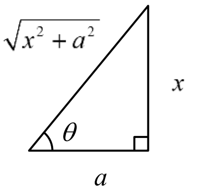
2025 ACJC H2 Math Preliminary Examination Paper 2 Markers' Report

Qn	Solution	Remarks
1(a)	$l_{OB}:\mathbf{r}=\lambda\overrightarrow{OB}=\lambda\begin{pmatrix}5\\-4\\3\end{pmatrix}$ Since R lies on the line OB, $\overrightarrow{OR}=\lambda\begin{pmatrix}5\\-4\\3\end{pmatrix}$. $\overrightarrow{AR}=\overrightarrow{OR}-\overrightarrow{OA}=\begin{pmatrix}5\lambda-3\\-4\lambda-1\\3\lambda-3\end{pmatrix}$ $\overrightarrow{AR}\cdot\begin{pmatrix}5\\-4\\3\end{pmatrix}=0\Rightarrow\begin{pmatrix}5\lambda-3\\-4\lambda-1\\3\lambda-3\end{pmatrix}\cdot\begin{pmatrix}5\\-4\\3\end{pmatrix}=0$ $25\lambda-15+16\lambda+4+9\lambda-9=0$ $\therefore\lambda=\frac{2}{5}$ $\overrightarrow{OR}=\frac{2}{5}\begin{pmatrix}5\\-4\\3\end{pmatrix}$ $\overrightarrow{OR}=\frac{\overrightarrow{OA}+\overrightarrow{OA'}}{2}=\frac{2}{5}\begin{pmatrix}5\\-4\\3\end{pmatrix}$ $\overrightarrow{OA'}=2\overrightarrow{OR}-\overrightarrow{OA}=\frac{4}{5}\begin{pmatrix}5\\-4\\3\end{pmatrix}-\begin{pmatrix}3\\1\\3\end{pmatrix}=\begin{pmatrix}1\\-\frac{21}{5}\\-\frac{3}{5}\end{pmatrix}=\frac{1}{5}\begin{pmatrix}5\\-21\\-3\end{pmatrix}$ **Alternatively (for finding foot of perpendicular from point to line),** $\overrightarrow{OR}=(\mathbf{a}.\mathbf{b})\mathbf{b}$ $=\left(\begin{pmatrix}3\\1\\3\end{pmatrix}.\frac{1}{\sqrt{50}}\begin{pmatrix}5\\-4\\3\end{pmatrix}\right)\frac{1}{\sqrt{50}}\begin{pmatrix}5\\-4\\3\end{pmatrix}$ $=\frac{1}{50}\left(\begin{pmatrix}3\\1\\3\end{pmatrix}.\begin{pmatrix}5\\-4\\3\end{pmatrix}\right)\begin{pmatrix}5\\-4\\3\end{pmatrix}=\frac{15-4+9}{50}\begin{pmatrix}5\\-4\\3\end{pmatrix}=\frac{2}{5}\begin{pmatrix}5\\-4\\3\end{pmatrix}$	Generally well done. Students understood that the procedure involves $\overrightarrow{AR}\cdot\overrightarrow{OB}=0$ or applying the projection vector formula to find $\overrightarrow{OR}$ and using the midpoint theorem to find $\overrightarrow{OA'}$. However, some students failed to recognize that points O, R, and B are collinear, hence, $\overrightarrow{OR}=\lambda\overrightarrow{OB}$. This oversight led to unnecessarily complicated expressions and careless mistakes in the calculations. eg $\overrightarrow{AR}=\begin{pmatrix}x-3\\y-1\\z-3\end{pmatrix}$ or $\overrightarrow{OR}=\begin{pmatrix}5\\-4\\3\end{pmatrix}+\lambda\begin{pmatrix}5\\-4\\3\end{pmatrix}$.

2(a)	$\mathrm{f}(x)=\dfrac{1}{\sqrt{1-4x^2}}=\left(1-4x^2\right)^{-\frac{1}{2}}$ $=1-\dfrac{1}{2}(-4x^2)+\frac{-\frac{1}{2}(-\frac{3}{2})}{2!}(-4x^2)^2+\frac{-\frac{1}{2}(-\frac{3}{2})(-\frac{5}{2})}{3!}(-4x^2)^3+\ldots$ $=1+2x^2+6x^4+20x^6+\ldots$	Most students knew the procedure to be carried out, i.e. express f(x) in power form and use the maclaurin expansion formula $\left(1+x\right)^n$ to obtain the first 4 non-zero terms. A handful of students used the repeated differentiation method; however, no marks were awarded as this did not follow the instructions of the question. Common mistakes: 1. Expressing $\mathrm{f}(x)$ in an incorrect power form due to carelessness, eg. $\left(1-4x^2\right)^{-1}$ or $\left(1-4x^2\right)^{\frac{1}{2}}$. 2. Omitting the negative sign in the expansion, eg $1-\frac{1}{2}(4x^2)+\frac{-\frac{1}{2}(-\frac{3}{2})}{2!}(4x^2)^2+\ldots$ 3. Taking out 4 and then applying the Maclaurin expansion formula directly, which is a misapplication of the formula.
2(b)	$\displaystyle\int\frac{1}{\sqrt{1-4x^2}}\ \mathrm{d}x=\int 1+2x^2+6x^4+20x^6\ \mathrm{d}x$ $\dfrac{1}{2}\sin^{-1}2x=x+\dfrac{2}{3}x^3+\dfrac{6}{5}x^5+\dfrac{20}{7}x^7+c$ $\sin^{-1}2x=2x+\dfrac{4}{3}x^3+\dfrac{12}{5}x^5+\dfrac{40}{7}x^7+d$ $\sin^{-1}0=0 \quad \therefore d=0$ $\sin^{-1}2x=2x+\dfrac{4}{3}x^3+\dfrac{12}{5}x^5+\dfrac{40}{7}x^7$	Most students recognized that $\dfrac{\mathrm{d}}{\mathrm{d}x}\sin^{-1}2x=\dfrac{2}{\sqrt{1-4x^2}}$; however, many either forgot to include the factor of 2 or were unsure how to use the results from 2(a) to proceed. Several students attempted repeated differentiation, which is a lengthy and tedious process. While some successfully obtained the result by integrating the previous answer, they forgot to include the constant of integration C, and consequently were unable to explain why C disappears in the final solution.

