Topic Y3 Correlation regression & chi-squared tests (Post-TT A) [48] MS

1.

(i)		$S_{xx} = 8700000 - \frac{7000^2}{6}$ (= 533333) $S_{xy} = 509900 - \frac{7000\times456}{6}$ (= –22100) $b = -\frac{"22100"}{"533333"}$ or $-\frac{663}{16000}$ (= –0.0414) $y - \frac{456}{6} = "-0.0414"(x - \frac{7000}{6})$ $y = -0.0414x + 124$ (3 sf)	M1 M1 M1 A1 [4]	Correct subst'n in any correct S formula Correct subst'n in any correct b formula from two correct S formulae ft their b except if using r or $y = -\frac{663}{16000}x + \frac{3979}{32}$ or $y = -0.041x + 124$	or $a = \frac{456}{6} - ("-0.0414")\times\frac{7000}{6}$ oe ft "b" Allow $y = -0.04x + 124$ if –0.041… seen above
(ii)		70 to 72	B1 [1]	or 71 per thousand, NOT 71000	No ft from (i) Ignore method
(iii)		Extrapolation oe Corr'n not high or small sample	B1 B1 [2]	Allow "2400 is beyond graph" } "Not shown on the graph" or } 1st B1 only "Line drops low, or below 0" } "Outlier" } Poor corr'n oe, or pts not close to line oe 2nd B1	"Line only allows for countries poorer than Nigeria" 1st B1 Allow "Value for Nigeria is –ve 1st B1 NOT "Other factors may apply" oe Ignore all else
(iv)		$S_{xx} = 8700000 + 1300^2 - \frac{(7000+1300)^2}{7}$ $S_{yy} = 36262 + 96^2 - \frac{(456+96)^2}{7}$ $S_{xy} = 509900 + 1300\times96 - \frac{8300\times552}{7}$ $r = \frac{"-19814.3"}{\sqrt{"548571"\times"1948.86"}}$ = –0.606 (3 sf)	M1 A1 M1 A1 [4]	or $10390000 - \frac{(8300)^2}{7} = \frac{3840000}{7}$ or 548571 or $45478 - \frac{552^2}{7} = \frac{13642}{7}$ or 1948.86 or $634700 - \frac{8300\times552}{7} = -\frac{138700}{7}$ or –19814.3 Correct subst'n in any correct r formula from 3 correct subs in 3 correct S formulae, ie all correct method	Correct sub in any correct S formula M1 Correct value of any S seen or implied by r A1 SC If $n = 6$, but otherwise correct allow M1A0M1A0 (ans $r = -0.574$, must see wking)
(v)		No effect oe	B1 [1]	Stay the same oe Allow just "No"	Ignore all else

2.

(i)		17.5 4.2 6.3 32.5 7.8 11.7 oe	M1 A1 A1 [3]	eg 50 × 28 ÷ 80 At least 2 correct. All correct.
(ii)		The E value of 4.2 < 5 Combine Biology and Chemistry (both sciences).	B1 B1 [2]	Need not mention 4.2 May need to look at (iii) to see which subjects combined.
(iii)		Ho: Subject and sex are independent H_1: They are not independent 21.7 6.3 40.3 11.7 $\chi^2 = (4.7 - 0.5)^2(21.7^{-1} + 6.3^{-1} + 40.3^{-1} + 11.7^{-1})$ = 5.558… ($v = 1$) (α) 2½ % CV = 5.024 5.558 > CV or in CR and reject H_0	B1 B1 M1M1 A1 B1 M1	oe. NOT 'variables', 'they' etc or 17.5 10.5 32.5 19.5 if C/A combined. No Yates(inc $v > 1$) or incorrect Yates (eg no modulus) M1M0. allow 6.96 or 6.79 Chem./Art combined B1B1M1M1A0B1M1A0. (TS = 3.75) ft TS & CV. Correct first conclusion. If C/A prob. accept H_0.
		(β) $P(\chi^2_1 \geq 5.558) = 0.0184$ < 0.025 and reject H_0	B1 M1	
		There is significant evidence that subject and sex are not independent	A1 [8]	cwo. NOT over-assertive. Thus no or incorrect Yates can score max 6/8 B1B1M1M0A1B1M1A0.

3.

