathematics marks, M, had means and standard deviations as follows.

	Mean	Standard deviation
G	36.42	6.87
M	42.65	10.25

Assuming that G and M have independent normal distributions, find the probability that a randomly chosen Geography candidate scores at least 10 marks more than a randomly chosen Mathematics candidate. Do not use a continuity correction. **[5]**

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

$$f(t) = \begin{cases} 0 & t < 0, \\ \dfrac{a}{e} & 0 \leqslant t < 2, \\ a e^{-\frac{1}{2}t} & t \geqslant 2, \end{cases}$$

where a is a positive constant.

(i) Show that $a = \frac{1}{4}e$. **[3]**

(ii) Find the upper quartile of T. **[4]**

4 A study in 1981 investigated the effect of water fluoridation on children's dental health. In a town with fluoridation, 61 out of a random sample of 107 children showed signs of increased tooth decay after six months. In a town without fluoridation the corresponding number was 106 out of a random sample of 143 children. The population proportions of children with increased tooth decay are denoted by p_1 and p_2 for the towns with fluoridation and without fluoridation respectively. A test is carried out of the null hypothesis $p_1 = p_2$ against the alternative hypothesis $p_1 < p_2$. Find the smallest significance level at which the null hypothesis is rejected. **[7]**

5 An experiment with hybrid corn resulted in yellow kernels and purple kernels. Of a random sample of 90 kernels, 18 were yellow and 72 were purple.

(i) Calculate an approximate 90% confidence interval for the proportion of yellow kernels produced in all such experiments. **[4]**

(ii) Deduce an approximate 90% confidence interval for the proportion of purple kernels produced in all such experiments. **[1]**

(iii) Explain what is meant by a 90% confidence interval for a population proportion. **[2]**

(iv) Mendel's theory of inheritance predicts that 25% of all such kernels will be yellow. State, giving a reason, whether or not your calculations support the theory. **[2]**

6 The continuous random variable X has (cumulative) distribution function given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < \frac{1}{2}, \\ \dfrac{2x-1}{x+1} & \frac{1}{2} \leqslant x \leqslant 2, \\ 1 & x > 2. \end{cases}$$

(i) Given that $Y = \dfrac{1}{X}$, find the (cumulative) distribution function of Y, and deduce that Y and X have identical distributions. **[6]**

(ii) Find $\mathrm{E}(X+1)$ and deduce the value of $\mathrm{E}\left(\dfrac{1}{X}\right)$. **[6]**

7 **(i)** When should Yates' correction be applied when carrying out a χ^2 test? **[1]**

Two vaccines against typhoid fever, A and B, were tested on a total of 700 people in Nepal during a particular year. The vaccines were allocated randomly and whether or not typhoid had developed was noted during the following year. The results are shown in the table.

	Vaccines	
	A	*B*
Developed typhoid	19	4
Did not develop typhoid	310	367

(ii) Carry out a suitable χ^2 test at the 1% significance level to determine whether the outcome depends on the vaccine used. Comment on the result. **[10]**

[Question 8 is printed overleaf.]

8 **(i)** State circumstances under which it would be necessary to calculate a pooled estimate of variance when carrying out a two-sample hypothesis test. **[1]**

(ii) An investigation into whether passive smoking affects lung capacity considered a random sample of 20 children whose parents did not smoke and a random sample of 22 children whose parents did smoke. None of the children themselves smoked. The lung capacity, in litres, of each child was measured and the results are summarised as follows.

For the children whose parents did not smoke: $n_1 = 20$, $\Sigma x_1 = 42.4$ and $\Sigma x_1^2 = 90.43$.

For the children whose parents did smoke: $n_2 = 22$, $\Sigma x_2 = 42.5$ and $\Sigma x_2^2 = 82.93$.

The means of the two populations are denoted by μ_1 and μ_2 respectively.

(a) State conditions for which a t-test would be appropriate for testing whether μ_1 exceeds μ_2. **[1]**

(b) Assuming the conditions are valid, carry out the test at the 1% significance level and comment on the result. **[11]**

