Paper 3: Further Statistics 1 Mark Schemes

<table>
<tr><th>Question</th><th colspan="2">Scheme</th><th>Marks</th><th>AOs</th></tr>
<tr><td rowspan="6">Q1</td><td colspan="2">$H_0 : \lambda = 5$ ($\lambda = 2.5$) $H_1 : \lambda > 5$ ($\lambda > 2.5$)</td><td>B1</td><td>2.5</td></tr>
<tr><td colspan="2">$X \sim \text{Po}(2.5)$</td><td>B1</td><td>3.3</td></tr>
<tr><td>Method 1:</td><td>Method 2:</td><td></td><td></td></tr>
<tr><td>$\text{P}(X \geqslant 7) = 1 - \text{P}(X \leqslant 6)$
$= 1 - 0.9858$</td><td>$\text{P}(X \geqslant 5) = 0.1088$
$\text{P}(X \geqslant 6) = 0.042$</td><td>M1</td><td>1.1b</td></tr>
<tr><td>$= 0.0142$</td><td>CR $X \geqslant 6$</td><td>A1</td><td>1.1b</td></tr>
<tr><td colspan="2">$0.0142 < 0.05$ $7 \geqslant 6$ or 7 is in critical region or 7 is significant
Reject H_0. There is evidence at the 5% significance level that the level of pollution has increased.
or
There is evidence to support the scientists claim is justified</td><td>A1cso</td><td>2.2b</td></tr>
<tr><td colspan="5" align="right">(5 marks)</td></tr>
</table>

Notes:

B1: Both hypotheses correct using λ **or** μ **and** 5 **or** 2.5

B1: Realising that the model Po(2.5) is to be used. This may be stated or used

M1: Using or writing $1 - \text{P}(X \leqslant 6)$ **or** $1 - \text{P}(X < 7)$

a correct CR **or** $\text{P}(X \geqslant 5)$ = awrt 0.109 **and** $\text{P}(X \geqslant 6)$ = awrt 0.042

A1: awrt 0.0142 **or** CR $X \geqslant 6$ **or** $X > 5$

M1: A fully correct solution and drawing a correct inference in context

Question	Scheme	Marks	AOs
Q2(a)	$\mathrm{P}(X \geqslant 1) = 1 - \mathrm{P}(X = 0)$ $1 - \mathrm{P}(X = 0) = 0.049$	B1	3.1b
	$\mathrm{P}(X = 0) = 0.951$	B1	1.1b
	$x^5 = 0.951$ $x = 0.99$	M1	3.1b
	$p = 0.01$	A1	1.1b
	$X \sim \mathrm{B}(1000, 0.01)$	M1	3.3
	Mean $= np = 10$	A1ft	1.1b
	Variance $= np(1-p) = 9.9$	A1ft	1.1b
		(7)	
(b)	$X \sim \mathrm{Po}(\text{“}10\text{”})$ then require: $\mathrm{P}(X > 6) = 1 - \mathrm{P}(X \leqslant 6)$	M1	3.4
	$= 1 - 0.1301$		
	$= 0.870$	A1	1.1b
		(2)	
(c)	The approximation is valid as : the number of calls is large	B1	2.4
	The probability of connecting to the wrong agent is small	B1	2.4
		(2)	
(d)	The answer is accurate to 2 decimal place	**B1**	3.2b
		(1)	
			(12 marks)

Notes:

(a)

B1: Realising that the P(at least 1 call) = $1 - \mathrm{P}(X = 0)$

B1: Calculating $\mathrm{P}(X = 0) = 0.951$

M1: Forming the equation $x^5 = \text{"their 0.951"}$ may be implied by $p = 0.01$

A1: 0.01 only

M1: Realising the need to use the model B(1000, 0.01) This may be stated or used

A1: Mean =10 or ft their p but only if $0 < p < 1$

A1: Var = 9.9 or ft their p but only if $0 < p < 1$

(b)

M1: Using the model Po(“their 10”) (this may be written or used) and $1 - \mathrm{P}(X \leqslant 6)$

A1: awrt 0.870 Award M1 A1 for awrt 0.870 with no incorrect working

Question 2 notes continued
(c) **B1:** Explaining why approximation is valid - need the context of number and calls **B1:** Need the context connecting, wrong agent
(d) **B1:** Evaluating the accuracy of their answer in (b). Allow 2 significant figures

