General Comments
• The rubric of the paper states that non-exact numerical answers should be given correct to 3 significant figures. • The use of graphing calculators is encouraged, but in questions where a calculator is prohibited, you need to show sufficient working in answering that question. • Need to be aware that every step shown in a **given answer question** needs to maintain an appropriate level of accuracy. • Read the question carefully.

Section A: Pure Mathematics (40 marks)

1 (a) Show that $\dfrac{1}{\sqrt{1+x^2}-\sqrt{1-x^2}}=\dfrac{1}{2x^2}\left(\sqrt{1+x^2}+\sqrt{1-x^2}\right)$. [1]

1a	$\dfrac{1}{\sqrt{1+x^2}-\sqrt{1-x^2}}\times\dfrac{\sqrt{1+x^2}+\sqrt{1-x^2}}{\sqrt{1+x^2}+\sqrt{1-x^2}}$ $=\dfrac{\sqrt{1+x^2}+\sqrt{1-x^2}}{(1+x^2)-(1-x^2)}$ $=\dfrac{1}{2x^2}\left(\sqrt{1+x^2}+\sqrt{1-x^2}\right)$ (shown)	Mostly well done 👏 Remember that: $(a+b)(a-b)=a^2-b^2$.

(b) Hence use appropriate expansions from the List of Formulae (MF26) to find the first two non-zero terms in the series expansion of $\dfrac{x^2}{\sqrt{1+x^2}-\sqrt{1-x^2}}$ in ascending powers of x for $x\neq 0$. [3]

1b	$\dfrac{x^2}{\sqrt{1+x^2}-\sqrt{1-x^2}}$ $=\dfrac{1}{2}\left[\left(1+x^2\right)^{\frac{1}{2}}+\left(1-x^2\right)^{\frac{1}{2}}\right]$ $=\dfrac{1}{2}\left[1+\dfrac{1}{2}x^2+\dfrac{\frac{1}{2}\left(-\frac{1}{2}\right)}{2!}\left(x^2\right)^2+...\right]+\dfrac{1}{2}\left[1-\dfrac{1}{2}x^2+\dfrac{\frac{1}{2}\left(-\frac{1}{2}\right)}{2!}\left(-x^2\right)^2+...\right]$ $=\dfrac{1}{2}\left[2-\dfrac{1}{4}x^4+...\right]$ $\approx 1-\dfrac{1}{8}x^4$	Mostly well done 👏 Do note that repeated differentiation is <u>not</u> allowed here as question says "use appropriate expansions from MF26".

(c) State the set of values of x for which the series expansion is valid. [1]

1c	$\lvert x^2 \rvert < 1,\ x \neq 0$ $x^2 < 1$ (or draw the graph) $x^2 - 1 < 0$ $(x+1)(x-1) < 0$ $\therefore -1 < x < 1$ Set of values of x: $\{\ x \in \mathbb{R} : -1 < x < 1,\ x \neq 0\ \}$	Learn how to solve inequalities properly and not "jump" the step.

(d) It is given that the two terms found in part **(b)** are equal to the first two terms in the series expansion of $\cos\left(ax^b\right)$. Find the possible value(s) of the constants a and b. [2]

1d	$\cos\left(ax^b\right) \approx 1 - \frac{\left(ax^b\right)^2}{2} \equiv 1 - \frac{1}{8}x^4$ $\frac{a^2}{2} = \frac{1}{8}$ and $2b = 4$ $a = \pm\frac{1}{2}$ and $b = 2$	Quite a few arithmetic errors spotted here, e.g. $\frac{\left(ax^b\right)^2}{2} \neq \frac{ax^{2b}}{2}$. Do practise well to gain fluency 💪

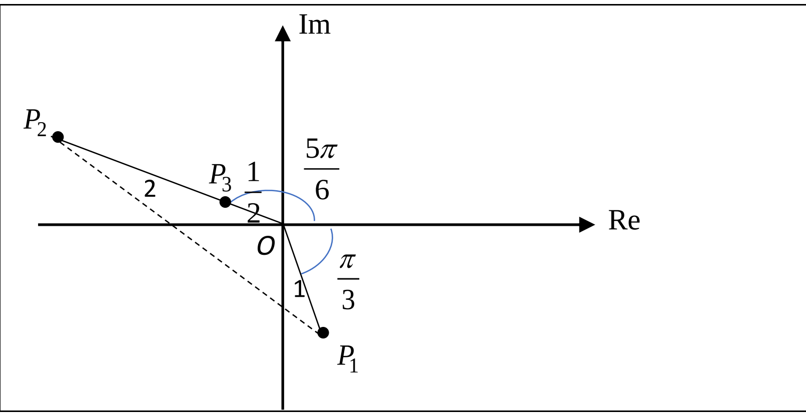
2 Do not use a calculator in answering this question.

The complex numbers z_1, z_2 and z_3 are such that $z_1 = -e^{i\frac{2\pi}{3}}$, $z_2 = -\sqrt{3} + i$ and $z_3 = \dfrac{z_1}{z_2}$.

