2018 June - Section A: STATISTICS

Qu 1	Scheme	Marks	AO
(a)	c: 0 1 2 3 4 5 6 7 8 P($C = c$): $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$ $\frac{1}{9}$	B1 B1ft (2)	1.2 1.2
(b)	P($C < 4$) = $\frac{4}{9}$ (accept 0.444 or better)	B1 (1)	3.4
(c)	Probability lower than expected suggests model is <u>not</u> good	B1ft (1)	3.5a
(d)	e.g. Cloud cover will vary from month to month and place to place So e.g. use a non-uniform distribution	B1 (1)	3.5c
		(5 marks)	

Notes

(a) 1st B1 for a correct set of values for c. Allow $\left\{\frac{1}{8}, \frac{2}{8}, \ldots \frac{8}{8}\right\}$

2nd B1ft for correct probs from their values for c, consistent with discrete uniform distrib'n
Maybe as a prob. function. Allow P($X = x$) = $\frac{1}{9}$ for $0 \leqslant x \leqslant 8$ provided $x = \{0, 1, 2, \ldots, 8\}$ is clearly defined somewhere.

(b) B1 for using correct model to get $\frac{4}{9}$ (o.e.)

SC **Sample space {1, …, 8}** If scored B0B1 in (a) for this allow P($C < 4$) = $\frac{3}{8}$ to score B1 in (b)

(c) B1ft for comment that states that the model proposed is or is not a good one based on their model in part (a) and their probability in (b)
|(b) – 0.315| > 0.05 Allow e.g. "it is not suitable"; "it is not accurate" etc
|(b) – 0.315| ⩽ 0.05 Allow a comment that suggests it <u>is</u> suitable
No prob in (b) Allow a comparison that mentions 50% or 0.5 and rejects the model
No prob in (b) and no 50% or 0.5 or **(b) > 1** scores B0
Ignore any comments about location or weather patterns.

(d) B1 for a sensible refinement considering variations in month or location
Just saying "not uniform" is B0
Context & "non-uniform" Allow mention of different locations, months <u>and</u> non-uniform <u>or</u> use more locations to form a new distribution with probabilities based on frequencies
Context & "binomial" Allow mention of different locations, months <u>and</u> binomial
Just refined model Model must be outlined and discrete and non-uniform
e.g. higher probabilities for more cloud cover <u>or</u> lower probabilities for less cloud cover
Continuous model Any model that is based on a continuous distribution. e.g. normal is B0

Qu 2	Scheme	Marks	AO
(a)	$H_0 : \rho = 0$ $H_1 : \rho < 0$	B1	2.5
	Critical value: -0.6215 (Allow any cv in range $0.5 < \|cv\| < 0.75$)	M1	1.1a
	$r < -0.6215$ so significant result and there is evidence of a negative correlation between *w* and *t*	A1	2.2b
		(3)	
(b)	e.g. As temperature increases people spend more time on the beach and less time shopping (o.e.)	B1	2.4
		(1)	
(c)	Since *r* is close to -1 , it **is** consistent with the suggestion	B1	2.4
		(1)	
(d)	*t* will be the explanatory variable since sales are likely to depend on the temperature	B1	2.4
		(1)	
(e)	Every degree rise in temperature leads to a drop in weekly earnings of £171	B1	3.4
		(1)	
		(7 marks)	

	Notes
(a)	B1 for both hypotheses in terms of ρ M1 for the critical value: sight of $\pm$ 0.6215 or any cv such that $0.5 < \|cv\| < 0.75$ A1 must reject H_0 on basis of comparing -0.915 with -0.6215 (if $-0.915 < 0.6215$ is seen then A0 but may use $\|r\|$ o.e. which is fine) and mention "negative", "correlation/relationship" and at least "*w*" and "*t*"
(b)	B1 for a suitable reason to explain negative correlation using the context given. e.g. "As temperature drops people are more likely to go shopping (than to the beach)" e.g. "As temperature increases people will be outside rather than in shops" A mere description in context of negative correlation is B0 SO e.g. "As temperature increases people don't want to go shopping/buy clothes" is B0 e.g. "Less clothes needed as temp increases" is B0
(c)	B1 for a suitable reason e.g. "strong"/"significant"/"near perfect" "correlation", $\|r\|$ close to 1 and saying it is consistent with the suggestion. Allow "yes" followed by the reason.
(d)	B1 For identifying *t* and giving a suitable reason. Need idea that "*w* depends on *t*" or "*w* responds to *t*" or "*t* affects *w*" (o.e.) Allow *t* (temperature) affects the other variable etc Just saying " *t* is the independent variable" or "*t* explains change in *w*" is B0 N. B. Suggesting causation is B0 e.g. "*t* causes *w* to decrease"
(e)	B1 for a description that conveys the idea of rate per degree Celsius. Must have 171, condone missing "£" sign.