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AOs</th></tr>
<tr><td>Q3(a)</td><td>Expected value for 2 = $150 \times \mathrm{P}(X = 2)$</td><td>M1</td><td>3.4</td></tr>
<tr><td></td><td>= 28.3015…</td><td>A1</td><td>1.1b</td></tr>
<tr><td></td><td>Expected value for 4 or more = 150 – (53.8 + 56.6 + 28.3 + 8.9)
= 2.4</td><td>A1ft</td><td>1.1b</td></tr>
<tr><td></td><td>H_0: Bin(20, 0.05) is a suitable model
H_1: Bin(20, 0.05) is not a suitable model</td><td>B1</td><td>2.5</td></tr>
<tr><td></td><td>Combining last two groups
<table>
<tr><th></th><th>⩾ 3</th></tr>
<tr><td>Observed frequency</td><td>19</td></tr>
<tr><td>Expected frequency</td><td>11.3</td></tr>
</table></td><td>M1</td><td>2.1</td></tr>
<tr><td></td><td>$\nu = 4 - 1 = 3$</td><td>B1</td><td>1.1b</td></tr>
<tr><td></td><td>Critical value, χ^2 (0.05) = 7.815</td><td>B1</td><td>1.1a</td></tr>
<tr><td></td><td>Test statistic = $\frac{(43-53.8)^2}{53.8} + \frac{(62-56.6)^2}{56.6} + \ldots$</td><td>M1</td><td>1.1b</td></tr>
<tr><td></td><td>= 8.117</td><td>A1</td><td>1.1b</td></tr>
<tr><td></td><td>In critical region, sufficient evidence to reject H_0, accept H_1
Significant evidence at 5% level to reject the manager’s model</td><td>A1</td><td>3.5a</td></tr>
<tr><td></td><td></td><td>(10)</td><td></td></tr>
<tr><td>(b)</td><td>$\nu = 4 - 2 = 2$</td><td></td><td></td></tr>
<tr><td></td><td>4 classes due to pooling</td><td>B1</td><td>2.4</td></tr>
<tr><td></td><td>2 restrictions (equal total and mean/proportion)</td><td>B1</td><td>2.4</td></tr>
<tr><td></td><td></td><td>(2)</td><td></td></tr>
<tr><td>(c)</td><td>H_0: Binomial distribution is a good model
H_1: Binomial distribution is not a good model</td><td>B1</td><td>3.4</td></tr>
<tr><td></td><td>Critical value, χ^2 (0.05) = 5.991
Test statistic is not in critical region, insufficient evidence to reject H_0
There is evidence that the Binomial distribution is a good model</td><td>B1</td><td>3.5a</td></tr>
<tr><td></td><td></td><td>(2)</td><td></td></tr>
<tr><td colspan="4" align="right">(14 marks)</td></tr>
</table>

Notes:

(a)

M1:	Using the binomial model $150 \times p^2 \times (1-p)^{18}$ may be implied by 28.3
A1:	awrt 28.3
A1:	awrt 2.4 or ft their “28.3”
B1:	Both hypotheses correct using the correct notation or written out in full
M1:	For recognising the need to combine groups
B1:	Number of degrees of freedom = 3 may be implied by a correct CV
B1:	awrt 7.82
M1:	Attempting to find $\sum \frac{(O_i - E_i)^2}{E_i}$ or $\sum \frac{O_i^2}{E_i} - N$ may be implied by awrt 8.12
A1:	awrt 8.12
A1:	Evaluating the outcome of a model by drawing a correct inference in context

(b)

B1:	Explaining why there are 4 classes
B1:	Explanation of why 2 is subtracted

(c)

B1:	Correct hypotheses for the refined model
B1:	The CV awrt 5.99 and drawing the correct inference for the refined model

Question	Scheme	Marks	AOs
Q4.	Po(2.3) $n = 100$ $\mu = 2.3$ $\sigma^2 = 2.3$		
	$\text{CLT} \Rightarrow \bar{X} \approx \text{N}\left(2.3, \frac{2.3}{100}\right)$	M1 A1	3.1a 1.1b
	$\text{P}(\bar{X} > 2.5) = \text{P}\left(Z > \frac{2.5 - 2.3}{\sqrt{0.023}}\right)$	M1	3.4
	$= \text{P}(Z > 1.318..)$		
	$= 0.09632\ldots$	A1	1.1b
		(4)	
		(4 marks)	