(a) Express each of z_1, z_2 and z_3 in the form $re^{i\theta}$, where $r > 0$ and $-\pi < \theta \le \pi$. [3]

2a	$z_1 = -e^{i\frac{2\pi}{3}}$ $= e^{i\pi}.e^{i\frac{2\pi}{3}}$ $= e^{i\frac{5\pi}{3}}$ $= e^{i\left(-\frac{\pi}{3}\right)}$ $z_2 = -\sqrt{3} + i$ $= 2e^{i\left(\pi - \tan^{-1}\frac{1}{\sqrt{3}}\right)}$ $= 2e^{i\left(\pi - \frac{\pi}{6}\right)}$ $= 2e^{i\frac{5\pi}{6}}$ $z_3 = \dfrac{z_1}{z_2}$ $= \dfrac{e^{i\left(-\frac{\pi}{3}\right)}}{2e^{i\frac{5\pi}{6}}}$ $= \frac{1}{2}e^{i\left(-\frac{\pi}{3} - \frac{5\pi}{6}\right)}$ $= \frac{1}{2}e^{i\left(-\frac{7\pi}{6}\right)}$ $= \frac{1}{2}e^{i\frac{5\pi}{6}}$	Polar form for complex numbers was poorly performed in general. Do note that $-1 = e^{i\pi}$ (i.e. modulus = 1, argument = π. When in doubt, please plot the complex number as a point on an argand diagram – this will help you to better determine the correct modulus and argument.

(b) Sketch an Argand diagram showing the points P_1, P_2 and P_3 where P_1, P_2 and P_3 represent the complex numbers z_1, z_2 and z_3 respectively. [2]

2b	 	This follows from (a). Do note that you are required to indicate the modulus in addition to the argument i.e. distance from the origin. Also, ensure that the relative positioning of the 3 points is correct.

(c) Find the area of triangle OP_1P_2. [2]

2c	Area of triangle OP_1P_2 $=\frac{1}{2}(1)(2)\sin\left(\frac{\pi}{6}+\frac{\pi}{2}+\frac{\pi}{6}\right)$ $=\sin\frac{5\pi}{6}$ $=0.5$	For those who have done (a) and (b) correctly, this was well done and students were able to apply the correct formula $\frac{1}{2}ab(\sin c)$.

(d) Find the smallest positive integer n for which $\left(z_2^*\right)^n$ is purely imaginary. [2]

2d	$\left(z_2^*\right)^n=\left(2e^{i\left(-\frac{5\pi}{6}\right)}\right)^n$ $=2^n e^{i\left(-\frac{5n\pi}{6}\right)}$ For $\left(z_2^*\right)^n$ to be purely imaginary, $-\frac{5n\pi}{6}=\frac{\pi}{2},-\frac{\pi}{2},\frac{3\pi}{2},-\frac{3\pi}{2},\frac{5\pi}{2},-\frac{5\pi}{2},...$ Smallest positive integer $n=3$	Most students were able to translate "purely imaginary" to the argument being an odd multiple of $\frac{\pi}{2}$. So, either write it in a general form e.g. $(2k+1)\frac{\pi}{2}$ or $(2k-1)\frac{\pi}{2}$, or since the question is asking for the smallest n, you can also list out the first few (negative) odd multiples of $\frac{\pi}{2}$ and check which value of n works.

3 The line l_1 has equation $\mathbf{r} = 3\mathbf{i} - 4\mathbf{j} - 5\mathbf{k} + \lambda(\mathbf{i} - 2\mathbf{j} - \mathbf{k})$, where λ is a real parameter. The point A has position vector $\mathbf{i} - 2\mathbf{j} + \mathbf{k}$.

(a) The plane p contains the line l_1 and the point A. Find a cartesian equation of the plane p. [3]

3a	$\underset{\sim}{r} = \begin{pmatrix} 3 \\ -4 \\ -5 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}, \lambda \in \mathbb{R}$; $\overrightarrow{OA} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$ Vector parallel to $p = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} - \begin{pmatrix} 3 \\ -4 \\ -5 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \\ 6 \end{pmatrix} = 2\begin{pmatrix} -1 \\ 1 \\ 3 \end{pmatrix}$ Normal vector $= \begin{pmatrix} -1 \\ 1 \\ 3 \end{pmatrix} \times \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix}$ Equation of p $\underset{\sim}{r} \bullet \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \bullet \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix} \Rightarrow \underset{\sim}{r} \bullet \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix} = 2$ Cartesian equation of p: $5x + 2y + z = 2$	Make sure you copy the vector correctly and not make any silly mistakes at the start 🤦‍♀️ The origin O may not be in plane p, hence you cannot assume that $\overrightarrow{OA}$ is a vector in the plane (as it turns out, O is not in the plane). Also, always check that the normal of the plane you have obtained after cross product is correct. A quick way to check is via dot product: $\begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \\ 3 \end{pmatrix} = ... = 0$ and $\begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix} = ... = 0$. (since the normal must be perpendicular to the 2 vectors used in the cross product.) Finally, if your normal is $\begin{pmatrix} 10 \\ 4 \\ 2 \end{pmatrix}$, it is a good idea to reduce it to $\begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix}$ first before finding the cartesian equation of the plane (so that the equation of the plane can be in the "reduced" form).

(b) Find the position vector of the point A', the reflection of the point A in the line l_1. [4]