3(a) (i)	$\sum_{r=1}^{n} r(r+1)^2 = \sum_{r=1}^{n} r(r^2+2r+1) = \sum_{r=1}^{n}(r^3+2r^2+r)$ $= \frac{n^2(n+1)^2}{4} + 2\left[\frac{n(n+1)(2n+1)}{6}\right] + 1+2+\ldots+n$ $= \frac{n^2(n+1)^2}{4} + \frac{n(n+1)(2n+1)}{3} + \frac{n}{2}(n+1)$ $= \frac{3n^2(n+1)^2}{12} + \frac{4n(n+1)(2n+1)}{12} + \frac{6n}{12}(n+1)$ $= \frac{n(n+1)}{12}\left[3n(n+1)+4(2n+1)+6\right]$ $= \frac{n(n+1)}{12}(3n^2+3n+8n+4+6)$ $= \frac{n(n+1)}{12}(3n^2+11n+10)$ $= \frac{n(n+1)(n+2)(3n+5)}{12}$	Mostly well done, except that some students were unfamiliar with the AP formula for $\sum_{r=1}^{n} r$. Students who combined the three fractions by expansion had difficulty arriving at the final result; a better approach is to factorize instead.
3(a) (ii)	$\sum_{r=5}^{n-1}(r+2)(r+3)^2$ Replace r with $r-2$ $\sum_{r-2=5}^{r-2=n-1}(r-2+2)(r-1+3)^2$ $= \sum_{r=7}^{n+1}(r)(r+1)^2$ $= \sum_{r=1}^{n+1} r(r+1)^2 - \sum_{r=1}^{6} r(r+1)^2$ $= \frac{(n+1)(n+2)(n+3)(3(n+1)+5)}{12}$ $- \frac{6(6+1)(6+2)(3(6)+5)}{12}$ $= \frac{(n+1)(n+2)(n+3)(3n+8)}{12} - 644$	Many students tried replacing r with $r+2$ in the expression from part (i) but failed to recognize the connection to part (ii). A more straightforward approach is to replace r with $r-2$ in the expression given in part (ii). Additionally, some students made careless mistakes when applying the change of limits formula or substituting values into the result from part (i).
3(b)	$u_2 = \frac{1}{1-2} = -1$ $u_3 = \frac{1}{1-(-1)} = \frac{1}{2}$ $u_4 = \frac{1}{1-\frac{1}{2}} = 2$ $u_{2025} = u_3 = \frac{1}{2}$	Very well done—almost all students were able to find the three terms. However, some students failed to recognize the pattern and were therefore unable to find u_{2025}.

4(a)	Let $y=\dfrac{1}{x^2+6x+5}$. $x^2+6x+5=\dfrac{1}{y}$ $(x+3)^2-4=\dfrac{1}{y}$ $(x+3)^2=\dfrac{1}{y}+4$ $x=-3+\sqrt{\dfrac{1}{y}+4}$ since $x\ge -3$ $f^{-1}(x)=-3+\sqrt{\dfrac{1}{x}+4}$	Most students were able to recognise the need to make x the subject in order to find the rule of the inverse function. Students need to be reminded that "±" must be considered on taking the square root, together with the reason for rejecting the negative root. The expression inside the square root must be simplified. The following mistakes should **NOT** be made: $(x+1)(x+5)=\dfrac{1}{y}\Rightarrow x+1=\dfrac{1}{y}$ or $x+5=\dfrac{1}{y}$ $(x+3)^2=\dfrac{1}{y}+4\Rightarrow x+3=\sqrt{\dfrac{1}{y}}+2$
4(b)	Let $y=\dfrac{1}{x^2+6x+5}$. $yx^2+6yx+5y=1$ $yx^2+6yx+5y-1=0$ This quadratic equation must have real roots for $y\in R_f$. Discriminant ≥ 0 $36y^2-4y(5y-1)\ge 0$ $16y^2+4y\ge 0$ $y(16y+4)\ge 0$ $y\le -\dfrac{1}{4}$ or $y>0$ $(\text{as } y\ne 0)$ $R_f=\left(-\infty,-\dfrac{1}{4}\right]\cup(0,\infty)$	Many students had no idea how to approach this question **algebraically**. Some students did partial fractions just to be able to do differentiation to obtain the stationary point. There were no tests to determine that this is a maximum point. There were also missing working for the interval $(0,\infty)$, though some students tried to explain graphically. Some students were not familiar with the set notation, especially the "$\cup$" sign and the interval notation. It is **WRONG** to write $(\infty,0)$. Some students did not handle inequalities with care, as "4*y*" should be included as a factor in $16y^2+4y\ge 0$. Thus, the result from $4y+1\ge 0$ was wrong and incomplete.
4(c)	$R_f=\left(-\infty,-\dfrac{1}{4}\right]\cup(0,\infty)\xrightarrow{g}\left(0,e^{-0.25}\right]\cup(1,\infty)$ $R_{gf}=\left(0,e^{-0.25}\right]\cup(1,\infty)$	Some students gave the wrong rule for the composite function gf, which should be $gf(x)=e^{\frac{1}{x^2+6x+5}}$. As exact range of gf is required, $gf(-3)=e^{\frac{1}{-4}}$ or $g\left(-\dfrac{1}{4}\right)=e^{-\frac{1}{4}}$ or $g(0)=1$ in **exact** form is expected.

