4769 Statistics 4 June 2009

Q1	Follow-through all intermediate results in this question, unless obvious nonsense.			
(i)	$P(X \geq 2) = 1 - \theta - \theta(1-\theta) = (1-\theta)^2$ [o.e.]	M1 A1		
	$L = [\theta]^{n_0} [\theta(1-\theta)]^{n_1} [(1-\theta)^2]^{n-n_0-n_1}$	M1 A1	Product form Fully correct	
	$= \theta^{n_0+n_1}(1-\theta)^{2n-2n_0-n_1}$	A1	BEWARE PRINTED ANSWER	5
(ii)	$\ln L = (n_0+n_1)\ln\theta + (2n-2n_0-n_1)\ln(1-\theta)$	M1 A1		
	$\frac{d\ln L}{d\theta}$	M1		
	$= \frac{n_0+n_1}{\theta} - \frac{2n-2n_0-n_1}{1-\theta}$	A1		
	$= 0$	M1		
	$\Rightarrow (1-\hat{\theta})(n_0+n_1) = \hat{\theta}(2n-2n_0-n_1)$			
	$\Rightarrow \hat{\theta} = \frac{n_0+n_1}{2n-n_0}$	A1		6
(iii)	$E(X) = \sum_{x=0}^{\infty} x\theta(1-\theta)^x$	M1		
	$= \theta\{0 + (1-\theta) + 2(1-\theta)^2 + 3(1-\theta)^3 + \ldots\}$ $= \frac{1-\theta}{\theta}$	A2	Divisible, for algebra; e.g. by "GP of GPs" BEWARE PRINTED ANSWER	
	So could sensibly use (method of moments)			
	$\tilde{\theta}$ given by $\frac{1-\tilde{\theta}}{\tilde{\theta}} = \bar{X}$	M1		
	$\Rightarrow \tilde{\theta} = \frac{1}{1+\bar{X}}$	A1	BEWARE PRINTED ANSWER	
	To use this, we need to know the exact numbers of faults for components with "two or more".	E1		6
(iv)	$\bar{x} = \frac{14}{100} = 0\cdot14$	B1		
	$\tilde{\theta} = \frac{1}{1+0.14} = 0{\cdot}8772$	B1		
	Also, from expression given in question, $\mathrm{Var}(\tilde{\theta}) = \frac{0\cdot8772^2(1-0\cdot8772)}{100}$ $= 0{\cdot}000945$	B1		
	CI is given by $0{\cdot}8772 \pm 1{\cdot}96 \times \sqrt{0\cdot000945} =$ (0·817, 0·937)	M1 B1 M1 A1	For 0.8772 For 1.96 For $\sqrt{0\cdot000945}$	7

Q2				
(i)	Mgf of Z = E$(e^{tZ}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tz-\frac{z^2}{2}} dz$	M1		
	Complete the square	M1		
	$tz - \frac{z^2}{2} = -\frac{1}{2}(z-t)^2 + \frac{1}{2}t^2$	A1 A1		
	$= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}} dt = e^{\frac{t^2}{2}}$	M1 M1 M1	For taking out factor $e^{\frac{t^2}{2}}$ For use of pdf of N(t,1) For $\int$ pdf = 1	
	Pdf of N(t,1) $\therefore \int = 1$	A1	For final answer $e^{\frac{t^2}{2}}$	8
(ii)	Y has mgf $M_Y(t)$ Mgf of $aY + b$ is $E[e^{t(aY+b)}]$ $= e^{bt}E[e^{(at)Y}] = e^{bt}M_Y(at)$	M1 1 1 1	For factor e^{bt} For factor $E[e^{(at)Y}]$ For final answer	4
(iii)	$Z = \frac{X-\mu}{\sigma}$, so $X = \sigma Z + \mu$ $\therefore M_X(t) = e^{\mu t}.e^{\frac{(\sigma t)^2}{2}} = e^{\mu t + \frac{\sigma^2 t^2}{2}}$	M1 1 1 1	For factor $e^{\mu t}$ For factor $e^{\frac{(\sigma t)^2}{2}}$ For final answer	4
(iv)	$W = e^X$ $E(W^k) = E[(e^X)^k] = E(e^{kX}) = M_X(k)$	M1 A1 A1	For $E[(e^X)^k]$ For $E(e^{kX})$ For $M_X(k)$	
	$\therefore E(W) = M_X(1) = e^{\mu + \frac{\sigma^2}{2}}$	M1 A1		
	$E(W^2) = M_X(2) = e^{2\mu + 2\sigma^2}$ $\therefore \text{Var}(W) = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} [= e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)]$	M1 A1 A1		8

