Y1S7 XMQs and MS

(Total: 43 marks)

1. P31(AS)_2018 Q3 . 7 marks - Y1S7 Hypothesis testing

2. P31(AS)_2019 Q5 . 6 marks - Y1S7 Hypothesis testing

3. P31(AS)_2020 Q4 . 7 marks - Y1S1 Data collection

4. P31(AS)_2021 Q4 . 10 marks - Y1S6 Statistical distributions

5. P31(AS)_2022 Q2 . 7 marks - Y1S6 Statistical distributions

6. P31_2022   Q4 . 6 marks - Y1S6 Statistical distributions

3. Naasir is playing a game with two friends. The game is designed to be a game of chance so that the probability of Naasir winning each game is $\frac{1}{3}$

Naasir and his friends play the game 15 times.

(a) Find the probability that Naasir wins

(i) exactly 2 games,

(ii) more than 5 games.

(3)

Naasir claims he has a method to help him win more than $\frac{1}{3}$ of the games. To test this claim, the three of them played the game again 32 times and Naasir won 16 of these games.

(b) Stating your hypotheses clearly, test Naasir's claim at the 5% level of significance.

(4)

Qu	Scheme	Marks	AO
3 (a)	Let N = the number of games Naasir wins $N \sim B(15, \frac{1}{3})$	M1	3.3
(i)	$P(N = 2) = 0.059946\ldots$ **awrt 0.0599**	A1	1.1b
(ii)	$P(N > 5) = 1 - P(N \leqslant 5) = 0.38162\ldots$ **awrt 0.382**	A1 (3)	1.1b
(b)	$H_0 : p = \frac{1}{3}$ $H_1 : p > \frac{1}{3}$	B1	2.5
	Let X = the number of games Naasir wins $X \sim B(32, \frac{1}{3})$	M1	3.3
	$P(X \geqslant 16) = 1 - P(X \leqslant 15) = 0.03765$ (< 0.05)	A1	3.4
	[Significant result so reject H_0 (the null model) and conclude:] There is evidence to support Naasir's claim (o.e.)	A1 (4)	3.5a
		(7 marks)	

	Notes
(a)	M1 for selecting a binomial model with correct n or p Award for sight of $B(15, \frac{1}{3})$ (o.e. e.g. in words) or implied by 1 correct answer 1st A1 for awrt 0.0599 (from a calculator). Allow 0.05995 2nd A1 for awrt 0.382 (from a calculator)
(b)	B1 for correctly stating both hypotheses in terms of p or π Accept $p = 0.\dot{3}$ or any exact equivalent. $H_1 : p \geqslant \frac{1}{3}$ is B0 M1 for selecting a suitable model to use for the test. Award for sight of $B(32, \frac{1}{3})$ (o.e. e.g. in words) or implied by 0.03765 1st A1 for use of the model to calculate an appropriate probability using calc. Sight of $P(X \geqslant 16)$ **and** answer awrt 0.0377
ALT	**CR** May use CR so award 1st A1 for CR of $X \geqslant 16$ must have seen some probabilities though: 1 of $P(X \leqslant 15) = 0.9623$ or $P(X \leqslant 14) = 0.9224$ or 0.9223 2nd A1 for conclusion in context that there is support for Naasir's claim Must mention "Naasir" or "his" and "claim" or "method" (o.e.) or e.g. probability of winning a game is $> \frac{1}{3}$ or has increased Dependent on M1 and 1st A1 but can ignore hypotheses.
SC	**Use of 0.3 for $\frac{1}{3}$** If used 0.3 instead of $\frac{1}{3}$ in (a) and score M0A0A0 can condone use of 0.3 in (b) 1st A1 ft needs $P(X \geqslant 16) = 0.0138$ or CR of $X \geqslant 15$ and sight of 1 of $P(X \geqslant 15) = 0.0327$ or $P(X \geqslant 14) = 0.0694$ 2nd A1 as before with 0.3 instead $\frac{1}{3}$ (if appropriate)

5. Past records show that 15% of customers at a shop buy chocolate. The shopkeeper believes that moving the chocolate closer to the till will increase the proportion of customers buying chocolate.

After moving the chocolate closer to the till, a random sample of 30 customers is taken and 8 of them are found to have bought chocolate.