5(a)	$\int \frac{1}{x\sqrt{x^2+a^2}}\,\mathrm{d}x$ $= \int \frac{1}{a\tan\theta\sqrt{a^2\tan^2\theta+a^2}}\left(a\sec^2\theta\right)\mathrm{d}\theta$ $= \int \frac{1}{a\tan\theta\left(a\sec\theta\right)}\left(a\sec^2\theta\right)\mathrm{d}\theta$ $= \int \frac{\sec\theta}{a\tan\theta}\,\mathrm{d}\theta$ $= \frac{1}{a}\int \operatorname{cosec}\theta\,\mathrm{d}\theta$ $= -\frac{1}{a}\ln\left\|\operatorname{cosec}\theta+\cot\theta\right\|+c$ $= -\frac{1}{a}\ln\left\|\frac{\sqrt{x^2+a^2}+a}{x}\right\|+c$ $= \frac{1}{a}\ln\left(\frac{x}{\sqrt{x^2+a^2}+a}\right)+c\ (\because x, a>0)$ $x = a\tan\theta$ $\frac{\mathrm{d}x}{\mathrm{d}\theta} = a\sec^2\theta$ $\tan\theta = \frac{x}{a}$  $\sin\theta = \frac{x}{\sqrt{x^2+a^2}}$	It's very strange to see solutions expanding $\sec^2\theta$ instead of simplifying $a^2\tan^2\theta+a^2$ to $a^2\sec^2\theta$. Some students were not able to simplify $\frac{\sec\theta}{\tan\theta}$ to $\operatorname{cosec}\theta$, while some others were not able to use MF27 for $\int \operatorname{cosec}\theta\,\mathrm{d}\theta = -\ln\left\|\operatorname{cosec}\theta+\cot\theta\right\|+c$ and the modulus sign was often missing. Many steps were wasted to convert $\operatorname{cosec}\theta+\cot\theta$ in terms of $\tan\theta$ for $\frac{x}{a}$ instead of using the right-angled triangle to do so. As this is a "show" question, it would be good to include the explanation that $x, a>0$ for converting the modulus sign to round brackets at the end.
5(b)(i)	$\frac{\mathrm{d}x}{\mathrm{d}t} = xy$ and $\frac{\mathrm{d}y}{\mathrm{d}t} = x^2 \Rightarrow \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{x^2}{xy} = \frac{x}{y}$ $\Rightarrow y\frac{\mathrm{d}y}{\mathrm{d}x} = x \Rightarrow \int y\,\mathrm{d}y = \int x\,\mathrm{d}x$ $\Rightarrow \frac{y^2}{2} = \frac{x^2}{2}+c \Rightarrow y^2 = x^2+d$ When $t=0, x=4, y=5,\ d=5^2-4^2=9$ $\therefore y^2 = x^2+9$ or $y=\sqrt{x^2+9}\ (\because y>0)$	Integral signs were missing when separating the variables. After multiplying throughout by 2 for $\frac{y^2}{2}=\frac{x^2}{2}+c$, the arbitrary constant should be changed to $2c$ or d where $d = 2c$. As this is a "show" question again, it would be required to include the reason that $y>0$ as "±" must be considered on taking the square root at the end.
5(b)(ii)	$\frac{\mathrm{d}x}{\mathrm{d}t} = xy = x\sqrt{x^2+9}$ $\Rightarrow \int \frac{1}{x\sqrt{x^2+9}}\,\mathrm{d}x = \int 1\,\mathrm{d}t$ Using **(a)** and $a=3$, $\Rightarrow t = \frac{1}{3}\ln\left(\frac{x}{\sqrt{x^2+9}+3}\right)+c$ When $t=0, x=4,\ c = 0-\frac{1}{3}\ln\left(\frac{4}{\sqrt{4^2+9}+3}\right) = \frac{1}{3}\ln 2$ $\Rightarrow t = \frac{1}{3}\ln\frac{x}{\sqrt{x^2+9}+3}+\frac{1}{3}\ln 2$	Some students could not understand the instructions to obtain a differential equation in terms of x and t only, with a few of them simply substitute the initial value of y to carry on wrongly. Some students could not see the need to apply part (a) after separating the variables, while some others took the value of a to be 9 instead of 3. The question asked for an expression of t in terms of x, so there was no need to waste extra steps to do the usual expressing x in terms of t, to find the arbitrary constant.

