4767 Statistics 2

Question 1

(i)	**EITHER:** $S_{xy} = \Sigma xy - \frac{1}{n}\Sigma x \Sigma y = 880.1 - \frac{1}{48} \times 781.3 \times 57.8$ $= -60.72$ $S_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 14055 - \frac{1}{48} \times 781.3^2 = 1337.7$ $S_{yy} = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 106.3 - \frac{1}{48} \times 57.8^2 = 36.70$ $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{-60.72}{\sqrt{1337.7 \times 36.70}} = -0.274$ **OR:** $\text{cov}(x,y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = 880.1/48 - 16.28 \times 1.204$ $= -1.265$ $\text{rmsd}(x) = \sqrt{\frac{S_{xx}}{n}} = \sqrt{(1337.7/48)} = \sqrt{27.87} = 5.279$ $\text{rmsd}(y) = \sqrt{\frac{S_{yy}}{n}} = \sqrt{(36.70/48)} = \sqrt{0.7646} = 0.8744$ $r = \frac{\text{cov(x,y)}}{rmsd(x)rmsd(y)} = \frac{-1.265}{5.279 \times 0.8744} = -0.274$	M1 for method for S_{xy} M1 for method for at least one of S_{xx} or S_{yy} A1 for at least one of S_{xy}, S_{xx}, S_{yy} correct M1 for structure of r A1 CAO (–0.27 to –0.28) M1 for method for cov (x,y) M1 for method for at least one msd A1 for at least one of cov/msd correct M1 for structure of r A1 CAO (–0.27 to –0.28)	5
(ii)	H_0: $\rho = 0$ H_1: $\rho < 0$ (one-tailed test) where ρ is the population correlation coefficient For $n = 48$, 5% critical value = 0.2403 Since $\|-0.274\| > 0.2403$ we can reject H_0: There is sufficient evidence at the 5% level to suggest that there is negative correlation between education spending and population growth.	B1 for H_0, H_1 in symbols B1 for defining ρ B1FT for critical value M1 for sensible comparison leading to a conclusion A1 for result (FT r<0) E1 FT for conclusion in words	6
(iii)	Underlying distribution must be bivariate Normal. If the distribution is bivariate Normal then the scatter diagram will have an elliptical shape.	B1 CAO for bivariate Normal B1 indep for elliptical shape	2
(iv)	• Correlation does not imply causation • There could be a third factor • increased growth could cause lower spending. Allow any sensible alternatives, including example of a possible third factor.	E1 E1 E1	3
(v)	Advantage – less effort or cost Disadvantage – the test is less sensitive (ie is less likely to detect any correlation which may exist)	E1 E1	2
			18

Question 2

(i)	(*A*) $P(X = 2) = e^{-0.37}\dfrac{0.37^2}{2!} = 0.0473$ (*B*) $P(X > 2)$ $= 1 - (e^{-0.37}\dfrac{0.37^2}{2!} + e^{-0.37}\dfrac{0.37^1}{1!} + e^{-0.37}\dfrac{0.37^0}{0!})$ $= 1 - (0.0473 + 0.2556 + 0.6907) = 0.0064$	M1 A1 (2 s.f.) M1 for $P(X = 1)$ and $P(X = 0)$ M1 for complete method A1 ***NB Answer given***	**5**
(ii)	P(At most one day more than 2) $= \binom{30}{1} \times 0.9936^{29} \times 0.0064 + 0.9936^{30} =$ $= 0.1594 + 0.8248 = 0.9842$	M1 for coefficient M1 for $0.9936^{29} \times 0.0064$ M1 for 0.993630 A1 CAO (min 2sf)	**4**
(iii)	$\lambda = 0.37 \times 10 = 3.7$ $P(X > 8) = 1 - 0.9863$ $= 0.0137$	B1 for mean (SOI) M1 for probability A1 CAO	**3**
(iv)	Mean no. per 1000ml = 200 × 0.37 = 74 Using Normal approx. to the Poisson, $X \sim N(74, 74)$ $P(X > 90) = P\left(Z > \dfrac{90.5 - 74}{\sqrt{74}}\right)$ $= P(Z > 1.918) = 1 - \Phi(1.918)$ $= 1 - 0.9724 = 0.0276$	B1 for Normal approx. with correct parameters (SOI) B1 for continuity corr. M1 for probability using correct tail A1 CAO (min 2 s.f.)**, (**but FT wrong or omitted CC)	**4**
(v)	P(questionable) = 0.0064 × 0.0137 × 0.0276 $= 2.42 \times 10^{-6}$	M1 A1 CAO	**2**
			18

Question 3

(i)	$X \sim N(27500, 4000^2)$ $P(X > 25000) = P\left(Z > \frac{25000 - 27500}{4000}\right)$ $= P(Z > -0.625)$ $= \Phi(0.625) = 0.7340$ (3 s.f.)	M1 for standardising A1 for -0.625 M1 *dep* for correct tail A1CAO (must include use of differences)	**4**
(ii)	P(7 of 10 last more than 25000) $= \binom{10}{7} \times 0.7340^7 \times 0.2660^3 = 0.2592$	M1 for coefficient M1 for $0.7340^7 \times 0.2660^3$ A1 FT (min 2sf)	**3**
(iii)	From tables $\Phi^{-1}(0.99) = 2.326$ $\frac{k - 27500}{4000} = -2.326$ $x = 27500 - 2.326 \times 4000 = 18200$	B1 for 2.326 seen M1 for equation in *k* and negative z-value A1 CAO for awrt 18200	**3**
(iv)	H_0: $\mu = 27500$; H_1: $\mu > 27500$ Where μ denotes the mean lifetime of the new tyres.	B1 for use of 27500 B1 for both correct B1 for definition of μ	**3**
(v)	Test statistic = $\frac{28630 - 27500}{4000/\sqrt{15}} = \frac{1130}{1032.8}$ = 1.094 5% level 1 tailed critical value of z = 1.645 1.094 < 1.645 so not significant. There is not sufficient evidence to reject H_0 There is insufficient evidence to conclude that the new tyres last longer.	M1 must include √ 15 A1 FT B1 for 1.645 M1 *dep* for a sensible comparison leading to a conclusion A1 for conclusion in words in context	**5**
			18

Question 4

(i)	H_0: no association between location and species. H_1: some association between location and species.	B1 for both	**1**
(ii)	Expected frequency = 38/160 × 42 = 9.975 Contribution = $(3 – 9.975)^2$ / 9.975 = 4.8773	M1 A1 M1 for valid attempt at $(O-E)^2/E$ A1 ***NB Answer given***	**4**
(iii)	Refer to χ_4^2 Critical value at 5% level = 9.488 Test statistic X^2 = 32.85 Result is significant There appears to be some association between location and species NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1or final E1	B1 for 4 deg of f(seen) B1 CAO for cv M1 Sensible comparison, using 32.85, leading to a conclusion A1 for correct conclusion (FT their c.v.) E1 conclusion in context	**5**
(iv)	• Limpets appear to be distributed as expected throughout all locations. • Mussels are much more frequent in exposed locations and much less in pools than expected. • Other shellfish are less frequent in exposed locations and more frequent in pools than expected.	E1 E1, E1 E1, E1	**5**
(v)	$\frac{24}{53} \times \frac{32}{65} \times \frac{16}{42} = 0.0849$	M1 for one fraction M1 for product of all 3 A1 CAO	**3**
			18