3b	Let F be the foot of the perpendicular from A to the line l_1, and B be the point on l_1 with position vector $3\mathbf{i}-4\mathbf{j}-5\mathbf{k}$ Let $\overrightarrow{BA}=\begin{pmatrix}1\\-2\\1\end{pmatrix}-\begin{pmatrix}3\\-4\\-5\end{pmatrix}=\begin{pmatrix}-2\\2\\6\end{pmatrix}$ $\overrightarrow{BF}=\left[\begin{pmatrix}-2\\2\\6\end{pmatrix}\bullet\frac{1}{\sqrt{6}}\begin{pmatrix}1\\-2\\-1\end{pmatrix}\right]\frac{1}{\sqrt{6}}\begin{pmatrix}1\\-2\\-1\end{pmatrix}=-2\begin{pmatrix}1\\-2\\-1\end{pmatrix}$ $\overrightarrow{OF}=\overrightarrow{OB}+\overrightarrow{BF}=\begin{pmatrix}3\\-4\\-5\end{pmatrix}-2\begin{pmatrix}1\\-2\\-1\end{pmatrix}=\begin{pmatrix}1\\0\\-3\end{pmatrix}$ $\overrightarrow{OF}=\frac{\overrightarrow{OA}+\overrightarrow{OA'}}{2}\Rightarrow\overrightarrow{OA'}=2\overrightarrow{OF}-\overrightarrow{OA}$ $\overrightarrow{OA'}=2\begin{pmatrix}1\\0\\-3\end{pmatrix}-\begin{pmatrix}1\\-2\\1\end{pmatrix}=\begin{pmatrix}1\\2\\-7\end{pmatrix}$	Most students were able to apply the method of finding the projection vector or finding the foot of the perpendicular but made several errors, e.g. For $\begin{pmatrix}-2\\2\\6\end{pmatrix}$, you will have to use $\begin{pmatrix}-2\\2\\6\end{pmatrix}$ and not $\begin{pmatrix}-1\\1\\3\end{pmatrix}$ which is only half of $\overrightarrow{BA}$ (contrast this with the idea of "reducing" the normal). Almost all students were familiar with applying the midpoint theorem which is good 💪

(c) The plane q is such that q is parallel to p and passes through the point with position vector $-3\mathbf{j}+\mathbf{k}$. Find a cartesian equation of q and the exact shortest distance between p and q. [3]

3c	$5x+2y+z=k$ Sub $-3\mathbf{j}+\mathbf{k}$ into the equation: $5(0)+2(-3)+1=k\Rightarrow k=-5$ Cartesian equation of q: $5x+2y+z=-5$ Exact shortest distance between p and q $=\left\|\left[\begin{pmatrix}0\\-3\\1\end{pmatrix}-\begin{pmatrix}1\\-2\\1\end{pmatrix}\right]\cdot\begin{pmatrix}5\\2\\1\end{pmatrix}\right\|\frac{1}{\sqrt{25+4+1}}$ $=\left\|\begin{pmatrix}-1\\-1\\0\end{pmatrix}\cdot\begin{pmatrix}5\\2\\1\end{pmatrix}\right\|\frac{1}{\sqrt{30}}$ $=\frac{7}{\sqrt{30}}$ **<u>Alternative</u>** Exact shortest distance between p and q $=\frac{5+2}{\sqrt{5^2+2^2+1^2}}=\frac{7}{\sqrt{30}}$	Always draw a simple diagram (if you need) to help you to determine the points to be used in the planes and whether dot or cross product should be used to find the required distance.

(d) The line l_2 has the equation $\frac{y-3}{2}=\frac{z-7}{3}$, $x=2$. Given that l_2 intersects p at point S, find the area of the triangle OAS. [4]

3d	Substitute $\underset{\sim}{r}=\begin{pmatrix}2\\3\\7\end{pmatrix}+\mu\begin{pmatrix}0\\2\\3\end{pmatrix}$ into $\underset{\sim}{r}\bullet\begin{pmatrix}5\\2\\1\end{pmatrix}=2$ $\left[\begin{pmatrix}2\\3\\7\end{pmatrix}+\mu\begin{pmatrix}0\\2\\3\end{pmatrix}\right]\bullet\begin{pmatrix}5\\2\\1\end{pmatrix}=2$ $(10+6+7)+(4+3)\mu=2$ $\mu=-3$ $\overrightarrow{OS}=\begin{pmatrix}2\\3\\7\end{pmatrix}+(-3)\begin{pmatrix}0\\2\\3\end{pmatrix}=\begin{pmatrix}2\\-3\\-2\end{pmatrix}$ Area of triangle OAS $=\frac{1}{2}\left\|\begin{pmatrix}2\\-3\\-2\end{pmatrix}\times\begin{pmatrix}1\\-2\\1\end{pmatrix}\right\|$ $=\frac{1}{2}\left\|\begin{pmatrix}-7\\-4\\-1\end{pmatrix}\right\|=\frac{\sqrt{66}}{2}$	Do note that it is **wrong** to write: $l_2=\begin{pmatrix}2\\3\\7\end{pmatrix}+\mu\begin{pmatrix}0\\2\\3\end{pmatrix}$ (Why?) Many students were also unable to convert the cartesian equation of l_2 to vector equation correctly, resulting in the wrong $\overrightarrow{OS}$ that was found. Please learn well from now 🙏

4 The curve C is defined by the parametric equations

$$x=a\left(1+\frac{1}{t}\right) \quad \text{and} \quad y=a\left(t-\frac{1}{t^2}\right)$$

where a is a positive constant and $t\neq 0$.

(a) Show that $\frac{\mathrm{d}y}{\mathrm{d}x}=-\left(\frac{2+t^3}{t}\right)$. [3]

4a	$\frac{\mathrm{d}x}{\mathrm{d}t}=-a\left(\frac{1}{t^2}\right)$ $\frac{\mathrm{d}y}{\mathrm{d}t}=a\left(1+\frac{2}{t^3}\right)$ $\frac{\mathrm{d}y}{\mathrm{d}x}=a\left(1+\frac{2}{t^3}\right)\div\left(-a\left(\frac{1}{t^2}\right)\right)$ $=\left(\frac{t^3+2}{t^3}\right)\times\left(-t^2\right)$ $=-\left(\frac{2+t^3}{t}\right)$	Mostly well done 👏

(b) Find, in terms of a, the coordinates of the turning point on C, and explain why it is a maximum.