6(a)

Given $\mathrm{P}(B \mid A) = 0.2$, $\mathrm{P}(B \mid A) = 0.2$ and $\mathrm{P}(A' \cap B') = 0.3$.

$$\mathrm{P}(B \mid A) = 0.2 \Leftrightarrow \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)} = 0.2 \Leftrightarrow \mathrm{P}(A) = 5\,\mathrm{P}(A \cap B) \quad (1)$$

$$\mathrm{P}(A \mid B) = 0.6 \Leftrightarrow \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = 0.6 \Leftrightarrow \mathrm{P}(B) = \frac{5}{3}\mathrm{P}(A \cap B) \quad (2)$$

Given $\mathrm{P}(A' \cap B') = 0.3$,

$\therefore \mathrm{P}(A \cup B) = 1 - \mathrm{P}(A' \cap B') = 0.7$

Now $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$ …(3)

Sub. (1) and (2) into (3)

$$0.7 = 5\,\mathrm{P}(A \cap B) + \frac{5}{3}\mathrm{P}(A \cap B) - \mathrm{P}(A \cap B)$$

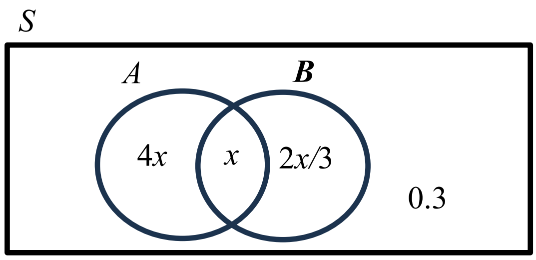
$$\mathrm{P}(A \cap B) = \frac{21}{170} \text{ or } 0.124 \text{ (3 sig fig)}$$

Alternatively,

$$\mathrm{P}(B \mid A) = 0.2 \Leftrightarrow \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)} = 0.2 \Leftrightarrow \mathrm{P}(A) = \frac{\mathrm{P}(A \cap B)}{0.2}$$

$$\mathrm{P}(A \mid B) = 0.6 \Leftrightarrow \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = 0.6 \Leftrightarrow \mathrm{P}(B) = \frac{\mathrm{P}(A \cap B)}{0.6}$$

Let $x = \mathrm{P}(A \cap B)$.



$$\mathrm{P}(A) = \frac{x}{0.2} = \frac{10x}{2} = 5x \quad \ldots(1)$$

$$\mathrm{P}(B) = \frac{x}{0.6} = \frac{10x}{6} = \frac{5x}{3} \quad \ldots(2)$$

Given $\mathrm{P}(A' \cap B') = 0.3$

From the Venn Diagram,

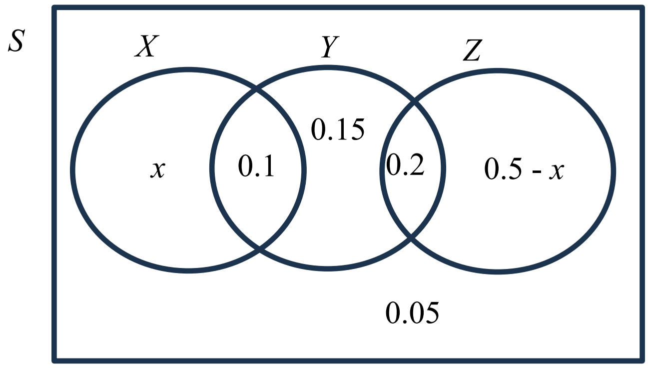
$$4x + x + \frac{2x}{3} + 0.3 = 1$$

$$\frac{17x}{3} = \frac{7}{10}$$

$$x = \frac{21}{170} \text{ or } 0.124 \text{ (3 s.f.)}$$

Generally, well-done.
The definition for conditional probability was well-known, and the usage of appropriate probability results led to many correct solutions.

The only issue that recurred a few times were students using $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$. This is <u>not true</u> unless the events A and B are independent. This was also an issue with (b).

6(b)	$\mathrm{P}\left[(X \cup Y \cup Z)'\right] = 1 - \mathrm{P}(X \cup Y \cup Z)$ $= 1 - 0.95 = 0.05$ Since X and Z are mutually exclusive, $\mathrm{P}(Y \cap X' \cap Z') = \mathrm{P}(Y \cap X') - \mathrm{P}(Y \cap Z)$ $= 0.35 - 0.2 = 0.15$ $y = \mathrm{P}(Z \cap Y')$ Let $x = \mathrm{P}(X \cap Y')$. $\mathrm{P}(Z \cap Y') = 1 - 0.5 - x = 0.5 - x$ We have the following Venn diagram:  $\therefore 0 \le x \le 0.5$ If $P(X \cap Y') = 0$ then $P(X) = 0.1$ (minimum). If $P(X \cap Y') = 0.5$ then $P(X) = 0.6$ (maximum).	Most students utilised a Venn diagram to visualise the situation, and the idea of mutual exclusivity was well-understood. Those who got this correct usually have one of the following 2 approaches: (1) assigning an unknown (method shown), or (2) considering visually how to maximise X's area (probability) by "moving Z into or out of Y". A more common careless mistake was the mistreatment of the 0.1 probability (forgetting that it contributes to the probability of X).
7(a)	Step 1: Select 4 people out of 12 people to play bridge: ${}^{12}C_4$ Step 2: Select 4 people out of the remaining 8 people to play carrom: ${}^{8}C_4$ Step 3: Select the remaining 4 people to play mahjong: ${}^{4}C_4$ By Principle of Multiplication, we have ${}^{12}C_4 \times {}^{8}C_4 \times {}^{4}C_4 = 34650$ Number of ways to break 12 people into 3 groups of 4 to play 3 different games = 34650	Students kept trying to permutate the people within each group (leading to 4!), or permutate the groups (leading to 3!).
7(b)	**Method 1: By complementation** Step 1: Put Aaron and Sandy into one of the 3 groups (Bridge, Carrom or UNO): 3 Step 2: Select 2 people from the other 10 people to complete the group Aaron and Sandy are in: ${}^{10}C_2$ Step 3: Select 4 out of the remaining 8 people to join one of the group that Aaron and Sandy are not in: ${}^{8}C_4$ Step 4: Select the remaining 4 people to form the last group: ${}^{4}C_4$ By complementation, we have $34650 - 3 \times {}^{10}C_2 \times {}^{8}C_4 \times {}^{4}C_4 = 25200$ Number of ways where Aaron and Sandy are not in the same group = 25200	Same issue as above occurred, but students generally could try to apply complementation, or by considering cases.