Julie carries out a hypothesis test, at the 5% level of significance, to test the shopkeeper's belief.

Julie's hypothesis test is shown below.

$H_0 : p = 0.15$

$H_1 : p \geqslant 0.15$

Let X = the number of customers who buy chocolate.

$X \sim B(30, 0.15)$

$P(X = 8) = 0.0420$

$0.0420 < 0.05$ so reject H_0

There is sufficient evidence to suggest that the proportion of customers buying chocolate has increased.

(a) Identify the first two errors that Julie has made in her hypothesis test.

(2)

(b) Explain whether or not these errors will affect the conclusion of her hypothesis test. Give a reason for your answer.

(1)

(c) Find, using a 5% level of significance, the critical region for a one-tailed test of the shopkeeper's belief. The probability in the tail should be less than 0.05

(2)

(d) Find the actual level of significance of this test.

(1)

Question	Scheme	Marks	AOs
5(a)	The alternative hypothesis should be $H_1 : p > 0.15$	B1	2.5
	The calculation of the test statistic should be $P(X \geq 8)$ [= 0.0698]	B1	2.3
		(2)	
(b)	These will affect the conclusion (as the null hypothesis should not be rejected) since $P(X \geq 8)$ [= 0.0698] is greater than 0.05	B1	2.4
		(1)	
(c)	$P(X \leq 8) = 0.9722\ldots > 0.95$ or $P(X \geq 9) = 0.0277\ldots < 0.05$	M1	2.1
	CR: $\{X \geq 9\}$	A1	1.1b
		(2)	
(d)	awrt **0.0278**	B1ft	1.1b
		(1)	
			(6 marks)

Notes	
(a)	**B1:** Identifying that $\geq$ should be $>$ in the alternative hypothesis **B1:** Identifying that $P(X = 8)$ should be $P(X \geq 8)$ Stating $P(X = 8)$ is incorrect on its own is insufficient Check for errors identified and corrected next to the question
(b)	**B1:** Will affect conclusion **and** correct supporting reason
(c)	**M1:** For use of tables to find probability associated with critical value [$P(X \leq 8)$ or $P(X \geq 9)$ with B(30, 0.15) (may be implied by either correct probability awrt 0.97 or awrt 0.03) or by the correct CR] **A1:** [$30 \geq$]$X \geq 9$ o.e. e.g. $X > 8$ Allow '9 or more' or 'CR $\geq$ 9'
(d)	**B1ft:** awrt 0.0278 (allow awrt 2.78%) or correct ft their one-tailed upper CR from B(30, 0.15) to 3s.f.

4. A lake contains three different types of carp.

There are an estimated 450 mirror carp, 300 leather carp and 850 common carp.

Tim wishes to investigate the health of the fish in the lake.

He decides to take a sample of 160 fish.

(a) Give a reason why stratified random sampling cannot be used.

(1)

(b) Explain how a sample of size 160 could be taken to ensure that the estimated populations of each type of carp are fairly represented.

You should state the name of the sampling method used.

(2)

As part of the health check, Tim weighed the fish.

His results are given in the table below.

Weight (w kg)	Frequency (f)	Midpoint (m kg)
$2 \leqslant w < 3.5$	8	2.75
$3.5 \leqslant w < 4$	32	3.75
$4 \leqslant w < 4.5$	64	4.25
$4.5 \leqslant w < 5$	40	4.75
$5 \leqslant w < 6$	16	5.5

(You may use $\sum fm = 692$ and $\sum fm^2 = 3053$)

(c) Calculate an estimate for the standard deviation of the weight of the carp.

(2)

Tim realised that he had transposed the figures for 2 of the weights of the fish.

He had recorded in the table 2.3 instead of 3.2 and 4.6 instead of 6.4

(d) Without calculating a new estimate for the standard deviation, state what effect

(i) using the correct figure of 3.2 instead of 2.3

(ii) using the correct figure of 6.4 instead of 4.6

would have on your estimated standard deviation.

Give a reason for each of your answers.

