Subject – Math AI(Higher Level)
Topic - Statistics and Probability
Year - May 2021 – Nov 2022
Paper -2
Answers

Question 1

(a) (i) $P(Y) = 0.8 \times 0.1 + 0.2 \times 0.3$ ***M1***

$= 0.14$ ***A1***

(ii) $P(\text{Star} \mid Y) = \dfrac{0.8 \times 0.1}{0.14}$ ***M1***

$= 0.571 \left(\dfrac{4}{7}, 0.571428\ldots\right)$ ***A1***

[4 marks]

(b) the colours of the sweets are distributed according to manufacturer specifications ***A1***

[1 mark]

(c)

Colour	Brown	Red	Green	Orange	Yellow	Purple
Expected Frequency	12	20	16	16	8	8

A2

> **Note:** Award ***A2*** for all 6 correct expected values, ***A1*** for 4 or 5 correct values, ***A0*** otherwise.

[2 marks]

(d) 5 ***A1***

[1 mark]

(e) 0.469 (0.4688117…) ***A2***

[2 marks]

(f) since $0.469 > 0.05$ ***R1***

fail to reject the null hypothesis. There is insufficient evidence to reject the manufacturer's specifications ***A1***

> **Note:** Award ***R1*** for a correct comparison of their correct p-value to the test level, award ***A1*** for the correct result from that comparison. Do not award ***R0A1***.

[2 marks]

Total [12 marks]

Question 2

(a) (i) evidence of correct probability *(M1)*
e.g. sketch **OR** correct probability statement $\mathrm{P}(X < 6.5)$

0.0151 ***A1***

(ii) 0.0228 ***A1***

Note: Answers should be given to 4 decimal place.

[3 marks]

(b) (i) multiplying **their** probability by 1000 *(M1)*
451.7 ***A1***

(ii) 510.5 ***A1***

[3 marks]

Note: Answers should be given to 4 sf.

(c) H_0 : stopping distances can be modelled by $\mathrm{N}(6.76,\ 0.12^2)$
H_1 : stopping distances cannot be modelled by $\mathrm{N}(6.76,\ 0.12^2)$ ***A1A1***

Note: Award ***A1*** for correct H_0, including reference to the mean and standard deviation. Award ***A1*** for the negation of their H_0.

[2 marks]

(d) 15.1 or 22.8 seen *(M1)*

0.0727 (0.0726542…, 7.27%) ***A2***

[3 marks]

(e) $0.05 < 0.0727$ ***R1***
there is insufficient evidence to reject H_0 (or "accept H_0") ***A1***

Note: Do not award ***R0A1***.

[2 marks]

Total [13 marks]

Question 3

(a) $X_1 \sim \text{Po}(3.1)$

$\text{P}(X_1 = 4) = 0.173 \ (0.173349...)$ ***A1***

[1 mark]

(b) (i) $X_2 \sim \text{Po}(3 \times 3.1) = \text{Po}(9.3)$ ***(M1)***

$\text{P}(X_2 = 12) = 0.0799 \ (0.0798950...)$ ***A1***

(ii) $(\text{P}(X_1 > 0))^2 \times \text{P}(X_1 = 0)$ ***(M1)***

$0.95495^2 \times 0.04505$ ***(A1)***

$= 0.0411 \ (0.0410817...)$ ***A1***

[5 marks]

(c) $\text{P}(X_1 = 0) = 0.04505$ ***(A1)***

$X_1 \sim \text{B}(12, 0.04505)$ ***(M1)(A1)***

Note: Award ***M1*** for recognizing binomial probability, and ***A1*** for correct parameters.

$= 0.0133 \ (0.013283....)$ ***A1***

[4 marks]

(d) **METHOD ONE**

n	λ	$\text{P}(X \geq 30)$
...	...	...
10	24.1	0.136705
11	26.2	0.253384

(M1)(A1)(A1)

Note: Award ***M1*** for evidence of a cumulative Poisson with $\lambda = 3.1 + 2.1n$, ***A1*** for 0.136705 and ***A1*** for 0.253384.