	Method 2: Direct Method Step 1: Assign Aaron to one of the 3 groups: ${}^{3}C_{1}$ Step 2: Assign Sandy to one of the 2 groups where Aaron is not in: ${}^{2}C_{1}$ Step 3: Select 3 people to complete the group Aaron is in: ${}^{10}C_{3}$ Step 4: Select 3 people to complete the group Sandy is in: ${}^{7}C_{3}$ Step 5: Select 4 people to form the last group: ${}^{4}C_{4}$ By Principle of Multiplication, we have ${}^{3}C_{1}\times{}^{2}C_{1}\times{}^{10}C_{3}\times{}^{7}C_{3}\times{}^{4}C_{4}=3\times2\times120\times35\times1=25200$ Number of ways where Aaron and Sandy are not in the same group = 25200	
7(c)	Let A = {Aaron and Billy are together in a group} Let B = {Aaron and Sandy are in different group} Step 1: Put Aaron and Billy in the same group: ${}^{3}C_{1}$ Step 2: Put Sandy in one of the other 2 groups: ${}^{2}C_{1}$ Step 3: Select 2 people from the remaining 9 people to complete the group Aaron and Billy are in: ${}^{9}C_{2}$ Step 4: Select 3 people from the remaining 7 people to complete the group Sand is in: ${}^{7}C_{3}$ Step 5: Select the remaining 4 people to form the 3$^{\text{rd}}$ group: ${}^{4}C_{4}$ By Principle of Multiplication, $n(A\cap B)={}^{3}C_{1}\times{}^{2}C_{1}\times{}^{9}C_{2}\times{}^{7}C_{3}\times{}^{4}C_{4}$ $=3\times2\times36\times35\times1=7560$ From **(b)**, $n(B)=25,200$ **Method 1:** $\mathrm{P}(A\mid B)=\dfrac{n(A\cap B)}{n(B)}=\dfrac{7560}{25200}=\dfrac{3}{10}$ or 0.3 **Method 2:** From **(a)**, $n(S)=34650$ $\mathrm{P}(A\cap B)=\dfrac{7560}{34650}=\dfrac{12}{55}$ $\mathrm{P}(B)=\dfrac{25,200}{34650}=\dfrac{8}{11}$ $\mathrm{P}(A\mid B)=\dfrac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}=\dfrac{12}{55}\div\dfrac{8}{11}=\dfrac{3}{10}$ or 0.3	As this result is a probability, students who managed to count wrongly (in the same way) for both the numerator and denominator could still have arrived at the correct answer. Generally, the conditional probability was well-considered, and students could process the event in the numerator well. There were a number of shorter solutions that were given full credit as well.

<table>
<tr>
<td>8</td>
<td>
Given the probability distribution of X:

x	-2	-1	0	1	3
$\mathrm{P}(X=x)$	0.1	a	0.3	b	c

$$\mathrm{E}(X)=(-2)(\frac{1}{10})+(-1)(a)+(0)(\frac{3}{10})+(1)(b)+(3)(c)$$
$$\frac{9}{10}=-\frac{2}{10}-a+b+3c$$
$$-a+b+3c=\frac{11}{10} \text{ .... (1)}$$
$$\mathrm{E}(X^2)=(-2)^2(\frac{1}{10})+(-1)^2(a)+(0)^2(\frac{3}{10})+(1)^2(b)+(3)^2(c)$$
$$=\frac{4}{10}+a+b+9c$$
$$\mathrm{Var}(X)=\mathrm{E}(X^2)-\left[\mathrm{E}(X)\right]^2$$
$$\frac{299}{100}=\left(\frac{4}{10}+a+b+9c\right)-\left(\frac{9}{10}\right)^2$$
$$a+b+9c=\frac{34}{10} \text{ .... (2)}$$

$$\sum \mathrm{P}(X=x)=1$$
$$\frac{1}{10}+a+\frac{3}{10}+b+c=1$$
$$a+b+c=\frac{6}{10} \text{ ..... (3)}$$

Solving (1), (2) and (3), we have
$a=\frac{1}{10}$ or 0.1, $b=\frac{3}{20}$ or 0.15, $c=\frac{7}{20}$ or 0.35
</td>
<td>
A small number of students squared the probability instead of x in computing $\mathrm{E}(X^2)$.

Most did not apply the fact that $\sum P(X=x)=1$. As a result, the 3 unknowns cannot be solved.