(c) Calculate a 99% confidence interval for $\mu_1 - \mu_2$. **[3]**

1(i)	Estμ = sample mean = 5.25	B1 **1**	
(ii)	Use (i) ± zSD SD = 0.19/√5 z = 1.96 5.083 < μ < 5.417	M1 B1 B1 A1 **4** **[5]**	 With √5 seen Rounding to 5.08, 5.42
2	Use $G - M \sim N(-6.23, \sigma^2)$ $\sigma^2 = 6.87^2 + 10.25^2$ $z = (16.23)/\sigma$ = 1.315 Probability = 0.0942 or 0.0943	M1 A1 M1 A1 A1 **[5]**	**Or G-M-10~ N(-16.23, σ²)** Accept 0.094
3(i)	$\int_0^2 a\mathrm{e}^{-1}\,\mathrm{d}t + \int_2^\infty a\mathrm{e}^{-\frac{1}{2}t}\,\mathrm{d}t = 1$ $[a\mathrm{e}^{-1}t] + [-2a\mathrm{e}^{-\frac{1}{2}t}]$ => $a = \frac{1}{4}\mathrm{e}$ AG	M1 A1 A1 **3**	 Properly obtained
(ii)	$\int_{q_3}^\infty \frac{1}{4}e^{1-\frac{1}{2}t}\,\mathrm{d}t = \frac{1}{4}$ $[-\frac{1}{2}\mathrm{e}^{1-\frac{1}{2}t}]$ $-\frac{1}{2}q_3 + 1 = -\ln 2$ => $q_3 = 2(\ln 2 + 1)$ or 3.39	M1 B1 M1 A1 **4** **[7]**	OR $\int_0^2 \frac{1}{4}\mathrm{d}t + \int_2^q \frac{1}{4}\mathrm{e}^{1-t/2}\mathrm{d}t = \frac{3}{4}$ AEF For taking logs (not ln(-)) AEF
4	$\hat{p}_2 = 106/143$, $\hat{p}_1 = 61/107$ = 0.7413 = 0.5701 Pooled est p = 167/250 Variance est = $(^{167}/_{250})(^{83}/_{250})(143^{-1}+107^{-1})$ Test statistic z = (0.7413 – 0.5701)/ SD = 2.84(35) Smallest significance level = 0.23% SR: No pe, B1B0B0M1A1(2.84)M1A1 Max 5/7	B1 B1 B1 M1 A1 M1 A1√ **[7]**	For both Only if used ART 0.22 or 0.23 Accept 0.0023 **√z** M1A0 if 0.25%
5(i)	$s^2 = 0.2 \times 0.8/90$ $p_s \pm zs$ z = 1.645 $0.1306 < p_y < 0.2693$	B1 M1 B1 A1 **4**	OR /89 Art (0.131, 0.269)
(ii)	$0.7306 < p_p < 0.8694$	B1ft **1**	ft (i) Art (0.731, 0**.869)**
(iii)	If a large number of such intervals were calculated from independent samples, approximately 90% of all such intervals would contain p	B2 **2**	Or: Probability that such an interval contains p is 0.9 B1 for right idea
(iv)	(0.131, 0.269) encloses 0.25 so Mendel's theory is supported	M1 A1 √ **2** **[9]**	 Or equivalent Ft CI(i)

6(i)	$G(y) = P(Y \le y)$ $= P(X \ge 1/y)$ $= 1 - F(1/y)$ $= (2y - 1)/(y+1)$ For $\frac{1}{2} \le 1/y \le 2 \Rightarrow \frac{1}{2} \le y \le 2$ X and Y have identical distributions SR: CDF not used. y decreases with x Use $g(y) = f(x(y)\|dx/dy\|)$ $f(x)= 3/(x+1)^2$ $\|dx/dy\|=1/y^2$ $g(y) = [3/(y^{-1}+1)^2][1/y^2] = 3/(y+1)^2$; for $\frac{1}{2}\le y\le 2$ So X and Y have identical distributions	M1 A1 M1 A1 B1 B1 **6** M1 M1A1 B1 M1A1B1 B1 **8**	 Seen
(ii)	$f(x) = F'(x) = 3/(x+1)^2$, $\frac{1}{2}\le x\le 2$ $E(X+1) = \int_{\frac{1}{2}}^{2} \frac{3}{x+1}\,dx$ $=3\ln 2$ (2.08) $E(1/X) = E(X)$ $= 3\ln 2 - 1$ (1.08)	M1A1 M1 A1 M1 A1 **6** **[12]**	Must have range of x AEF Not if awarded in (i)
7(i)	In a 2×2 contingency table	B1 **1**	Or equivalent Accept df=1
(ii)	H_0: Vaccine type and outcome are independent H_1: They are not independent E-values: 10.81 12.19 318.19 358.81 $\chi^2 = 7.69^2(10.81^{-1}+12.19^{-1}+318.19^{-1}+358.81^{-1})$ =10.67 CV = 6.635 **10.67** > CV Reject H_0, there is sufficient evidence at the 1% significance level that the outcome of the test depends on the vaccine used The results is significant at a level less than ½ %, so the evidence is very strong	B1M*dep M1 A1 M1 M1 A1 B1 M1 A1√ dep*M A1 √ **10** **[11]**	Accept omission of H_1 1 correct E value Accept 1 dp 1 correct χ^2 value ft E values Using Yates' correctly Accept 10.7 √ 10.67 Sensible comment.√ 10.67

8(i)	When independent samples are drawn from populations having a common variance	B1 **1**	For common variance
(ii)	**(a)** Lung capacities should have normal distributions with a common variance	B1 **1**	Normal distributions required In context here
	(b) $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 > \mu_2$	B1	Or equivalent
	$s_1^2 = \frac{1}{19}(90.43 - 42.4^2/20)$	M1	For 1 correct s^2
	$s_2^2 = \frac{1}{21}(82.93 - 42.5^2/22)$		
	$\bar{x}_1 = 2.12$ $\bar{x}_2 = 1.93(2)$	B1	For both
	PEV, $s^2 = (19s_1^2 + 21s_2^2)/(20+22-2)$ $= 0.03424(3)$	M1 A1	
	Test statistic = $\frac{2.12-1.932}{\sqrt{s^2(20^{-1}+22^{-1})}}$	M1A1	ft s^2
	$= 3.29(15)$	A1√	Accept answer rounding to 3.3
	CV = 2.423	B1	If z used the B0M0A0
	TS > CV	M1	Compare with CV
	There is sufficient evidence at the 1% SL that the mean lung capacity is greater for children whose parents do not smoke than for children whose parents do smoke	A1 **11**	Or equivalent, in context
	SR1: For 2-tail test Lose 1st B1 and last 3. Max 8/11 SR2: If $s^2 = s_1^2/20 + s_2^2/22$, B1M1A0A0M1A0A1(3.32) B1M1A1 Max 8/11		
	(c) $t = 2.704$	B1	
	$0.1882\ldots \pm ts(20^{-1}+22^{-1})^{1/2}$	M1	Accept 0.19
	(0.0336, 0.3423)	A1 **3** **[16]**	(0.033-0.036,0.342-0.346)