so require 12 magpies (including Bill) ***A1***

METHOD TWO

evidence of a cumulative Poisson with $\lambda = 3.1 + 2.1n$ ***(M1)***

sketch of curve and $y = 0.2$ ***(A1)***

(intersect at) $10.5810\ldots$ ***(A1)***

rounding up gives $n = 11$

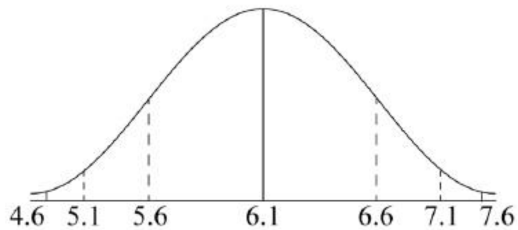
so require 12 magpies (including Bill) ***A1***

[4 marks]

Total [14 marks]

Question 4

(a)



A1A1

Note: Award ***A1*** for a normal curve with mean labelled 6.1 or μ, ***A1*** for indication of SD (0.5): marks on horizontal axis at 5.6 and/or 6.6 **OR** $\mu - 0.5$ and/or $\mu + 0.5$ on the correct side and approximately correct position.

[2 marks]

(b) $X \sim \mathrm{N}(6.1,\ 0.5^2)$

$\mathrm{P}(5.5 < X < 6.5)$ **OR** labelled sketch of region ***(M1)***

$= 0.673\ \ (0.673074\ldots)$ ***A1***

[2 marks]

(c) $(\mathrm{P}(X < 5.3) =)\ 0.0547992\ldots$ ***(A1)***

$0.0547992\ldots \times 80$ ***(M1)***

$= 4.38\ \ (4.38393\ldots)$ ***A1***

[3 marks]

(d) (i) $Y \sim \mathrm{N}(4.5,\ 0.45^2)$,

$(\mathrm{P}(Y > 4.62) =)\ 0.394862\ldots$ ***(A1)***

use of binomial seen or implied ***(M1)***

using $\mathrm{B}(10,\ 0.394862\ldots)$ ***(M1)***

$0.0430\ \ (0.0429664\ldots)$ ***A1***

(ii) $np(1-p) = 2.39\ \ (2.38946\ldots)$ ***A1***

[5 marks]

(e) $\mathrm{P}(F \cap (W > 4.7)) = 0.5 \times 0.3284\ \ (= 0.1642)$ ***(A1)***

attempt use of tree diagram **OR** use of $\mathrm{P}(F \mid W > 4.7) = \dfrac{\mathrm{P}(F \cap (W > 4.7))}{\mathrm{P}(W > 4.7)}$ ***(M1)***

$$\frac{0.5 \times 0.3284}{0.5 \times 0.9974 + 0.5 \times 0.3284}$$ ***(A1)***

$= 0.248\ \ (0.247669\ldots)$ ***A1***

[4 marks]

Total [16 marks]

Question 7

(a) (i) let X be the random variable "number of patients arriving in a minute", such that $X \sim \mathrm{Po}(m)$.

$\mathrm{H}_0 : m = 1.5$ ***A1***

$\mathrm{H}_1 : m > 1.5$ ***A1***

> **Note:** Allow a value of 270 for m. Award at most ***A0A1*** if it is not clear that it is the population mean being referred to e.g
>
> H_0 : *The number of patients is equal to 1.5 every minute*
>
> H_1 : *The number of patients exceeds 1.5 every minute.*
>
> Referring to the "expected" number of patients or the use of μ or λ is sufficient for ***A1A1***.

(ii) under H_0 let Y be the number of patients in 3 hours

$Y \sim \mathrm{Po}(270)$ ***(A1)***

$\mathrm{P}(Y \geq 320) \left(= 1 - \mathrm{P}(Y \leq 319)\right) = 0.00166 \quad (0.00165874)$ ***(M1)A1***

since $0.00166 < 0.05$ ***R1***

(reject H_0)

Loreto should employ more staff ***A1***

[7 marks]

(b) (i) H_0 : The probability of a patient waiting less than 20 minutes is 0.95 ***A1***

H_1 : The probability of a patient waiting less than 20 minutes is less than 0.95

A1

(ii) under H_0 let W be the number of patients waiting more than 20 minutes

$W \sim \mathrm{B}(150,\ 0.05)$ ***(A1)***

$\mathrm{P}(W \geq 11) = 0.132 \quad (0.132215...)$ ***(M1)A1***

since $0.132 > 0.1$ ***R1***

(fail to reject H_0)

insufficient evidence to suggest they are not meeting their target ***A1***

> **Note:** Do not accept "they are meeting target" for the ***A1***.
> Accept use of $\mathrm{B}(150,\ 0.95)$ and $\mathrm{P}(W \leq 139)$ and any consistent use of a random variable, appropriate p-value and significance level.