[4]

4b

At stationary point, $\frac{\mathrm{d}y}{\mathrm{d}x} = 0 \Rightarrow \frac{2+t^3}{t} = 0$

$t^3 = -2 \Rightarrow t = -\sqrt[3]{2}$

$x = a\left(1 - \frac{1}{\sqrt[3]{2}}\right) = 0.2062995a$

$y = a\left(-\sqrt[3]{2} - \frac{1}{2^{\frac{2}{3}}}\right) = -1.88988a$

Coordinates of turning point are $(0.206a, -1.89a)$.

To determine that it is a max turning point, use first derivative sign test.

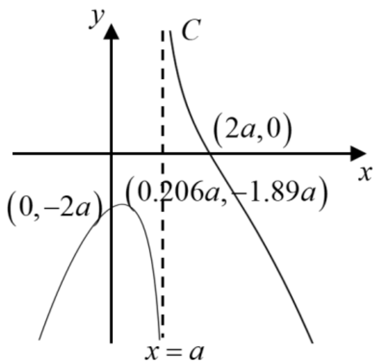
t	$t = -1.25$	$t = -\sqrt[3]{2} = -1.25992$	$t = -1.27$
x	$x = 0.2a$ (left)	$x = a\left(1 - \frac{1}{\sqrt[3]{2}}\right) = 0.206a$	$x = 0.213a$ (right)
$\frac{\mathrm{d}y}{\mathrm{d}x}$	0.0375	0	–0.0381
Sign of $\frac{\mathrm{d}y}{\mathrm{d}x}$	Positive	Zero	Negative

Therefore, turning point is a maximum.

Since there is no requirement to give the coordinates in exact form, please give the answer as $(0.206a, -1.89a)$ as it will help you to position the turning point when sketching the graph in (c). (Note a is positive.)

Also, do note that finding the second derivative for parametric equations is <u>out of the syllabus</u> – i.e. it is not simply to differentiate $\frac{\mathrm{d}y}{\mathrm{d}x}$ once more with respect to x (i.e. chain rule is required). Hence, for explaining why the turning point is a maximum, you will have to perform the first derivative test, showing <u>a table of values of x and t</u> that were used and the corresponding values of $\frac{\mathrm{d}y}{\mathrm{d}x}$ that were found.

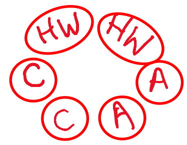
(c) Sketch C. [3]

4c	 Asymptote at $x=a$: As $t\to+\infty,\ x\to a^{+},\ y\to+\infty$ As $t\to-\infty,\ x\to a^{-},\ y\to-\infty$ Axial intercepts: At $x=0,\ t=-1,\ y=-2a$ At $y=0,\ t=1,\ x=2a$	Do not forget to adjust t to include negative values. If you forget, you should realise something is not right, as you would be missing the maximum turning point that was mentioned in (b). Please learn how to find the asymptote and axial intercepts and not just rely on the GC. The GC may give you the general shape, but not the details that you may need especially when an unknown a is present in the parametric equations.

Section B: Statistics (60 marks)

5 Two married couples, two single adults and two children form a team of 8 to take part in a series of games.

(a) In the first game, the team sits in a circle. Find the number of arrangements that can be formed if each married couple must be seated together. [2]

5a	N (team seated in a circle & each couple together) $= (6-1)! \times 2! \times 2!$ $= 480$  	2 couples – hence do not forget it is $2! \times 2!$

(b) A group of three people are to be selected from the team for the second game. Find the number of different groups that can be formed if there must **not** be a married couple in the group. [2]

5b	N (select group of 3, husband & wife cannot both be selected) = N (no restriction) – N (1 couple & 1 other person) $= \binom{8}{3} - \binom{2}{1} \times \binom{6}{1}$ $= 44$ **<u>Alternative</u>** Case 1: N(no married person) = $\binom{4}{3} = 4$ Case 2: N(1 married person) = $\binom{4}{1} \times \binom{4}{2} = 24$ Case 3: N(1 married person from each couple) $= 2 \times \binom{2}{1} \times \binom{4}{1} = 16$ N (group of 3, husband & wife cannot both be selected) $= 4+24+16 = 44$	A common mistake for those who did by the complement method was: $\binom{8}{3} - \binom{2}{1} \times \binom{4}{1}$ resulting in the wrong answer.

(c) In the third game, each team member selects a unique number from the set $\{1, 2, \ldots, 8\}$. Find the number of different ways this can be done if the numbers selected by the children are both greater than the numbers selected by the two single adults. [2]

5c	Number of ways $= {}^{8}C_{4} \times 2! \times 2! \times 4!$ $= 6720$	This was meant to be the differentiating question so kudos to the few who got it! As for the majority of you, fret not, there are still 98 marks in the paper to be earned 💪 So stay calm and focused in the A levels and not be thrown off.

