# 4767 Statistics 2

Question 1

**(i)**

| $x$ | 18 | 43 | 52 | 94 | 98 | 206 | 784 | 1530 |
|---|---|---|---|---|---|---|---|---|
| $y$ | 1.15 | 0.97 | 1.26 | 1.35 | 1.28 | 1.42 | 1.32 | 1.64 |
| Rank $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Rank $y$ | 2 | 1 | 3 | 6 | 4 | 7 | 5 | 8 |
| $d$ | -1 | 1 | 0 | -2 | 1 | -1 | 2 | 0 |
| $d^2$ | 1 | 1 | 0 | 4 | 1 | 1 | 4 | 0 |

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6\times 12}{8\times 63}$$

= 0.857 (to 3 s.f.) [ *allow* 0.86 to 2 s.f.]

M1 for attempt at ranking (allow all ranks reversed)

M1 for $d^2$

A1 for $\Sigma d^2 = 12$

M1 for method for $r_s$

A1 f.t. for $|r_s| < 1$

NB No ranking scores zero

**5**

**(ii)**

$H_0$: no association between $X$ and $Y$ in the population

$H_1$: some association between $X$ and $Y$ in the population

Two tail test critical value at 5% level is 0.7381

Since 0.857> 0. 7381, there is sufficient evidence to reject $H_0$,

i.e. conclude that the evidence suggests that there is association between population size $X$ and average walking speed $Y$.

B1 for $H_0$

B1 for $H_1$

B1 for population SOI

NB $H_0$ $H_1$ <u>not</u> ito $\rho$

B1 for $\pm 0.7381$

M1 for sensible comparison with c.v., provided $|r_s| < 1$

A1 for conclusion in words f.t. their $r_s$ and sensible cv

**6**

**(iii)**

$\bar{t} = 45$, $\bar{w} = 2.2367$

$$b = \frac{Stw}{Stt} = \frac{584.6 - 270\times 13.42/6}{13900 - 270^2/6} = \frac{-19.3}{1750} = -0.011$$

OR $b = \dfrac{584.6/6 - 45\times 2.2367}{13900/6 - 45^2} = \dfrac{-3.218}{291.6667} = -0.011$

hence least squares regression line is:

$$w - \bar{w} = b(t - \bar{t})$$

$$\Rightarrow w - 2.2367 = -0.011(t - 45)$$

$$\Rightarrow w = -0.011t + 2.73$$

B1 for $\bar{t}$ and $\bar{w}$ used (SOI)

M1 for attempt at gradient ($b$)

A1 CAO for -0.011

M1 for equation of line

A1 FT for complete equation

**5**

| | | | |
|---|---|---|---|
| **(iv)** | (*A*) For $t = 80$, predicted speed<br>$= -0.011 \times 80 + 2.73 = 1.85$<br>(*B*) The relationship relates to adults, but a ten year old will not be fully grown so may walk more slowly.<br>NB Allow E1 for comment about extrapolation not in context | M1<br>A1 FT provided b < 0<br>E1 extrapolation o.e.<br>E1 sensible contextual comment | **4** |
| | | **TOTAL** | **20** |

Question 2

| | | | |
|---|---|---|---|
| **(i)** | Binomial(5000,0.0001) | B1 for binomial<br>B1 dep, for parameters | **2** |
| **(ii)** | $n$ is large and $p$ is small<br>$\lambda = 5000 \times 0.0001 = 0.5$ | B1, B1<br>(Allow appropriate numerical ranges)<br>B1 | **3** |
| **(iii)** | $P(X \geq 1) = 1 - e^{-} \frac{0.5^0}{0!} = 1 - 0.6065 = 0.3935$<br>or from tables $= 1 - 0.6065 = 0.3935$ | M1 for correct calculation or correct use of tables<br>A1 | **2** |
| **(iv)** | P(9 of 20 contain at least one)<br>$= \binom{20}{9} \times 0.3935^9 \times 0.6065^{11}$<br>$= 0.1552$ | M1 for coefficient<br>M1 for $p^9 \times (1-p)^{11}$, $p$ from part (iii)<br>A1 | **3** |
| **(v)** | Expected number $= 20 \times 0.3935 = 7.87$ | M1 A1 FT | **2** |
| **(vi)** | Mean $= \frac{\Sigma xf}{n} = \frac{7+4}{20} = \frac{11}{20} = 0.55$<br>Variance $= \frac{1}{n-1}\left(\Sigma fx^2 - n\bar{x}^2\right)$<br>$= \frac{1}{19}\left(15 - 20 \times 0.55^2\right) = 0.471$ | B1 for mean<br>M1 for calculation<br>A1 CAO | **3** |
| **(vii)** | Yes, since the mean is close to the variance,<br>and also as the expected frequency for 'at least one', i.e. 7.87,<br>is close to the observed frequency of 9. | B1<br>E1 for sensible comparison<br>B1 for observed frequency $= 7 + 2 = 9$ | **3** |
| | | **TOTAL** | **18** |