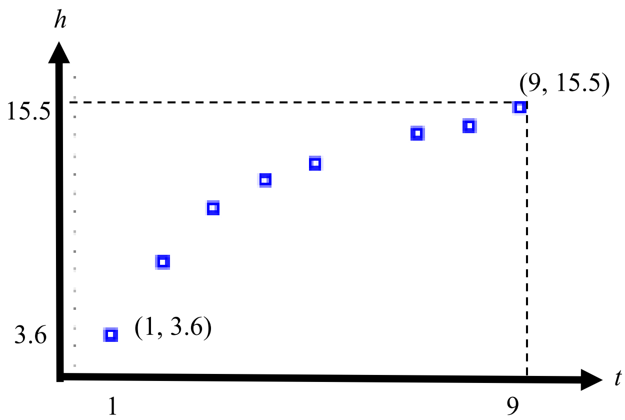
Many input E(X) = 1.1 instead of E(X) = 0.9 into the statement $\mathrm{Var}(X)=\mathrm{E}(X^2)-[\mathrm{E}(X)]^2$.
</td>
</tr>
<tr>
<td>9(a)</td>
<td>
The probability that an orange is rotten is a constant of 0.01.

Whether an orange is rotten is independent of whether any other oranges are rotten.
</td>
<td>
The most common mistake is using the term "probability" instead of "event" when stating the assumption for independence, an error that has been emphasized repeatedly.

Instead of stating the "probability that an orange is rotten", quite a number stated "probability that the number of oranges …", whereas some stated "probability that all oranges …"

Some stated the assumptions needed for sampling to be random.

Some stated that there is only 2 outcomes "success" and "failure" for each trial which is a fact and not an assumption.
</td>
</tr>
</table>

9(b)	Let the random variable X denote the number of rotten oranges in a box of 20 oranges. $X \sim \mathrm{B}(20, 0.01)$ $\mathrm{P}(2 \le X \le 5) = \mathrm{P}(X \le 5) - \mathrm{P}(X \le 1) = 0.0169$	Wrote $X \sim \mathrm{B}(0.01, 20)$ instead of $X \sim \mathrm{B}(20, 0.01)$. Computed $\mathrm{P}(2 < X < 5)$ instead. Many did not know that $\mathrm{P}(2 \le X \le 5) = \mathrm{P}(X \le 5) - \mathrm{P}(X \le 1)$
9(c)	$\mathrm{E}(X) = (20)(0.01) = 0.2$	Used $p = 0.1$ or $p = 0.0169$. Did not know that $\mathrm{E}(X) = np$. For some reason, some felt compelled to round off the answer for $\mathrm{E}(X)$ into an integer. Despite having obtained the answer $\mathrm{E}(X) = 0.2$, some still did not realise that this was the answer and went further to multiply 0.2 with some other numbers such as 20.
9(d)	Let the random variable Y denote the number of substandard boxes out of the entire truckload of 10 boxes. $Y \sim \mathrm{B}(10, p)$ where $p = 0.0169$ $\mathrm{P}(Y \le 2) = 0.999$ (3 s.f.)	Many did not get the right answer here because $p = 0.0169$ was wrongly computed in 9(b).
9(e)	Let A = {he rejects the entire truckload} $\mathrm{P}(A) = \mathrm{P}(X \ge 2) + \mathrm{P}(X = 1)\mathrm{P}(X \ge 2)$ $= [1 - \mathrm{P}(X \le 1)] + \mathrm{P}(X = 1)[1 - \mathrm{P}(X \le 1)]$ $= 0.0196$	Probabilities were computed using random variable Y in 9(d) instead of X. In the case of opening 2 boxes, this case was computed using conditional probability. Some simply could not figure out the 2 cases. Some broke down into 3 cases.
10(a)		Treated last 3 points as a point of inflexion when the last point is merely a point that is slightly higher (an overestimate). Many still did not indicate the scales on both axes. Some wrote the scale on one axis but not the other. Some just labelled the first and the last points but it is still important to show the scales on the axes.

	From the scatter diagram, there appears to be a non-linear relationship between h and t. Thus linear model $h=a+bt$ is not suitable. As t increases, h increases at a decreasing rate (h increases by decreasing amounts), whereas the model $h=a+bt^2$, $(b>0)$ h increases at an increasing rate as t increases. Thus, relationship between h and t should not be modelled by $h=a+bt^2$ (where $b>0$).	Explained using the values of r for each model. Some computed the values of r whereas some did not even show the values of r. However, the question asked candidates to explain by referring to the scatter diagram that was sketched. The condition $b>0$ is essential and important but has been left out when quoting the model $h=a+bt^2$. Without computing the values of r, the scatter diagram could well be perceived as $h=a+bt^2$, $b<0$. Some claim that the model $h=a+bt^2$, $b>0$ must first decrease and then increase, which is not necessarily so for a restricted range of values for t. Many described a concave upward curve as exponential or exponentially increasing.
10(b)	For $h=c+d\sqrt{t}$: $r=0.97364=0.974$ (3 s.f.) For $h=c+d\ln t$: $r=0.99658=0.997$ (3 s.f.) The product moment correlation coefficient of the model $h=c+d\ln t$ <u>is closer to one</u>, hence, $h=c+d\ln t$ is a better model. $h=3.890166185+5.311381407\ln t$ $h=3.890+5.311\ln t$ (3 d.p)	Many chose a model without computing and comparing the values of r for both models, but by some alternative argument. As a result, many chose model A for a variety of reasons including carelessness, despite having computed the values of r. For those who have chosen model B, the equation of the regression line was wrong possibly due to entering wrong values for the data into the GC. Some mixed up the values of c and d, giving the answer $h=3.890\ln t+5.311$.
10(c)	When $t=6$, $h=3.890+5.311\ln 6=$ 13.4 This estimate is likely to be reliable as $t=6$ lies within the t data range $(1\le t\le 9)$ and $r=0.997$ is close to 1 indicating a strong linear relationship between h and ln t.	Wrong values of h was computed mainly because of wrong model chosen in 10(b) or wrong equation. A common mistake was computing $h=3.890+5.311(6)$ instead of $h=3.890+5.311\ln 6$. Almost every candidate left out the condition that r is very close to 1.