M1: For realising the need to use the CLT to set $\bar{X} \approx$ normal with correct mean
May be implied by using the correct normal distribution

A1: For fully correct normal stated or used

M1: Use of the normal model to find $\text{P}(\bar{X} > 2.5)$. Can be awarded for $\frac{2.5-2.3}{\sqrt{0.023}}$

or awrt 1.32

A1: awrt 0.0963

Question	Scheme	Marks	AOs
Q5(a)	$\binom{7}{1} \times 0.15^2 \times (0.85)^6$	M1	3.3
	$= 0.05940\ldots$ = awrt **<u>0.0594</u>**	A1	1.1b
		(2)	
(b)	The model is only valid if:		
	the games (trials) are **independent**	B1	3.5b
	the probability of winning a prize, 0.15, is **constant** for each game	B1	3.5b
		(2)	
(c)	$18 = \frac{r}{p}$ and $6^2 = \frac{r(1-p)}{p^2}$	M1 A1	3.1b 1.1b
	Solving: $2p = 1 - p$	M1	1.1b
	$p = \frac{1}{3}$ (> 0.15) so Mary has the greater chance of winning a prize	A1	3.2a
		(4)	
	(8 marks)		

Notes:

5(a)

M1: For selecting an appropriate model negative binomial or B(7, 0.15) with an extra success in 8th trial e.g.

$\binom{7}{1} 0.15 \times (0.85)^6 \times 0.15$ Allow $\binom{7}{1} 0.85 \times (0.15)^6 \times 0.85$ may be implied by awrt 0.0594

A1: awrt 0.0594

(b)

B1: Stating the first assumption that games are independent

B1: Stating the second assumption that the probability remains constant

(c)

M1: Forming an equation for the mean or for the standard deviation

A1: Both equations correct

M1: Solving the 2 equations leading to $2p = 1 - p$

A1: For $p = \frac{1}{3}$ followed by a correct deduction

Question	Scheme	Marks	AOs
Q6(a)	$G_X(1)=1$ gives	M1	2.1
	$k\times 6^2=1$ so $k=\frac{1}{36}$ *	A1*cso	1.1b
		(2)	
(b)	P(X = 3) = coefficient of t^3 so $G_X(t)=k\left(...+4t^3...\right)$	M1	1.1b
	[P(X = 3) =] $\underline{\frac{1}{9}}$	A1	1.1b
		(2)	
(c)	$G'_X(t)=2k\left(3+t+2t^2\right)\times\left(1+4t\right)$	M1	2.1
	E(X) = $G'_X(1)=2k\left(3+1+2\right)\times\left(1+4\right)$	M1	1.1b
	$=\frac{5}{3}$	A1	1.1b
	$G''_X(t)=2k\left[\left(3+t+2t^2\right)\times4+\left(1+4t\right)^2\right]$	M1 A1	2.1 1.1b
	$G''_X(1)=2k[6\times4+5^2]$ $\left\{=\frac{49}{18}\right\}$	M1	1.1b
	Var(X) = $G''_X(1)+G'_X(1)-\left[G'_X(1)\right]^2=\frac{49}{18}+\frac{5}{3}-\frac{25}{9}$	M1	2.1
	$=\underline{\frac{29}{18}}$ *	A1*cso	1.1b
		(8)	
(d)	$G_{2X+1}(t)=\frac{t}{36}\left(3+t^2+2\left(t^2\right)^2\right)^2$ [$\times t$ or sub t^2 for t]	M1	3.1a
	$=G_{2X+1}(t)=\frac{t}{36}\left(3+t^2+2t^4\right)^2$	A1	1.1b
		(2)	
		(14 marks)	

Notes:

(a)

M1: Stating $G_X(1)=1$

A1*: Fully correct proof with no errors cso

(b)

M1: Attempting to find the coefficient of t^3. May be implied by obtaining $\frac{1}{9}$ or awrt 0.11

A1: $\frac{1}{9}$, allow awrt 0.111

Question 6 notes continued:

(c)

M1: Attempting to find $G_X'(t)$. Allow Chain rule or multiplying out the brackets and differentiating

M1: Substituting $t = 1$ into $G_X'(t)$

A1: $\frac{5}{3}$, allow awrt 1.67

M1: Attempting to find $G_X''(t)$

A1: $2k\left[\left(3+t+2t^2\right)\times 4+\left(1+4t\right)^2\right]$ **or** $k(48t^2+24t+26)$ o.e.