Question 3

| | | | |
|---|---|---|---|
| **(i)** | (*A*) P( $X < 120$) = $\mathrm{P}\left( Z < \frac{120-115.3}{21.9} \right)$<br>= P( $Z < 0.2146$)<br>= $\Phi(0.2146) = 0.5849$ | M1 for standardizing<br>A1 for $z = 0.2146$<br>A1 CAO (min 3 sf, to include use of difference column) | **3** |
| | (*B*) P($100 < X < 110$) = $\mathrm{P}\left( \frac{100-115.3}{21.9} < Z < \frac{110-115.3}{21.9} \right)$<br>= P($-0.6986 < Z < -0.2420$)<br>= $\Phi(0.6986) - \Phi(0.2420)$<br>= $0.7577 - 0.5956$<br>= 0.1621 | M1 for standardizing both 100 & 110<br>M1 for correct structure in calc$^{n}$<br>A1 CAO | **3** |
| | (*C*) From tables $\Phi^{-1}(0.1) = -1.282$<br>$\frac{k-115.3}{21.9} = -1.282$<br>$k = 115.3 - 1.282 \times 21.9 = 87.22$ | B1 for $\pm 1.282$ seen<br>M1 for equation in $k$ and negative z-value<br>A1 CAO | **3** |
| **(ii)** | From tables,<br>$\Phi^{-1}(0.70) = 0.5244$, $\Phi^{-1}(0.15) = -1.036$<br>$180 = \mu + 0.5244\,\sigma$<br>$140 = \mu - 1.036\,\sigma$<br>$40 = 1.5604\,\sigma$<br>$\sigma = 25.63$, $\mu = 166.55$ | B1 for 0.5244 or $\pm 1.036$ seen<br>M1 for at least one equation in $\mu$ and $\sigma$ and $\Phi^{-1}$ value<br>M1 dep for attempt to solve two equations<br>A1 CAO for both | **4** |
| **(iii)** | $\Phi^{-1}(0.975) = 1.96$<br>$a = 166.55 - 1.96 \times 25.63 = 116.3$<br>$b = 166.55 + 1.96 \times 25.63 = 216.8$ | B1 for $\pm 1.96$ seen<br>M1 for either equation<br>A1<br>A1<br>[Allow other correct intervals] | **4** |
| | | **TOTAL** | **17** |

Question 4

**(i)**

$H_0$: no association between growth and type of plant;
$H_1$: some association between growth and type of plant;

| EXPECTED | Good | Average | Poor |
|---|---|---|---|
| Coriander | 12.10 | 24.93 | 17.97 |
| Aster | 10.56 | 21.76 | 15.68 |
| Fennel | 10.34 | 21.31 | 15.35 |

| CONTRIBUTION | Good | Average | Poor |
|---|---|---|---|
| Coriander | 0.0008 | 0.3772 | 0.4899 |
| Aster | 1.2002 | 0.6497 | 3.4172 |
| Fennel | 1.2955 | 0.0226 | 1.2344 |

$X^2 = 8.69$

Refer to $\chi^2_4$

Critical value at 5% level = 9.488

Result is not significant
There is not enough evidence to suggest that there is some association between reported growth and type of plant;
NB if $H_0$ $H_1$ reversed, or 'correlation' mentioned, do not award first B1or final A1

B1 (in context)

M1 A2 for expected values (to 2 dp)
(allow A1 for at least one row or column correct)

M1 for valid attempt at $(O-E)^2/E$
A1 for all correct
NB These M1A1 marks cannot be implied by a correct final value of $X^2$

M1 for summation
A1 for $X^2$ CAO

B1 for 4 d.o.f.

B1 CAO for cv

M1
A1

**12**

**(ii)**

Test statistic = $\dfrac{49.2-47}{8.5/\sqrt{50}} = \dfrac{2.2}{1.202} = 1.830$

1% level 1 tailed critical value of z = 2.326

1.830 < 2.326 so not significant.
There is not sufficient evidence to reject $H_0$

There is insufficient evidence to conclude that the flowers are larger.

M1 correct denominator
A1

B1 for 2.326
M1 (dep on first M1) for sensible comparison leading to a conclusion

A1 for fully correct conclusion in words in context

**5**

**TOTAL** **17**