<table>
<tr><td>10(d)</td><td>As $t \to \infty$, $\ln t \to \infty$, $h \to \infty$.
The model is not suitable in the long run as the height of bamboo stalk will not increase indefinitely.

Alternative answer:

The model is only appropriate for the range $1 \le t \le 9$. In the long run, the model is unsuitable as $t > 9$.</td><td>Reasons given for the model being suitable or in some cases unsuitable, are contradictory. For example, “the model is suitable because the bamboo will not grow infinitely tall”.

Many thought that the logarithmic function has a horizontal asymptote and thus, the model is suitable.</td></tr>
<tr><td rowspan="2">10(e)</td><td>$H = 3.890 + 5.311\ln t$ where H is the predicted value of height of bamboo stalk using the least squares regression line:</td><td>Many did not know how to compute S. For some who did know how to, did so inaccurately.</td></tr>
<tr><td colspan="2"><table>
<tr><td>t</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>7</td><td>8</td><td>9</td></tr>
<tr><td>h</td><td>3.6</td><td>7.4</td><td>10.2</td><td>11.7</td><td>12.6</td><td>14.1</td><td>14.5</td><td>15.5</td></tr>
<tr><td>H</td><td>3.89</td><td>7.5713</td><td>9.7247</td><td>11.253</td><td>12.438</td><td>14.225</td><td>14.934</td><td>15.559</td></tr>
</table>
$S = \sum (h - H)^2 = 0.77$ (2 dp)</td></tr>
<tr><td>10(f)</td><td>For $H' = p + q\ln t$, $\sum (h - H')^2 \ge \sum (h - H)^2 = S$ since S is the <u>minimum sum of squared residuals</u> as H is calculated from the least squares regression line.
$\therefore \sum (h - H')^2$ will be greater than (or equal to) S.</td><td>Many answered, “less” and a handful stated “same”.

Many used the term “lesser”. “Less” is used for comparing quantity whereas “lesser” is used for comparing quality.

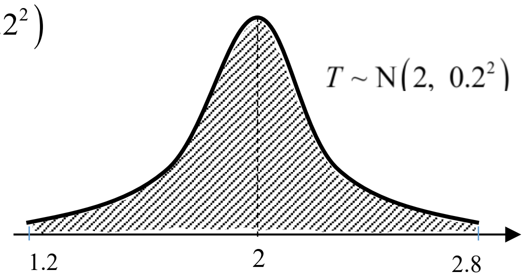
Explanation was cumbersome instead of stating the key term, “least squares regression line”.</td></tr>
<tr><td>11(a)</td><td>μ represent the <u>population mean</u> time taken to test an electric circuit board.

To test $H_0 : \mu = 5.82$
Against $H_1 : \mu < 5.82$ at 5% significance level</td><td>Some students did not define μ, as required by the question. The word ‘population’ must be seen in defining μ.

Incorrect hypothesis:
(a) $H_0 : \mu_0 = 5.82$
$H_1 : \mu_1 < 5.82$
(testing μ_0 and μ_1 instead of μ)

(b) $H_0 : \mu = 5.74$
$H_1 : \mu < 5.74$
(confused between population mean 5.82 and sample mean 5.74)
It is important to read carefully to distinguish between population mean & sample mean for hypothesis question.</td></tr>
</table>

11(b)	Under H_0, $\bar{X} \sim N\left(5.82, \frac{0.63^2}{150}\right)$ approx. by Central Limit Theorem since n = 150 is large. Value of test statistic, $z = \frac{5.74-5.82}{\sqrt{\frac{0.63^2}{150}}} = -1.56$ (3 sf) p – value $= P(\bar{X} > 5.74) = 0.0599 > 0.05$ $\therefore$ Do not reject H_0. There is insufficient evidence at 5% level of significance to conclude that the inspector carries out the test more quickly.	Common mistakes • Putting the sample mean value 5.74 in the distribution. Should be $\bar{X} \sim N\left(5.82, \frac{0.63^2}{150}\right)$. • Writing X instead of $\bar{X}$ • Writing 0.63 instead of 0.63^2 • Finding $s^2 = \frac{150}{149}\left(0.63^2\right)$ (This is not needed since the population variance 0.63^2 is known) • Failing to state CLT • Some students left out the 5% sig. level at the conclusion part. • Some conclusion were over-assertive, stating that 'there is sufficient evidence that mean time taken is 5.82', instead of 'insufficient evidence that mean time is less than 5.82.'
11(c)	There is no need for the laboratory supervisor to know anything about the population distribution of X. Since n is large, by Central Limit Theorem, $\bar{X}$ will be approximately normally distributed.	Many wrote that 'by CLT, the population distribution will be normally distributed'. It must be stated clearly that '$\bar{X}$ will be normally distributed.'
11(d)	(1) The time taken to test an electric circuit board follows a normal distribution (since sample size of 20 is small). (2) The population standard deviation of the time taken to test an electric circuit board is still 0.63.	As the question already states that random sample is taken, there is no need to assume that the time taken is independent of each other. It must be stated clearly that 'the population distribution **of *X*** follows a normal distribution'. There were many incorrect responses that wrote 'distribution of mean time taken follows a normal distribution'.