(2)

Question	Scheme	Marks	AOs
4(a)	It is not possible to have a sampling frame	B1	2.3
		(1)	
(b)	Quota sampling **and** (catch 85 common carp, 45 mirror carp and 30 leather carp) **or** (ignore any fish caught of a type where the quota is full)	M1	1.1a
	Quota sampling **and** catch 85 common carp, 45 mirror carp and 30 leather carp **and** ignore any fish caught of a type where the quota is full	A1	1.1b
		(2)	
(c)	$\sigma = \sqrt{\frac{3053}{160} - \left(\frac{692}{160}\right)^2}$	M1	1.1b
	$= 0.6129\ldots$ awrt 0.613	A1	1.1b
		(2)	
(d)(i)	This would have no effect as the piece of data would remain in the same class	B1	2.2a
(ii)	This would increase the standard deviation as change in mean is small and $6.4 - 4.6 \approx 3\sigma$ therefore estimate of standard deviation will increase	B1	2.2a
		(2)	
			(7 marks)

Notes		
(a)	**B1:**	For the idea there cannot be a sampling frame/list
(b)	**M1:**	Quota sampling **and** either for the correct numbers of each type **or** for the idea that if quota full ignore the fish.
	A1:	Quota sampling **and** both the correct numbers of each type **and** for the idea that if quota full ignore the fish or sample until all quotas are full
(c)	**M1:**	A correct expression for σ
	A1:	Awrt 0.613 allow s = awrt 0.615
(d)	**B1:**	Correct deduction with suitable explanation Allow range for class. Do not allow there is no differences
	B1:	Correct deduction with suitable explanation. so would increase the standard deviation and a suitable reason. Allow the value is bigger than any others in the table **oe**

4. A nursery has a sack containing a large number of coloured beads of which 14% are coloured red.

Aliya takes a random sample of 18 beads from the sack to make a bracelet.

(a) State a suitable binomial distribution to model the number of red beads in Aliya's bracelet.

(1)

(b) Use this binomial distribution to find the probability that

(i) Aliya has just 1 red bead in her bracelet,

(ii) there are at least 4 red beads in Aliya's bracelet.

(3)

(c) Comment on the suitability of a binomial distribution to model this situation.

(1)

After several children have used beads from the sack, the nursery teacher decides to test whether or not the proportion of red beads in the sack has changed.
She takes a random sample of 75 beads and finds 4 red beads.

(d) Stating your hypotheses clearly, use a 5% significance level to carry out a suitable test for the teacher.

(4)

(e) Find the p-value in this case.

(1)

Qu	Scheme	Marks		AO
4. (a)	[R = no. of red beads in Aliya's bracelet] $R \sim B(18, 0.14)$	B1 **(1)**		3.3
(b)(i)	P(R = 1) = 0.19403… awrt **0.194**	B1		1.1b
(ii)	P(R …4) = 1 – P(R ,, 3) = 1 – [0.76184…]	M1		3.4
	= 0.2381588… awrt **<u>0.238</u>**	A1 **(3)**		1.1b
(c)	Requires p = 0.14 to be constant so need a large number of beads in the sack to ensure that removing 18 beads does not appreciably affect this probability, then it could be suitable.	B1 **(1)**		3.5b
(d)	$H_0 : p = 0.14$ $H_1 : p \neq 0.14$	B1		2.5
	[X = number of red beads in the sample] $X \sim B(75, 0.14)$	M1		3.3
	P(X ,, 4) = 0.01506… or if B(75, 0.14) seen awrt 0.02	A1		3.4
	{0.02 < 0.025 so significant <u>or</u> reject H_0 } There is evidence that the proportion of red beads has changed	A1 **(4)**		2.2b
(e)	p-value is $2\times$"0.01506..." = 0.030123... = awrt 0.03	B1ft **(1)**		1.1b
		(10 marks)		