2024 VJC Prelim Paper 2 Solutions

6 A random variable X has the probability distribution given in the following table.

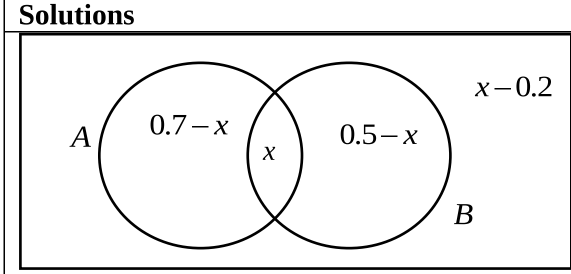
x	1	4	6	8
$\mathrm{P}(X=x)$	a	b	c	d

Given that $\mathrm{E}(X)=4$, $\mathrm{Var}(X)=\frac{19}{4}$ and $\mathrm{P}(X<4)=\mathrm{P}(X>4)$, find the values of a, b, c and d. [5]

6	$\sum_{\text{all } r}\mathrm{P}(X=r)=1$ $a+b+c+d=1 \quad ---(1)$ $\mathrm{P}(X<4)=\mathrm{P}(X>4)$ $a=c+d$ $a-c-d=0 \quad ---(2)$ $\mathrm{E}(X)=4$ $a+4b+6c+8d=4 \quad ---(3)$ $\mathrm{Var}(X)=\frac{19}{4}$ $a+16b+36c+64d-4^2=\frac{19}{4}$ $a+16b+36c+64d=\frac{83}{4} \quad ---(4)$ Solving, $a=\frac{1}{4}$, $b=\frac{1}{2}$, $c=\frac{1}{8}$ and $d=\frac{1}{8}$	Mostly well done 👏 Do note that X is a discrete random variable here, i.e. it only takes on possible values of 1, 4, 6 and 8. Some students mistook or assumed X to be normal which is not true. Also, always remember that if you have 4 variables to solve completely (i.e. a, b, c and d), you will need (at least) 4 equations. So don't forget the equation that comes from the sum of the probabilities is 1.

7 For events A, B and C, it is given that $\mathrm{P}(A)=0.7$, $\mathrm{P}(B)=0.5$, $\mathrm{P}(C \mid A')=0.6$ and $\mathrm{P}(A \mid C')=0.76$.

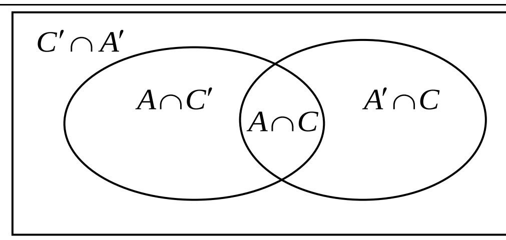
(a) Find the greatest and least possible values of $\mathrm{P}(A\cap B)$. [2]

Solutions	**Comments**
 Let $\mathrm{P}(A\cap B)=x$. $0\le x-0.2\le 1 \quad \Rightarrow 0.2\le x\le 1$ $0\le 0.7-x\le 1 \quad \Rightarrow 0\le x\le 0.7$ $0\le 0.5-x\le 1 \quad \Rightarrow 0\le x\le 0.5$ Hence, greatest and least values of $\mathrm{P}(A\cap B)$ are 0.5 and 0.2 respectively.	Since set C is not involved in this part, Venn diagram drawn includes only sets A and B.

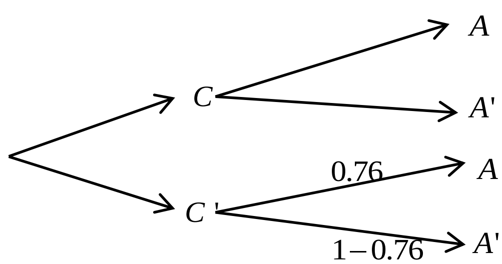
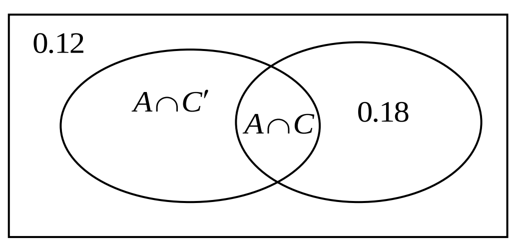
(b) Find $\mathrm{P}(C \cap A')$. [1]

Solutions	Comments
$\mathrm{P}(C \cap A') = \mathrm{P}(C \mid A') \times \mathrm{P}(A')$ $= (0.6)(1-0.7) = 0.18$	Mostly well done Apply conditional probability $\mathrm{P}(C \mid A') = \dfrac{\mathrm{P}(C \cap A')}{\mathrm{P}(A')}$

(c) Find $\mathrm{P}(C' \cap A')$. [2]

Solutions	Comments
$\mathrm{P}(C' \cap A') = 1 - \mathrm{P}(A \cup C)$ $= 1 - [\mathrm{P}(A) + \mathrm{P}(C \cap A')]$ $= 1 - (0.7 + 0.18)$ $= 0.12$	 No need to include set *B* in the Venn diagram

(d) Find $\mathrm{P}(C)$. [3]

Solutions	Comments
$\mathrm{P}(A \mid C') = 0.76$ $\mathrm{P}(A' \mid C') = 1 - 0.76 = 0.24$ $\dfrac{\mathrm{P}(A' \cap C')}{\mathrm{P}(C')} = 0.24$ $\mathrm{P}(C') = \dfrac{0.24}{\mathrm{P}(A' \cap C')} = \dfrac{0.24}{0.12} = 0.5$ $\mathrm{P}(C) = 1 - 0.5 = 0.5$ **<u>Alternative</u>** $\mathrm{P}(A \mid C') = 0.76$ $\dfrac{\mathrm{P}(A \cap C')}{\mathrm{P}(C')} = 0.76$ $\mathrm{P}(A \cap C') = 0.76[1 - \mathrm{P}(C)]$ $\mathrm{P}(A \cap C') + 0.76\mathrm{P}(C) = 0.76 \quad ---(1)$ $\mathrm{P}(A \cup C) = 0.88$ $\mathrm{P}(A \cap C') + \mathrm{P}(C) = 0.88 \quad ---(2)$ Solving (1) & (2): $\mathrm{P}(C) = 0.5$	 **<u>Alternative</u>** 

8 A small company makes wine glasses. Each day, n randomly cosen wine glasses are checked and the number of wine glasses found to be cracked is denoted by X.