[7 marks]

Total: [14 marks]

Question 6

(a) (i) Let X be the random variable "distance from O".

$X \sim \mathrm{N}(10,\ 3^2)$

$\mathrm{P}(X < 13) = 0.841\ \ (0.841344...)$ ***(M1)A1***

(ii) $(\mathrm{P}(X > 15) =)\ 0.0478\ \ (0.0477903)$ ***A1***

[3 marks]

(b) $\mathrm{P}(X > 15) \times \mathrm{P}(X > 15)$ ***(M1)***

$= 0.00228\ \ (0.00228391...)$ ***A1***

[2 marks]

(c) $1 - (0.8143)^3$ ***(M1)***

$= 0.460\ \ (0.460050...)$ ***A1***

[2 marks]

(d) (i) let Y be the random variable "number of points scored"

evidence of use of binomial distribution ***(M1)***

$Y \sim \mathrm{B}(10,\ 0.539949...)$ ***(A1)***

$(\mathrm{E}(Y) =) 10 \times 0.539949...$ ***(M1)***

$= 5.40$ ***A1***

(ii) $(\mathrm{P}(Y \geq 5) =)\ 0.717\ \ (0.716650...)$ ***A1***

(iii) $\mathrm{P}(5 \leq Y < 8)$ ***(M1)***

$= 0.628\ \ (0.627788...)$ ***A1***

Note: Award ***M1*** for a correct probability statement or indication of correct lower and upper bounds, 5 and 7.

(iv) $\dfrac{\mathrm{P}(5 \leq Y < 8)}{\mathrm{P}(Y \geq 5)} \left(= \dfrac{0.627788...}{0.716650...} \right)$ ***(M1)***

$= 0.876\ \ (0.876003...)$ ***A1***

[9 marks]

Total: [16 marks]

Question 7

(a) (let T be the number of passengers who arrive)

$(\mathrm{P}(T > 72) =)\ \mathrm{P}(T \geq 73)$ **OR** $1 - \mathrm{P}(T \leq 72)$ ***(A1)***

$T \sim \mathrm{B}(74,\ 0.9)$ **OR** $n = 74$ ***(M1)***

$= 0.00379\ \ (0.00379124...)$ ***A1***

> **Note:** Using the distribution $\mathrm{B}(74,\ 0.1)$, to work with the 10% that do not arrive for the flight, here and throughout this question, is a valid approach.

[3 marks]

(b) (i) 72×0.9 ***(M1)***

64.8 ***A1***

(ii) $n \times 0.9 = 72$ ***(M1)***

80 ***A1***

[4 marks]

(c) **METHOD 1**

EITHER

when selling 74 tickets

	$T \leq 72$	$T = 73$	$T = 74$
Income minus compensation (I)	11100	10800	10500
Probability	0.9962…	0.003380…	0.0004110…

top row ***A1A1***

bottom row ***A1A1***

> **Note:** Award ***A1A1*** for each row correct. Award ***A1*** for one correct entry and ***A1*** for the remaining entries correct.

$\mathrm{E}(I) = 11100 \times 0.9962... + 10800 \times 0.00338... + 10500 \times 0.000411 \approx 11099$ ***(M1)A1***

OR
income is $74 \times 150 = 11100$ ***(A1)***

expected compensation is
$0.003380... \times 300 + 0.0004110... \times 600 \ (= 1.26070...)$ ***(M1)A1A1***

Note: The ***(M1)*** is for an attempt to work out expected compensation by multiplying a probability for tickets sold by either 300 or 600.

expected income when selling 74 tickets is $11100 - 1.26070\ldots$ ***(M1)***

Note: Award ***(M1)*** for subtracting their expected compensation from 11100.

$= 11098.73..\ (= \$11099)$ ***A1***

THEN
income for 72 tickets $= 72 \times 150 = 10800$ ***(A1)***
so expected gain $\approx 11099 - 10800 = \$299$ ***A1***

METHOD 2
for 74 tickets sold, let C be the compensation paid out
$P(T = 73) = 0.00338014\ldots, \ P(T = 74) = 0.000411098\ldots$ ***A1A1***

$E(C) = 0.003380... \times 300 + 0.0004110... \times 600 \ (= 1.26070...)$ ***(M1)A1A1***

extra expected revenue $= 300 - 1.01404... - 0.246658\ldots \ (300 - 1.26070\ldots)$
(A1)(M1)

Note: Award ***A1*** for the 300 and ***M1*** for the subtraction.