	Notes
(a)	B1 for B(18, 0.14) accept in words e.g. <u>binomial</u> with <u>n = 18</u> and <u>p = 0.14</u>
(b)(i)	B1 for awrt 0.194
(ii)	M1 for interpreting "at least 4" Need 1 – P(R ,, 3) <u>and</u> 1 – p [0 < p < 1] P(R = 3) = 0.233.. OK A1 for awrt 0.238
(c)	B1 for mention of <u>large number of beads</u> and need for <u>p = 0.14 to be constant</u> for it to be suitable. Do NOT accept e.g. "events are independent"
(d)	B1 for both hypotheses correct with use of p or π M1 for selecting a suitable model: sight or correct use of B(75, 0.14) May be implied by sight of 0.015 or better <u>or</u> [P(X > 4) =] 0.9849… i.e. 0.985 or better 1st A1 for use of the correct model awrt 0.015 (accept awrt 0.02 following a correct expression) Allow 1st A1 for awrt 0.985 <u>only if</u> correct comparison with 0.975 is seen. Sight of B(75, 0.14) and P(X ,, 4) = awrt 0.02 scores M1A1 <u>No sight</u> of B(75, 0.14) <u>but</u> sight of awrt 0.015 scores M1($\Rightarrow$)A1[Condone P(X = 4) =..] 2nd A1 (**dep on M1A1)** for a correct conclusion in context mentioning "proportion", "red" and "changed" If there is a statement about H_0 or significance it must be compatible.
NB	May see CR i.e. X ,, 4 (mark when prob seen) and X …18 (prob = 0.01406..) Ignore upper limit NB for information P(X = 4) = 0.0104… and can only score M1A0A0 if B(75, 0.14) seen
(e)	B1ft for awrt 0.03 Allow ft of their probability in (d) provided at least 3sf used NB an answer of 0.02 in (d) leading to 0.04 in (e) is B0
SC	Use of CR will give significance level of 0.01506…+ 0.01406… = 0.029… score B1 **no ft**

2. A manufacturer of sweets knows that 8% of the bags of sugar delivered from supplier *A* will be damp.
A random sample of 35 bags of sugar is taken from supplier *A*.

(a) Using a suitable model, find the probability that the number of bags of sugar that are damp is

(i) exactly 2

(ii) more than 3

(3)

Supplier *B* claims that when it supplies bags of sugar, the proportion of bags that are damp is less than 8%

The manufacturer takes a random sample of 70 bags of sugar from supplier *B* and finds that only 2 of the bags are damp.

(b) Carry out a suitable test to assess supplier *B*'s claim.
You should state your hypotheses clearly and use a 10% level of significance.

(4)

Qu	Scheme	Mark	AO
2. (a)	[D = number of bags that are damp] $D \sim$ B(35, 0.08) NB 0.08 = $\frac{2}{25}$	M1	3.3
(i)	P(D = 2) = 0.2430497… awrt **0.243**	A1	3.4
(ii)	P(D > 3) = $\left[1-\text{P}(D \text{ ,, } 3) = 1-0.69397...\right]$ = 0.30602… awrt **0.306**	A1	1.1b
		(3)	
(b)	$H_0 : p = 0.08$ $H_1 : p < 0.08$	B1	2.5
	[$X \sim$] B(70, 0.08)	M1	2.1
	$\left[\text{P}(X \text{ ,, } 2)\right] = 0.0739756…$ awrt **0.074**	A1	1.1b
	[0.074 < 0.10 so significant, reject H_0 so…]		
	there is evidence to support supplier B's claim (o.e.)	A1	2.2b
		(4)	
		(7 marks)	

Notes

(a) M1 for selecting a correct model: sight of or use of B(35, 0.08) [Condone B(0.08, 35)]
May be implied by one correct answer or sight of P(D ,, 3) = awrt 0.694 (or allow 0.693)
or seeing $\binom{35}{2}0.08^2 \times (1-0.08)^{35-2}$
Saying B(35, 8%) without a correct calculation would score M0

(i) 1st A1 for awrt 0.243

(ii) 2nd A1 for awrt 0.306 (Condone poor use of notation e.g. P(D = 3) = 0.306… i.e. just mark ans)

NB $\text{P}(D \ldots 3) = 0.539$ scores 2nd A0 but would of course score M1

(b) B1 for both hypotheses correct in terms of p or π [Condone 8% for 0.08]

M1 for sight or correct use of B(70, 0.08) [Condone B(0.08, 70)]
May be implied by prob of 0.074 or better