(a) State, in context of the question, two assumptions needed for X to be well modelled by a binomial distribution. [2]

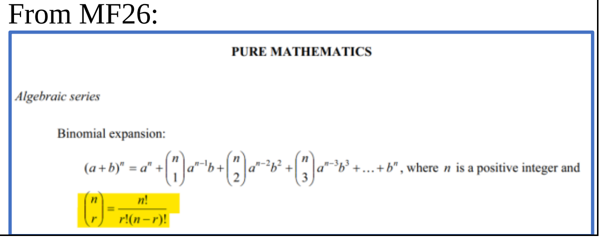
Solutions	**Comments**
The assumptions are 1. Cracked wine glasses occur independently of one 2. The probability that a wine glass is cracked remains constant.	These are **INCORRECT** statements • wine glasses are independent of each other • probability (number) of cracked wine glass is independent of each other • probability that a wine glass is cracked is constant for **all *n* wine glasses (OR each day)** • selecting / choosing / finding/ getting a cracked wine glass is independent.

Assume now that X has the distribution $\mathrm{B}(n,\ p)$, where $n \geq 3$

(b) Given that the mean of X and the variance of X are 1.8 and 1.773 respectively, find the value of n and the value of p. [2]

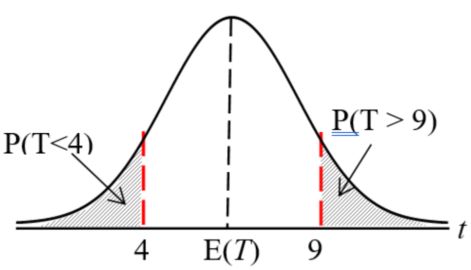
Solutions	**Comments**
$X \sim \mathrm{B}(n,\ p)$ $\mathrm{E}(X)=1.8 \qquad \mathrm{Var}(X)=1.773$ $np=1.8 \quad ---(1) \qquad np(1-p)=1.773 \quad ---(2)$ $\frac{(2)}{(1)}: \frac{np(1-p)}{np}=\frac{1.773}{1.8}=0.985$ $1-p=0.985$ $p=0.015$ $n=\frac{1.8}{0.015}=120$	Mostly well done $p = 0.015$ is exact and should be left as such.

(c) Given instead that the probability of finding 2 cracked wine glasses is thrice the probability of finding 3 cracked wine glasses, find p in terms of n. [2]

Solutions	**Comments**
$X \sim \mathrm{B}(n,\ p)$	From MF26:  PURE MATHEMATICS *Algebraic series* Binomial expansion: $(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \binom{n}{3}a^{n-3}b^3 + \ldots + b^n$, where n is a positive integer and $\binom{n}{r} = \frac{n!}{r!(n-r)!}$

$P(X=2)=3P(X=3)$ $\binom{n}{2}p^2(1-p)^{n-2}=3\binom{n}{3}p^3(1-p)^{n-3}$ $\frac{n(n-1)}{2}p^2(1-p)^{n-2}=\frac{3n(n-1)(n-2)}{3!}p^3(1-p)^{n-3}$ Since $p>0,\ 1-p>0,\ n>0$ and $n-1>0$ $\frac{1-p}{2}=\frac{3(n-2)p}{6}$ $(n-2+1)p=1$ $p=\frac{1}{n-1}$	

9 **(a)** S and T are independent random variables with the distributions $N(18,3^2)$ and $N(\mu,\sigma^2)$ respectively. It is given that $P(T<4)=P(T>9)$ and $P(S<3T)=0.65$. Calculate the values of μ and σ. [4]

Solutions	**Comments**
Since $P(T<4)=P(T>9)$, by symmetry, $\mu=\frac{4+9}{2}=6.5$ $E(S-3T)=E(S)-3E(T)=18-3\mu=-1.5$ $\text{Var}(S-3T)=\text{Var}(S)+3^2\text{Var}(T)=3^2+3^2\sigma^2=9+9\sigma^2$ $S-3T\sim N(-1.5,9+9\sigma^2)$ $P(S<3T)=0.65$ $P(S-3T<0)=0.65$ $P\left(Z<\frac{0-(-1.5)}{\sqrt{9+9\sigma^2}}\right)=0.65$ From GC, $P(Z<0.38532)=0.65$ $\frac{1.5}{\sqrt{9+9\sigma^2}}=0.38532$ Solving, $\sigma=0.82694\approx0.827$	

(b) A fruit stall sells grapes that are packed in packets with masses in grams that follow the distribution $\mathrm{N}\left(850, 30^2\right)$. The grapes are sold at \$18 per kilogram.

(i) Find the probability that a customer pays more than \$30 for two packets of grapes. [2]

Solutions	Comments
Let X g be the mass of a packet of grapes. $X \sim \mathrm{N}\left(850, 30^2\right)$ $\therefore X_1 + X_2 \sim \mathrm{N}(1700, 1800)$ $\mathrm{P}\left(\frac{18}{1000}(X_1 + X_2) > 30\right) = \mathrm{P}\left(X_1 + X_2 > \frac{30000}{18}\right)$ $= 0.78397$ ≈ 0.784 **<u>Alternative</u>** Let Y be the total cost of 2 packets of grapes $\therefore Y = \frac{18}{1000}(X_1 + X_2)$ $\mathrm{E}(Y) = \frac{18}{1000}\left[2\mathrm{E}(X)\right] = 30.6$ $\mathrm{Var}(Y) = \left(\frac{18}{1000}\right)^2\left[2\mathrm{Var}(X)\right] = 0.5832$ $\therefore Y \sim \mathrm{N}(30.6, 0.5832)$ $\mathrm{P}(Y > 30) = 0.78397 \approx 0.784$	<u>Selling price:</u> \$18 per kg → \$ $\frac{18}{1000}$ per g