(i)	1 2 3 4 5 or 5 4 3 2 1 3 5 4 1 2 3 1 2 5 3 Σd^2 (= 32) $1 - {}^{6\times"32"}/_{5(25-1)}$ = - 0.6	M1 A1 M1dep M1dep A1 5	attempt ranks correct ranks S_{xx} or $S_{yy} = 55 - 15^2/5 (=10)$ or $S_{yy} = 39 - 15^2/5 (= -6)$ $^{-6}/\sqrt{(10\times10)}$
(ii)	1 & 3 Largest neg r_s or large neg r_s or strong neg corr'n or close(st) to -1 or lowest r_s	B1ind B1dep 2	ft if -1 < (i) < -0.9, ans 1 & 2 NOT: furthest from 0 or closest to ±1 little corr'n most disagreement

4.

<table>
<tr><td>4 (a)</td><td>$\mathbf{H_0}$: Geo(1/3) is a suitable model for the number of rolls to 1st 5 or 6
$\mathbf{H_1}$: Geo(1/3) is not a suitable model for the number of rolls to 1st 5 or 6</td><td>B1</td><td>2.5</td></tr>
<tr><td></td><td>Assuming $\mathbf{H_0}$ is true , expected values are $\mathbf{E_i = 100 \times \frac{1}{3} \times \left(\frac{2}{3}\right)^{i-1}}$</td><td>M1</td><td>3.4</td></tr>
<tr><td></td><td><table><tr><td>Number of rolls to 1st 5 or 6</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7 or more</td></tr><tr><td>Expected frequency</td><td>33.3…</td><td>22.2…</td><td>14.8…</td><td>9.87…</td><td>6.58…</td><td>4.38…</td><td>8.77….</td></tr></table></td><td>A1
A1</td><td>1.1b
1.1b</td></tr>
<tr><td></td><td><table><tr><td>Number of rolls to 1st 5 or 6</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6 or more</td></tr><tr><td>Observed frequency</td><td>39</td><td>17</td><td>17</td><td>10</td><td>3</td><td>14</td></tr><tr><td>Expected frequency</td><td>33.3…</td><td>22.2…</td><td>14.8…</td><td>9.87…</td><td>6.58…</td><td>13.1…</td></tr></table>Combining last two groups</td><td>M1</td><td>2.1</td></tr>
<tr><td></td><td>$\mathbf{\nu = 6 - 1 = 5}$</td><td>B1ft</td><td>1.1b</td></tr>
<tr><td></td><td>Critical value, $\mathbf{\chi^2(0.05) = 11.070}$</td><td>B1ft</td><td>1.1a</td></tr>
<tr><td></td><td>Test statistic = $\mathbf{\frac{(39-33.3)^2}{33.3} + \frac{(17-22.2)^2}{22.2} + \frac{(17-14.8)^2}{14.8} + \cdots}$</td><td>M1</td><td>1.1b</td></tr>
<tr><td></td><td>= 4.5181… awrt 4.5</td><td>A1</td><td>1.1b</td></tr>
<tr><td></td><td>Not in critical region, insufficient evidence to reject $\mathbf{H_0}$, no significant evidence at the 5% level that the die is biased.</td><td>A1cso</td><td>3.5a</td></tr>
<tr><td></td><td></td><td></td><td>(10)</td></tr>
<tr><td>(b)(i)</td><td>e.g. Mai’s test measures results against the frequencies of all six possible scores, whereas Desmond’s doesn’t.</td><td>B1</td><td>3.5b</td></tr>
<tr><td></td><td>Mai’s test needs a fixed number of rolls, whereas number of rolls for Desmond’s test is unknown.</td><td>B1</td><td>3.5b</td></tr>
<tr><td>(ii)</td><td>Desmond’s test is likely to collect more data</td><td>B1</td><td>3.5b</td></tr>
</table>

5.

Question	Scheme	Marks	AOs
1 (a)	H_0; recycling plastics and the distance to the nearest recycling point are independent H_1; recycling plastics and the distance to the nearest recycling point are not independent	B1	1.2
	Degrees of freedom, $\nu = (3-1)\times(2-1) = 2$ Therefore critical value = $\chi^2(0.05) = 5.991$	B1	3.1b
	Test statistic $= \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(64-58.3)^2}{58.3} + \frac{(42-47.7)^2}{47.7} + \frac{(32-29.7)^2}{29.7} + \cdots =$	M1	1.1b
	$8.0988\ldots\ldots =$ awrt 8.10	A1	1.1b
	In critical region, therefore sufficient evidence to reject H_0. Data does **not** support Barbara's belief at the 5% significance level that people recycling plastics is independent of the distance to the nearest recycling point.	A1	3.2a
			(5)
(b)	• Test statistic is halved and • Critical value stays the same	B1	2.4
	This is not in the critical region and therefore Barbara's original conclusions are not valid, data suggests independence.	dB1	2.3
			(2)
			(7)