Q3			
(i)	$\bar{x} = 126\cdot 2$ s = 8·7002 $s^2 = 75\cdot 69\dot{3}$ $\bar{y} = 133\cdot 9$ s = 10·4760 $s^2 = 109\cdot 74\dot{6}$	A1	A1 if all correct. [No mark for use of s_n, which are 8·2537 and 9·6989 respectively.]
	$H_0 : \mu_A = \mu_B$ $H_0 : \mu_A \neq \mu_B$ Where μ_A, μ_B are the population means.	1 1	<u>Do not</u> accept $\overline{X} = \overline{Y}$ or similar.
	Pooled s^2 $= \dfrac{9 \times 75\cdot 69\dot{3} + 6 \times 109\cdot 47\dot{6}}{15} = \dfrac{681\cdot 24 + 658\cdot 48}{15}$ $= 89\cdot 314\dot{6}$ [√ = 9·4506]	B1	
	Test statistic is $\dfrac{126\cdot 2 - 133\cdot 9}{\sqrt{89\cdot 314\dot{6}}\sqrt{\dfrac{1}{10}+\dfrac{1}{7}}} = -\dfrac{7\cdot 7}{4\cdot 6573} = -1\cdot 653$	M1 A1	
	Refer to t_{15}	1	No FT if wrong
	Double-tailed 10% point is 1·753 Not significant No evidence that population mean concentrations differ.	1 1 1	No FT if wrong 10
(ii)	There may be consistent differences between days (days of week, types of rubbish, ambient conditions,…) which should be allowed for.	E1 E1	
	Assumption: Normality of population of <u>differences.</u>	1	
	Differences are 7·4 -1·2 11·1 5·5 6·2 3·7 −0·3 1·8 3·6 [$\bar{d}$ = 4·2, s = 3·862 (s^2 =14·915)] Use of $s_n (= 3\cdot 641)$ is <u>not</u> acceptable, even in a denominator of $s_n / \sqrt{n-1}$]	M1	A1 Can be awarded here if NOT awarded in part (i)
	Test statistic is $\dfrac{4\cdot 2 - 0}{3\cdot 862/\sqrt{9}}$ = 3·26	M1 A1	
	Refer to t_8 Double-tailed 5% point is 2·306 Significant Seems population means differ	1 1 1 1	No FT if wrong No FT if wrong 10

(iii)	Wilcoxon rank sum test Wilcoxon signed rank test H_0: $\text{median}_A = \text{median}_B$ H_1: $\text{median}_A \neq \text{median}_B$	B1 B1 1 1	[Or more formal statements]	4
Q4				
(i)	Description must be in <u>context</u>. If no context given, mark according to scheme and then give half-marks, rounded down. Clear description of "rows". And "columns" As extraneous factors to be taken account of in the design, with "treatments" to be compared. Need same numbers of each Clear contrast with situations for completely randomised design and randomised trends.	E1 E1 E1 E1 E1 E1 E1 E1 E1		9
(ii)	$e_{ij} \sim \text{ind N}(0, \sigma^2)$ α_i is population mean effect by which ith treatment differs from overall mean	1 1 1 1 1	Allow uncorrelated For 0 For σ^2	5
(iii)	see ANOVA table below	1 1 1 1 1 1 1		
	Refer to $F_{4,15}$	1	No FT if wrong	
	Upper 1% point is 4·89	1	No FT if wrong	
	Significant, seems treatments are not all the same	1		10

Source of Variation	SS	df	MS	MS ratio
Between Treatments	92·30	4	23·075	5·034
Residual	68·76	15	4·584	
Total	161·06	19		