(ii) The fruit stall accepts payment by cash or PayNow. The number of customers who pay by PayNow in a day is a random variable with mean 12 and variance 4.8. In a month of 30 days, find the probability that the average number of customers per day who pay by PayNow is more than 12.3. [3]

Solutions	Comments
Let Y be the number of customers who pay by PayNow in a day. $\mathrm{E}(Y) = 12, \quad \mathrm{Var}(Y) = 4.8$ Then $\overline{Y} = \frac{Y_1 + Y_2 + \ldots + Y_{30}}{30}$ is the average number of customers per day who pay by PayNow. By Central Limit Theorem, $\overline{Y} \sim \mathrm{N}\left(12, \frac{4.8}{30}\right)$ approx . $\mathrm{P}\left(\overline{Y} > 12.3\right) = 0.22663$ ≈ 0.227	**<u>Incorrect to assume</u>** $Y \sim \mathrm{N}(12, 4.8)$ OR $Y \sim \mathrm{B}(n, p)$ It is necessary to state "by Central Limit Theorem, $\overline{Y} \sim \mathrm{N}\left(12, \frac{4.8}{30}\right)$ approx "

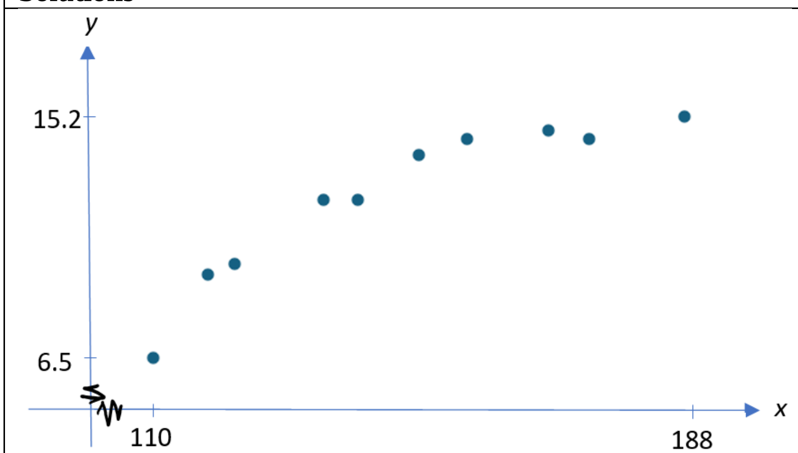
10 The yield per hectare, y kg, of a crop is believed to depend on the average rainfall, x mm, in the month of June. For 10 regions, records are kept of the values of x and y, and these are shown in the table below. The yield from the tenth region was accidentally deleted from the records after the data was analysed, and this is indicated by the value p.

Average rainfall (x mm)	149	110	188	135	156	140	168	118	122	174
Yield of crop (y kg)	13.8	6.5	15.2	12.2	14.4	12.2	14.7	9.5	9.9	p

Given that the equation of the regression line of y on x is $y=-2.5652+0.10168x$, show that $p=14.4$. [2]

Solutions	Comments
$\sum x=1460$, $\sum y=108.4+p$, $n=10$ Since $(\bar{x},\bar{y})$ lies on the regression line of y on x, $\bar{y}=-2.5652+0.10168\bar{x}$ $\frac{108.4+p}{10}=-2.5652+0.10168\frac{1460}{10}$ $=12.28008$ $p=14.4008$ ≈ 14.4 (shown)	$(174, p)$ does not lie on the regression line, hence you will get an **<u>approximated</u>** value for p when you substitute $x=174$ into the regression line. "**Show**" → you need to give at least the 5 s.f value of $p=14.401$ before concluding that $p=14.4$.

(a) Draw a scatter diagram for these values, labelling the axes clearly. Calculate the product moment correlation coefficient between x and y. [2]

Solutions	Comments
 From GC, product moment correlation coefficient, $r=0.92147\approx 0.921$	☐ Label minimum and maximum x and y values. ☐ Check that the 4th and 5th points are level as they have the same y-value. ☐ Make sure you have **<u>10</u>** points sketched in your scatter diagram.

(b) It is thought that a model of the form $y = a + b\ln x$ may also be a suitable fit to the data. Calculate least square estimates of a and b, and find the value of the product moment correlation coefficient between y and $\ln x$. [3]

Solutions	**Comments**
From GC, the regression line of y on $\ln x$ is $y = -63.028 + 15.154\ln x$ $a = -63.028 \approx -63.0$ $b = 15.154 \approx 15.2$ $r = 0.94660 \approx 0.947$	The rubric of the paper states that non-exact numerical answers should be given correct to 3 significant figures.

(c) Use your answers to parts **(a)** and **(b)** to explain which of

Need to comment on (1) scatter diagram in (a) & (2) compare both r values

$$y = -2.5652 + 0.10168x \quad \text{or} \quad y = a + b\ln x$$

is the better model. [2]

Solutions	**Comments**
1) From the scatter diagram, it is observed that as x increases, y increases by decreasing amounts, and 2) product moment correlation coefficient between y and $\ln x$ is 0.947, which is closer to 1 than that of x and y, which is 0.921 . Hence $y = a + b\ln x$ is the better model.	□ Describe how your y values change as x increases as seen in the scatter diagram in (a) □ You should compare which of the two r values is **<u>closer to 1</u>**.