11(e)	To test $H_0: \mu = \mu_0$ Against $H_1: \mu \neq \mu_0$ at 1% level of significance Under H_0, $\overline{X} \sim N\left(\mu_0, \dfrac{0.63^2}{20}\right)$ Value of test statistic, $z = \dfrac{5.88-\mu_0}{\sqrt{\dfrac{0.63^2}{20}}}$ In order not to reject H_0, $-2.57583 < \dfrac{5.88-\mu_0}{\sqrt{\dfrac{0.63^2}{20}}} < 2.57583$ $-2.57583\left(\dfrac{0.63}{\sqrt{20}}\right) < 5.88-\mu_0$ and $5.88-\mu_0 < 2.57583\left(\dfrac{0.63}{\sqrt{20}}\right)$ $\Rightarrow \mu_0 < 6.24$ and $5.52 < \mu_0$ $\Rightarrow 5.52 < \mu_0 < 6.24$	Many were confused with what to test in the hypotheses. Common mistakes • $H_0: \mu_0 = 5.82$ $H_1: \mu_0 \neq 5.82$ • Writing the range of values as $-0.12533 < z < 0.12533$. Incorrect critical values due to incorrect interpretation $P(-0.12533 < z < 0.12533) = 0.01$ • Writing the range of values as $z < -2.57583$ or $z > 2.57583$. Note that 'H_0 is **not** rejected' does not correspond to tail-ends
12(a)	$T \sim N\left(2,\ 0.2^2\right)$  Note: $P(1.2 < T < 2.8) = 0.99994$	Since the interval (1.2, 2.8) is within 4σ from the mean 2 (note that $P(1.2 < T < 2.8) = 0.99994$), the markings for 1.2 and 2.8 should be drawn such that the curve is almost parallel and very close to the horizonal axis. Only very negligible unshaded region should be seen under the graph.
12(b)	Let the random variable T denote the time taken in minutes to categorise an incoming call. Let the random variable X denote the time taken in minutes to resolve a routine call. Let the random variable Y denote the time taken in minutes to resolve a complex call. $T \sim N\left(2, 0.2^2\right)$, $X \sim N(5, k)$, $Y \sim N\left(20, 6^2\right)$ $\therefore T + X \sim N\left(7,\ k + 0.2^2\right)$ Given: $P(T + X > 8) = 0.254$ $\Rightarrow P\left(Z > \dfrac{8-7}{\sqrt{k+0.2^2}}\right) = 0.254$ $\Rightarrow \dfrac{1}{\sqrt{k+0.2^2}} = 0.66196$ $\therefore k = 2.24$ (to 3 s.f.)	Mostly done well, except for minority who failed to standardize to show the result.

<table>
<tr>
<td>12(c)</td>
<td>

$\bar{Y} = \dfrac{Y_1 + Y_2 + \ldots + Y_n}{n} \sim \mathrm{N}\left(20,\ \dfrac{6^2}{n}\right)$

Given: $\mathrm{P}(\bar{Y} > 24) \leq 0.01$

$\Rightarrow \mathrm{P}\left(Z > \dfrac{24-20}{\sqrt{\dfrac{6^2}{n}}}\right) \leq 0.01$

$\Rightarrow \dfrac{24-20}{\sqrt{\dfrac{6^2}{n}}} \geq 2.32635$

$\Rightarrow \dfrac{4\sqrt{n}}{6} \geq 2.32635$

$\Rightarrow n \geq 12.18$

Since $n \leq 20$, Set of values of $n = \left\{13 \leq n \leq 20, n \in Z^+\right\}$

Alternatively (by GC – table),

<table>
<tr><th>n</th><th>$\mathrm{P}(\bar{Y} > 24)$</th></tr>
<tr><td>12</td><td>$0.0105 > 0.01$</td></tr>
<tr><td>13</td><td>$0.0081 < 0.01$</td></tr>
<tr><td>14</td><td>$0.0063 < 0.01$</td></tr>
</table>

Since $n \leq 20$, Set of values of $n = \left\{13 \leq n \leq 20, n \in Z^+\right\}$

</td>
<td>

Many mis-read the question and consider the distribution of $T+Y$, instead of $\bar{Y}$.

For those who considered the distribution of $Y_1 + Y_2 + \ldots + Y_n$, $\mathrm{P}(\bar{Y} > 24) \leq 0.01$ is to be interpreted as $\mathrm{P}(Y_1 + Y_2 + \ldots + Y_n > 24n) \leq 0.01$

</td>
</tr>
<tr>
<td>12(d)</td>
<td>

Let $W = 0.8\left(Y_1 + Y_2\right) - 2X$

$\mathrm{E}(W) = 0.8(2)(20) - 2(5) = 22$

$\mathrm{Var}(W) = 0.8^2(2)\left(6^2\right) + 2^2(2.24) = 55.04$

$W \sim \mathrm{N}(22, 55.04)$

$\mathrm{P}(W > 0) = 0.998$ (3 s.f.)

</td>
<td>

Some students factored in the reduced time for the Mean and forgot to do that for the Variance.

<u>Common mistakes</u>

- Writing $W = 0.8(Y + Y) - 2X$
- Writing $\mathrm{Var}(0.8Y) = 0.8\mathrm{Var}(Y)$

</td>
</tr>
</table>