Qu 3	Scheme	Marks	AO
(a)	The <u>probability</u> of a dart hitting the target is <u>constant</u> (from child to child and for each throw by each child) (o.e.)	B1	1.2
	The <u>throws</u> of each of the darts are <u>independent</u> (o.e.)	B1 (2)	1.2
(b)	[P($H \geqslant 4$) = 1 – P($H \leqslant 3$) = 1 – 0.9872 = 0.012795.. =] awrt **<u>0.0128</u>**	B1 (1)	1.1b
(c)	P($F = 5$) = $0.9^4 \times 0.1$, = 0.06561 = awrt **<u>0.0656</u>**	M1, A1 (2)	3.4 1.1b
(d)	<table><tr><td>n</td><td>1</td><td>2</td><td>…</td><td>10</td></tr><tr><td>P($F = n$)</td><td>0.01</td><td>$0.01 + \alpha$</td><td>…</td><td>$0.01+9\alpha$</td></tr></table>	M1	3.1b
	Sum of probs = 1 $\Rightarrow \frac{10}{2}[2\times 0.01+9\alpha]=1$	M1A1	3.1a 1.1b
	[i.e. $5(0.02 + 9\alpha) = 1$ or $0.1 + 45\alpha = 1$] so α = **<u>0.02</u>**	A1 (4)	1.1b
(e)	P($F = 5$| Thomas' model) = **<u>0.09</u>**	B1ft (1)	3.4
(f)	<u>Peta's</u> model assumes the <u>probability</u> of hitting target is <u>constant</u> (o.e.) **and** <u>Thomas</u>' model assumes this <u>probability increases</u> with each attempt(o.e.)	B1 (1)	3.5a
		(11 marks)	

Notes

(a) 1st B1 for stating that the <u>probability</u> (or possibility or chance) is <u>constant</u> (or fixed or same)
2nd B1 for stating that <u>throws</u> are <u>independent</u> ["trials" are independent is B0]

(b) B1 for awrt 0.0128 (found on calculator)

(c) M1 for a probability expression of the form $(1-p)^4 \times p$ where $0 < p < 1$
A1 for awrt 0.0656
SC Allow M1A0 for answer only of 0.066

(d) 1st M1 for setting up the distribution of F with at least 3 correct values of n and P($F = n$) in terms of α. (Can be implied by 2nd M1 or 1st A1)
2nd M1 for use of sum of probs = 1 **and** clear summation or use of arithmetic series formula (allow 1 error or missing term). (Can be implied by 1st A1)
1st A1 for a correct equation for α
2nd A1 for α = 0.02 (must be exact and come from correct working)

(e) B1ft for value resulting from $0.01 + 4\times$"their α" (provided α and the answer are probs)
Beware If their answer is the same as their (c) (or a rounded version of their (c)) score B0

(f) B1 for a suitable comment about the <u>probability</u> of hitting the target
ALT Allow idea that Peta's model suggests the dart may never hit the target but Thomas' says that it will hit at least once (in the first 10 throws).

Qu 4	Scheme	Marks	AO
(a)	Convenience <u>or</u> opportunity [sampling]	B1 (1)	1.2
(b)	Quota [sampling] e.g. Take 4 people every 10 minutes	B1 B1 (2)	1.1a 1.1b
(c)	Census	B1 (1)	1.2
(d)	[58 – 26 =] **<u>32</u>** (min)	B1 (1)	1.1b
(e)	$\mu = \frac{4133}{95} = 43.505263\ldots$ **awrt <u>43.5</u> (min)** $\sigma_x = \sqrt{\frac{202\,294}{95} - \mu^2} = \sqrt{236.7026...}$ = 15.385… **awrt <u>15.4</u> (min)**	B1 M1 A1 (3)	1.1b 1.1b 1.1b
(f)	There are outliers in the data (or data is skew) which will affect mean and sd Therefore use median and IQR	B1 dB1 (2)	2.4 2.4
(g)	Value of 20, LQ at 26 and outliers will not change <u>or</u> state that median and upper quartile are the values that <u>do</u> change <u>More values now below 40 than above</u> so Q_2 <u>or</u> Q_3 will change and be lower Both Q_2 <u>and</u> Q_3 will be lower	B1 M1 A1 (3)	1.1b 2.1 2.4
		(13 marks)	

Notes

(b)
1st B1 for quota (sampling) mentioned ("Stratified" or "systematic" or "random" are B0B0)
2nd B1 for a description of how such a system might work, requires suitable strata or categories
e.g. time slots, departments, gender, age groups, distance travelled etc
Suggestion of randomness is B0

(e)
B1 for a correct mean (awrt 43.5)
M1 for a correct expression for the sd (including $\sqrt{\ }$)ft their mean
A1 for awrt 15.4 (Allow s = 15.4667… awrt 15.5)

(f)
1st B1 for acknowledging <u>outliers</u> or <u>skewness</u> are a problem for <u>mean and sd</u>
"extreme values"/"anomalies" OK May be implied by saying median and IQR not affected by..
We need to see mention of "outliers", "skewness" and the problem so "data is skewed so use median and IQR" is B0 unless mention that they are not affected by extreme values <u>or</u> mean and standard deviation can be "inflated" by the positive skew etc
2nd dB1 dep on 1st B1 for therefore choosing <u>median and IQR</u>