(d) Using an appropriate regression line, estimate the yield for a region that experienced 200 mm of rainfall in June. Comment on the reliability of your estimate. [2]

Solutions	**Comments**
$y = -63.028 + 15.154\ln 200$ $= 17.263$ ≈ 17.3 Since $x = 200$ lies outside the given range of x values, $110 \le x \le 188$, the estimated yield may not be reliable.	Remember to note down the specific range of x i.e. state $110 \le x \le 188$

(e) In some regions, rainfall is measured in inches instead of in mm. Given that there are 25.4 mm in an inch, show how the regression line found in part **(b)** can be re-written so that it can be used when x, the average rainfall in June, is given in inches. [1]

Solutions	**Comments**
Replace x by $25.4x$, $y = -63.028 + 15.154\ln(25.4x)$ $= -63.028 + 15.154[\ln 25.4 + \ln x]$ $= -14.009 + 15.154\ln x$ $= -14.0 + 15.2\ln x$	All x values (in mm) need to be multiply by $\frac{1}{25.4}$ to convert to inches $\Rightarrow$ i.e. stretch // x-axis by factor $\frac{1}{25.4}$ $\Rightarrow$ replace x by $25.4x$. Answer has to be simplified.

11 In the swimming training school AquaV, the time taken to swim a lap of the pool by the trainees is found to have a mean of 35 seconds. The school adopted a new international training programme Breakthru for 3 months and wanted to analyse if Breakthru is effective in improving the timings of the trainees.

A sample of 30 trainees is taken and the times taken, x seconds, to swim a lap of the pool by the trainees are summarised by

$$\sum(x-30)=94, \quad \sum(x-30)^2=758.$$

(a) Test, at the 5% significance level, whether there is any evidence that the mean time taken to swim a lap of the pool has improved after the trainees underwent 3 months of Breakthru, defining any parameters you use. [7]

i.e. swim faster → $H_1: \mu < 35$

Need to define μ

Solutions	**Comments**
Let μ be the population mean time taken by the trainees to swim a lap. $H_0: \mu = 35$ $H_1: \mu < 35$ Level of significance: 5 % Test Statistic: Since n = 30 is large, by Central Limit Theorem, $\bar{X}$ is approximately normally distributed. When H_0 is true, $Z=\dfrac{\bar{X}-35}{\frac{S}{\sqrt{n}}} \sim N(0,1)$ approximately Computation: $n=30$, $\bar{x}=\dfrac{\sum(x-30)}{30}+30=\dfrac{94}{30}+30=33\dfrac{2}{15}\approx 33.133$ $s^2=\dfrac{1}{29}\left[\sum(x-30)^2-\dfrac{\left(\sum(x-30)\right)^2}{30}\right]$ $=\dfrac{1}{29}\left[758-\dfrac{94^2}{30}\right]=15.982$ From GC, p-value = 0.00526 Since p-value = 0.00526 < 0.05, H_0 is rejected at 5% significance level. There is sufficient evidence that the trainees' mean time taken to swim a lap has improved.	When carrying out a hypothesis test, need to write down Step 1: Definition of μ, Step 2: Correct Hypotheses statements Step 3: Test Statistics Step 4: computation of $\bar{x}$, s^2 (because population variance is not known) and p-value. Step 5: Conclusion 1. Compare p-value to sig. level → H_0 rejected (or not rejected) at the level of significance. 2. There is sufficient (or insufficient) evidence that mean time has improved.

(b) State an assumption used in carrying out the test. [1]

Solutions	Comments
Assumption: The sample of 30 trainees taken is a random sample. OR Assumption: The sample of 30 trainees taken is such that every trainee has an equal probability of being selected for the sample **and** each trainee is selected independently.	Need a **random** sample to carry out a hypothesis test.

In another swimming training school AquaZ, the time taken to swim a lap of the pool by the trainees is normally distributed with a mean of 38 seconds. AquaZ similarly adopted Breakthru for 3 months and then also carried out a test at the 5% significance level to determine whether there is an improvement in the swimming times. A sample of 30 trainees was taken and their timings were measured. The sample standard deviation was found to be 4 seconds and the mean time was denoted by $\bar{x}$

sample variance $= 4^2$

(c) Find the set of values of $\bar{x}$ for which the result of the test would be to reject the null hypothesis. [4]

Solutions	Comments
$H_0: \mu = 38$ $H_1: \mu < 38$ Level of significance: 5 % Test Statistic: When H_0 is true, $Z = \dfrac{\bar{X} - 38}{\frac{S}{\sqrt{n}}} \sim N(0,1)$ approximately Computation: $s^2 = \dfrac{n}{n-1}[\text{sample variance}] = \dfrac{30}{29}\left[4^2\right] = 16.552$ Rejection region: $z \le -1.64485$ Since H_0 is rejected $\Rightarrow z - \text{calculated} \le -1.64485$ $\dfrac{\bar{x} - 38}{\frac{\sqrt{16.552}}{\sqrt{30}}} \le -1.64485$ $\bar{x} \le 38 + \dfrac{\sqrt{16.552}}{\sqrt{30}}(-1.64485)$ $\bar{x} \le 36.778$ Set of values of $\bar{x}$: $\{\bar{x} \in \mathbb{R} : \bar{x} \le 36.8\}$	Note the difference between - population variance (σ^2) - unbiased estimate of population variance (s^2) - sample variance

(d) If the times taken by the 30 trainees is summarised by $\sum(x-30)=234$, determine the conclusion of the test. [2]

Solutions	**Comments**
$\bar{x}=\frac{234}{30}+30=37.8$ Since 37.8 > 36.778, from result in **(c)**, H_0 is not rejected at 5% significance level. There is insufficient evidence that the trainees' mean time taken to swim a lap has improved.	