A1: $2k[6\times 4+5^2]$ o.e.

M1: Using $G_X''(1)+G_X'(1)-\left[G_X'(1)\right]^2$ to find the Variance

A1*: $\frac{29}{18}$ cso

(d)

M1: Realising the need to $\times t$ **or** sub t^2 for t

A1: $\frac{t}{36}\left(3+t^2+2t^4\right)^2$, or $\frac{t}{36}\left(9+6t^2+13t^4+4t^6+4t^8\right)$ o.e.

Question	Scheme	Marks	AOs
Q7(a)	$X \sim \mathrm{B}(20, 0.2)$ and seek c such that $\mathrm{P}(X \leqslant c) < 0.10$	M1	3.3
	[$\mathrm{P}(X \leqslant 1) = 0.0692$] CR is $X \leqslant 1$	A1	1.1b
		(2)	
(b)	Size = **0.0692**	B1ft	1.2
		(1)	
(c)	Y = no. of spins until red obtained so $Y \sim \mathrm{Geo}(0.2)$	M1	3.3
	$\mu = \frac{1}{p}$ so if $p < 0.2$ then mean is larger so seek d so that $\mathrm{P}(Y \geqslant d) < 0.10$	M1	2.4
	$\mathrm{P}(Y \geqslant d) = (0.8)^{d-1}$	M1	3.4
	$(0.8)^{d-1} < 0.10 \Rightarrow d - 1 > \frac{\log(0.1)}{\log(0.8)}$	M1	1.1b
	$d > 11.3..$	A1	1.1b
	CR is $\boldsymbol{Y \geqslant 12}$	A1	2.2b
		(6)	
(d)	Size = [$0.8^{11} = 0.085899\ldots$] = **0.0859**	B1	1.1b
		(1)	
(e)(i)	Power = P(reject H0 when it is false) = $\mathrm{P}(X \leqslant 1 \mid X \sim \mathrm{B}(20, p))$	M1	2.1
	$= (1-p)^{20} + 20(1-p)^{19}p$	M1	1.1b
	$= (1-p)^{19}(1+19p)$ *	A1*cso	1.1b
(ii)	Power = $(1-p)^{11}$	B1	1.1b
		(4)	
(f)	Sam's test has smaller P(Type I error) (or size) so is better	B1	2.2a
	Power of Sam's test = 0.1755…	B1	1.1b
	Power of Tessa's test = 0.85^{11} = 0.1673…	B1	1.1b
	So for $p = 0.15$ **Sam's test** is recommended	B1	2.2b
		(4)	
		(18 marks)	

Notes:

(a)

M1: Realising the need to use the model Using B(20,0.2) with method for finding the CR or implied by a correct CR

A1: $X \leqslant 1$ or $X < 2$

(b)

B1: awrt 0.0692

(c)

M1: Realising that the model Geo(0.2)is needed. This may be written or used

M1: Realising the key step that they need to find $P(Y \geqslant d) < 0.10$

M1: Using the model $(0.8)^{d-1}$

M1: Using the model $(0.8)^{d-1} < 0.10$ and finding a method to solve leading to a value/range of values for d

A1: For $d > 11.3..$

A1: For $Y \geqslant 12$ or $Y > 11$ (a correct inference)

(d)

B1ft: awrt 0.0692. ft their answer to part (c)

(e)(i)

M1: Using B(20, p) and realizing they need to find $P(X \leqslant 1)$ o.e. This may be used or written

M1: Using $P(X = 0) + P(X = 1)$

A1*: Fully correct proof (no errors) cso

(ii)

B1: For $(1 - p)^{11}$

(f)

B1: Making a deduction about the tests using the answers to part(b) and (d)

B1: awrt 0.0176

B1: awrt 0.167

B1: A correct inference about which test is recommended