1st A1 for final answer awrt 0.074 can condone poor notation e.g. P(X = 2) = awrt 0.074
Can allow this mark for CR of X ,, 2 provided $\left[\text{P}(X \text{ ,, } 2)\right] = 0.074$ (or better) is seen
[Can allow 0.07 if $X \sim$B(70, 0.08) and P(X ,, 2) are both seen]

2nd A1 (**dep on** M1A1 but independent of hypotheses) for a correct inference in context
Must mention claim or B **and** idea of support for …
or proportion/probability (of damp bags) **and** idea of less than 8% or A
2nd A0 for contradictory statements e.g. "accept H_0 so evidence to support B's claim"
2nd A0 if you see 0.0739… < 0.08 so significant/ reject H_0 etc

MR **0.8 for 0.08**
In (a) allow M1 for B(35, 0.8) then A0A0
In (b) allow B1 for Hypotheses and M1 for B(70, 0.8) seen, then A0A0

4. A dentist knows from past records that 10% of customers arrive late for their appointment.

A new manager believes that there has been a change in the proportion of customers who arrive late for their appointment.

A random sample of 50 of the dentist's customers is taken.

(a) Write down

- a null hypothesis corresponding to no change in the proportion of customers who arrive late
- an alternative hypothesis corresponding to the manager's belief

(1)

(b) Using a 5% level of significance, find the critical region for a two-tailed test of the null hypothesis in (a)
You should state the probability of rejection in each tail, which should be less than 0.025

(3)

(c) Find the actual level of significance of the test based on your critical region from part (b)

(1)

The manager observes that 15 of the 50 customers arrived late for their appointment.

(d) With reference to part (b), comment on the manager's belief.

(1)

Question	Scheme	Marks	AOs
4(a)	$H_0: p = 0.1$ $H_1: p \neq 0.1$	B1	2.5
		(1)	
(b)	Use of $X \sim B(50, 0.1)$ implied by sight of one of awrt 0.0052 or awrt 0.9755 or awrt 0.0245	M1	3.4
	Critical regions $X = 0$ or $X \geqslant 10$	A1	1.1b
	$X = 0$ and $X \geqslant 10$ plus $P(X = 0) =$ awrt 0.0052 and $P(X \geqslant 10) =$ awrt 0.0245	A1	1.1b
	SC: Both CR correct with no probabilities and no distribution seen scores M0A1A0		
		(3)	
(c)	0.0297	B1ft	1.1b
		(1)	
(d)	15 is <u>in the critical region</u> therefore there is evidence to support the **<u>manager</u>**'s belief	B1ft	2.2b
		(1)	
			(6 marks)

Notes

(a)	**B1**	For both hypotheses in terms of p or π. Connected to H_0 and H_1 correctly Condone 10% but not 10
(b)	**M1**	Using correct distribution to find the probability associated with one tail of the CR If the correct distribution is <u>stated</u> (may be seen in part(a)) allow for one tail of the correct CR **or** one of (awrt 0.025 or awrt 0.005 or awrt 0.975) seen connected to a correct probability statement
	A1	Lower CR $X = 0$ / $X < 1$ / $X \leqslant 0$ / [condone eg $P(X = 0)$ labelled as CR] Or Upper CR $X \geqslant 10$ or $X > 9$ [condone $P(X \geqslant 10)$ oe labelled as CR]
	A1	Both CR's correct with the relevant probabilities Allow $\cup$ for "and" and $X > 9$, $X < 1$, $X \leqslant 0$ [**do not allow** $P(X = 0)$ or $P(X \geqslant 10)$ oe] Allow CR in different form eg $(9, \infty)$, $[10, \infty)$
(c)	**B1ft**	awrt 0.0297 or 2.97% or ft for the sum of the probabilities in (b) for "their 2 critical regions" if seen. If none seen it must be awrt 0.0297 **SC M0 in (b) for a one tail test** Allow B1ft for their one tail CR in (b) eg 0.0338 or 0.0245 or 0.0579
(d)	**B1ft**	A correct statement about 15 and "their CR" or sight $P(X \geqslant 15) = 0.0000738...$ and comparison with "their 0.0245" **and** a compatible correct statement in context. eg There is evidence that there has been a change in the **<u>proportion/probability</u>** arriving **<u>late</u>** Condone increase rather than change Do not allow contradicting statements. NB No CR given in (b) then B0