(g)
B1 for identifying 2 of these 3 groups of unchanged values or stating only Q_2 and Q_3 change
M1 for <u>explaining</u> that median or UQ should be lower.
E.g. the 2 values have moved to below 40 (or 58) and therefore more than 50% below 40 or (more than 75% below 58) <u>or</u> an argument to show that the other 3 values are the same. (o.e.)
Allow arrows on box plot provided statement in words about increased % below 40 or 58 etc
A1 for stating median <u>and</u> UQ are both lower with clear evidence of M1 scored

[If lots of values on 40 then median might not change but, since two values <u>do</u> change then UQ would change. If this meant that 92 became an outlier then we would have a new value for upper whisker and an extra outlier so effectively 3 values are altered. So median changes]

Qu 5	Scheme	Marks	AO
(a)	$P(L > 16) = 0.69146\ldots$ **awrt 0.691**	B1 (1)	1.1b
(b)	$P(L > 20 \mid L > 16) = \dfrac{P(L > 20)}{P(L > 16)}$	M1	3.1b
	$= \dfrac{0.308537...}{(a)}$ or $\dfrac{1-(a)}{(a)}, = 0.44621\ldots$	A1ft, A1	1.1b 1.1b
	For calc to work require $(0.44621...)^4 = 0.03964\ldots$ **awrt <u>0.0396</u>**	dM1 A1 (5)	2.1 1.1b
(c)	Require: $[P(L > 4)]^2 \times [P(L > 20 \mid L > 16)]^2$	M1	1.1a
	$= (0.99976...)^2 \times ("0.44621...")^2$	A1ft	1.1b
	$= 0.19901\ldots$ **awrt <u>0.199</u> (*)**	A1cso*	1.1b
		(3)	
(d)	$H_0 : \mu = 18 \quad H_1 : \mu > 18$	B1	2.5
	$\bar{L} \sim N\left(18, \left(\dfrac{4}{\sqrt{20}}\right)^2\right)$	M1	3.3
	$P(\bar{L} > 19.2) = P(Z > 1.3416\ldots) = 0.089856\ldots$	A1	3.4
	(0.0899 > 5%) <u>or</u> (19.2 < 19.5) <u>or</u> 1.34 < 1.6449 so not significant	A1	1.1b
	Insufficient evidence to support Alice's claim (or belief)	A1 (5)	3.5a
		(14 marks)	

	Notes
(a)	B1 for evaluating probability using their calculator (awrt 0.691) Accept 0.6915
(b)	1st M1 for a first step of identifying a suitable conditional probability (either form) 1st A1ft for a ratio of probabilities with numerator = awrt 0.309 or 1 – (a) and denom = their (a) 2nd A1 for awrt 0.446 (o.e.) Accept 0.4465 (from $\frac{0.3085}{0.691} = 0.44645...$) NB $\frac{P(16<L<20)}{P(L>16)} = 0.5538\ldots$ scores M1A1A1 when they do 1 – 0.5538 = 0.4462… 2nd M1 (dep on 1st M1) for 2nd correct step i.e. (their 0.446…)^4 <u>or</u> X~B(4, "0.446") and P(X = 4) 3rd A1 for awrt 0.0396
(c)	1st M1 for a correct approach to solving the problem (May be implied by A1ft) 1st A1ft for P(L > 4) = awrt 0.9998 used <u>and</u> ft their 0.44621 in correct expression If use P(L > 20) = 0.3085.. as 0.446.. in (b) then M1 for $(0.3085..)^2 \times [P(L > 4)]^2$; A1ft as above * 2nd A1cso for 0.199 or better with clear evidence of M1 [NB $(0.4662..)^2 = 0.199\ldots$ is M0A0A0] **Must see M1 scored by correct expression in symbols or values (M1A1ft)**
(d)	B1 for both hypotheses in terms of μ. M1 for selecting a suitable model. Sight of <u>normal</u>, <u>mean</u> 18, <u>sd</u> $\frac{4}{\sqrt{20}}$ (o.e.) or <u>variance</u> = 0.8 1st A1 for using the model correctly. Allow awrt 0.0899 <u>or</u> 0.09 from correct prob. statement **ALT** **CR** $(\bar{L}) > 19.471\ldots$ (accept awrt 19.5) <u>or</u> **CV** of 1.6449 (or better: calc 1.6448536..) 2nd A1 for correct non-contextual conclusion. Wrong comparison or contradictions A0 Error giving 2nd A0 implies 3rd A0 but just a correct contextual conclusion can score A1A1 3rd A1 dep on M1 and 1st A1 for a correct contextual conclusion mentioning <u>Alice's claim</u> /<u>belief</u> <u>or</u> there is insufficient evidence that the mean <u>lifetime</u> is more than 18 hours