$= \$299$ (to the nearest dollar) ***A1***

METHOD 3
let D be the change in income when selling 74 tickets.

	$T \le 72$	$T = 73$	$T = 74$
Change in income	300	0	–300

(A1)(A1)

Note: Award ***A1*** for one error, however award ***A1A1*** if there is no explicit mention that $T = 73$ would result in $D = 0$ and the other two are correct.

$P(T \le 73) = 0.9962..., \ P(T = 74) = 0.000411098\ldots$ ***A1A1***

$E(D) = 300 \times 0.9962... + 0 \times 0.003380... - 300 \times 0.0004110$ ***(M1)A1A1***

$= \$299$ ***A1***

[8 marks]
[Total 15 marks]

Question 8

(a) quota **A1**

[1 mark]

(b) (i) $27.125 \approx 27.1$ ***(M1)A1***

(ii) $8.29815\ldots \approx 8.30$ ***A1***

[3 marks]

(c) (let μ be the national mean)

H_0: $\mu = 25.2$

H_1: $\mu > 25.2$ ***A1***

> **Note:** Accept hypotheses in words if they are clearly expressed and '**population** mean' or 'school mean' is referred to. Do not accept H_0: $\mu = \mu_0$ unless μ_0 is explicitly defined as "national standard mark" or given as 25.2.

recognizing t-test ***(M1)***

p-value $= 0.279391\ldots$ ***A1***

$0.279391\ldots > 0.05$ ***R1***

> **Note:** The ***R1*** mark is for the comparison of their p-value with 0.05.

insufficient evidence to reject the null hypothesis (that the mean for the school is 25.2) ***A1***

> **Note:** Award the final ***A1*** only if the **null** hypothesis is also correct (e.g. $\mu_0 = 25.2$ or (population) mean $= 25.2$) and the conclusion is consistent with both the direction of the inequality and the alternative hypothesis.

[5 marks]

(d) **EITHER**

the sampling process is not random ***R1***

For example:

the school asked for volunteers

the students were selected from a single class

OR

the quota might not be representative of the student population ***R1***

For example:

the school may have only 4 boys and 400 girls.

> **Note:** Do not accept 'the sample is too small'.

[1 mark]

(e) (i) $(28.1 \times 2 + 20 =)\ 76.2$ ***A1***

(ii) 8.4×2 ***(A1)***

$= 16.8$ ***A1***

[3 marks]

[Total 13 marks]

Question 8

(a) wood layer, $W \sim \mathrm{N}(7, 0.3^2)$; plastic, $P \sim \mathrm{N}(3, 0.16^2)$

door: $X = W + P$

$\mathrm{E}(X) = 10$ (mm) ***(A1)***

$\mathrm{Var}(X) = \mathrm{Var}(W) + \mathrm{Var}(P) = 0.1156$ (mm^2) ***(M1)(A1)***

recognizing the distribution is Normal, with their mean and variance ***(M1)***

$X \sim \mathrm{N}(10, 0.34^2)$

$\mathrm{P}(X < 9.5) = 0.0707$ $(0.07070125...)$ ***A1***

[5 marks]

(b) $\mathrm{E}(T) = 80$ ***(A1)***

$\mathrm{Var}(T)(= 0.1156 \times 8) = 0.9248$ ***(M1)(A1)***

$T \sim \mathrm{N}(80, 0.9248)$

$\mathrm{P}(T > 82) = 0.0188$ $(0.0187753\ldots)$ ***A1***

[4 marks]

(c) (i) 6.93 mm $(6.93428\ldots)$ ***A1***

(ii) $(s_{n-1} =)$ 0.404 ***(A1)***

$(s_{n-1}^2 =)$ 0.163 mm^2 $(0.162928\ldots)$ ***A1***

[3 marks]

(d) $\mathrm{H}_0: \mu_A = \mu_B$ and $\mathrm{H}_1: \mu_A > \mu_B$ ***A1A1***

> **Note:** Award ***A1*** for use of μ or in words "population mean", and ***A1*** for both correct equality in null hypothesis and correct inequality in alternative hypothesis. Accept an equivalent statement in words, must include mean and reference to "**population** mean" / "mean for **all** Machine B layers" for the first ***A1*** to be awarded.

use a two-sample t-test ***(M1)***

p-value $= 0.406975...$ ***A1***

since $0.406975... > 0.05$ **OR** p-value > 0.05 ***R1***

Do not reject H_0 (Insufficient evidence to support the employee's claim) ***A1***

> **Note:** Accept a p-value of $0.415861\ldots$ from use of 3sf values from part (c). Follow through within the question for the final ***R1*** and ***A1*** for their p-value provided $0 \le p \le 1$. Do not award ***R0A1***.

[6 marks]

Total [18 marks]

Question 9

(a) (i) $\frac{370+472}{2}$ ***(M1)***

Note: This ***(M1)*** can also be awarded for either a correct Q_3 or a correct Q_1 in part (a)(ii).

$Q_3 = 421$ ***A1***

(ii) their part (a)(i) – their Q_1 (clearly stated) ***(M1)***

$\text{IQR} = (421-318=)\ 103$ ***A1***

[4 marks]

(b) $(Q_3 + 1.5(\text{IQR}) =)\ 421 + (1.5 \times 103)$ ***(M1)***

$= 575.5$

since $498 < 575.5$ ***R1***

Netherlands is not an outlier ***A1***

Note: The ***R1*** is dependent on the ***(M1)***. Do not award ***R0A1***.

[3 marks]

(c) not appropriate ("no" is sufficient) ***A1***

as r is too close to zero / too weak a correlation ***R1***

[2 marks]

(d) (i) 6 ***A1***

(ii) 4.5 ***A1***

(iii) 4.5 ***A1***

[3 marks]

(e) (i) $r_s = 0.683\ \ (0.682646\ldots)$ ***A2***

(ii) **EITHER**

there is a (positive) association between the population size and the score ***A1***

OR

there is a (positive) linear correlation between the ranks of the population size and the ranks of the scores (when compared with the PMCC of 0.249) ***A1***

[3 marks]

(f) lowering the top score by 20 does not change its rank so r_s is unchanged ***R1***

Note: Accept "this would not alter the rank" or "Netherlands still top rank" or similar. Condone any statement that clearly implies the ranks have not changed, for example: "The Netherlands still has the highest score."

[1 mark]

[Total 16 marks]

Question 10

(a) (i) $(m=)\ 54(\%)$ ***A1***

(ii) $(n=)\ 14(\%)$ ***A1***

(iii) $(p=)\ 22(\%)$ ***A1***

(iv) $(q=)\ 10(\%)$ ***A1***

Note: Based on their n, follow through for parts (i) and (iii), but only if it does not contradict the given information. Follow through for part (iv) but only if the total is 100%.

[4 marks]

(b) (i) $0.54\ \left(\frac{54}{100}, \frac{27}{50}, 54\%\right)$ ***A1***

(ii) $\frac{54}{64}\ \left(0.844, \frac{27}{32}, 84.4\%, 0.84375\right)$ ***A1A1***

Note: Award ***A1*** for a correct denominator (0.64 or 64 seen), ***A1*** for the correct final answer.

[3 marks]

(c) (i) recognizing Binomial distribution with correct parameters ***(M1)***

$X \sim \mathrm{B}(10, 0.68)$

$(\mathrm{P}(X=5)=)\ 0.123\ \ (0.122940..., 12.3\%)$ ***A1***

(ii) $1-\mathrm{P}(X \le 3)$ **OR** $\mathrm{P}(X \ge 4)$ **OR** $\mathrm{P}(4 \le X \le 10)$ ***(M1)***

$0.984\ (0.984497..., 98.4\%)$ ***A1***

(iii) $(0.68)^9 \times 0.32$ ***(M1)***

recognition of two possible cases ***(M1)***

$2\times\left((0.68)^9 \times 0.32\right)$

$0.0199\ \ (0.0198957..., 1.99\%)$ ***A1***

[7 marks]

(d) **EITHER**

the probability is not constant ***A1***

OR

the events are not independent ***A1***

OR

the events should be modelled by the hypergeometric distribution instead ***A1***

[1 mark]

Total [15 marks]