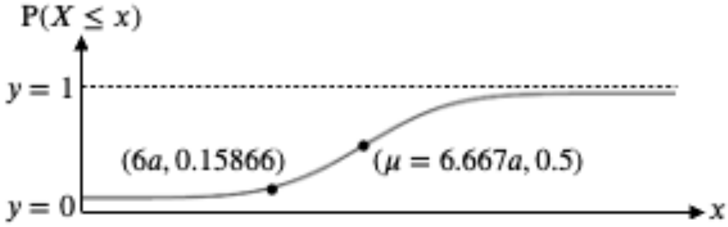
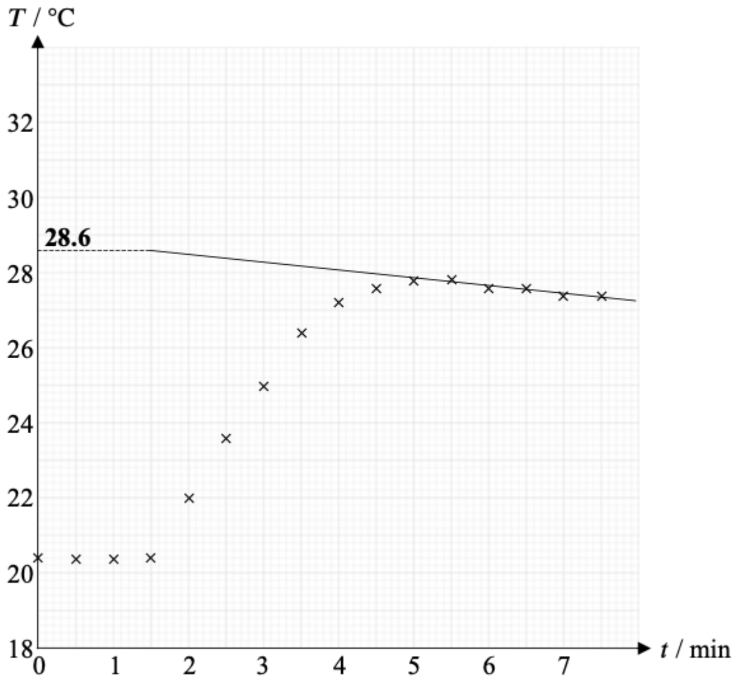
MATHEMATICS

Paper 9758/02
Set I – Paper 2

Topic identification and short answers

Qn	Topic(s)	Part	Answers
			Section A: Pure Mathematics [40 marks]
1	Vectors (two dimensions); Graphs and transformations		Required sequence of transformation: 1st: Scaling by a scale factor of k^2 parallel to the $x-$axis 2nd: Translation of $k^2 - 1 - \frac{1}{k} + k$ units in the positive $y-$direction
2	Graphs (sketching using GC); Inequalities		• If $0 < a < 1$, then $-1 < x \le -a$ or $a \le x < 2$. • If $a = 1$, then $1 \le x < 2$. • If $1 < a < 2$, then $-a \le x < -1$ or $a \le x < 2$. • If $a = 2$, then $-2 \le x < -1$. • If $a > 2$, then $-a \le x < -1$ or $2 < x \le a$.
3	Differentiation (maxima and minima)		$h = \sqrt{3}r$ gives minimum volume. Minimum volume $= 4\sqrt{3}r^3$ units^3
4	Complex numbers (cartesian form, conjugate)	**(a)**	[shown]
		(b)	$y = 1 + 2x + \left(2 - \frac{a^2}{2}\right)x^2 + \left(\frac{4}{3} - a^2\right)x^3 + \cdots$
		(c)	$e^{2x} \sin ax = ax + 2ax^2 + \cdots$
5	Sequences and series (geometric series)	**(a)**	End of July 2026
		(b)	[shown] Amount $= \$9,477.56$
		(c)	August 2029

Section B: Probability and Statistics [60 marks]			
6	Normal distribution	**(a)**	$\mu = 6.667a$; $\sigma = 0.667a$
		(b)	
7	Binomial distribution	**(a)**	[shown]
		(b)	$p \approx 0.968$ or 0.252
		(c)	$\frac{1}{2} < p < 1$
8	Discrete random variables	**(a)**	[shown]
		(b)	$\mathrm{E}(X) = \dfrac{4N^2 + 3N - 1}{6N}$
		(c)	[shown] $m = 71$
9	Probability	**(a)**	$\mathrm{P}(B) = \frac{2}{5}$; $\mathrm{P}(A \cap B) = \frac{7}{30}$
		(b)	[shown]
		(c)	$\mathrm{P}(A \cap B \cap C) = \frac{1}{15}$
		(d)	$\frac{7}{20} \le \mathrm{P}(A' \cap B' \cap C') \le \frac{13}{30}$
10	Sampling; Hypothesis testing	**(a)**	• The sample is taken from the population of 1 000 chips of the newest model. • The sample size (or number of chips) n is greater than 30. • The n chips are obtained randomly (or independently of one another with equal probability).
		(b)	[shown] $\mathrm{Var}(x) = 98.75$
		(c)	$H_1: \mu < 40$; $6.66 \le \alpha < 13.3$

<table>
<tr><td>11</td><td>Linear regression</td><td>(a)</td><td>
</td></tr>
<tr><td></td><td></td><td>(b)</td><td>[shown]
The square of the residual horizontal distances between the relevant data points and the best-fit line is minimised.</td></tr>
<tr><td></td><td></td><td>(c)</td><td>Required value = 8.2 degree Celsius (allow marginal error)
This value is unreliable.</td></tr>
<tr><td></td><td></td><td>(d)</td><td>$t = 417.40 - \frac{25}{9}T$
$r = -0.9561828875 \approx\ 0.956$</td></tr>
<tr><td></td><td></td><td>(e)</td><td>$k = 3$ (allow marginal error)
The data is selected in the range of t where T is strictly increasing and such that the product moment correlation coefficient is as close to 1 as possible.</td></tr>
</table>

Suggested solutions and post-mortem

Qn	Suggested Solutions	Comments
	Section A: Pure Mathematics [40 marks]	
1 **[4]**	Finding the cartesian equation of L_1, $\mathbf{r} = \begin{pmatrix} k \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ k \end{pmatrix}$ $\rightarrow x - k = \dfrac{y-1}{k}$ $\rightarrow y = kx - k^2 + 1$ Finding the cartesian equation of L_2, $\mathbf{r} = \begin{pmatrix} 1 \\ k \end{pmatrix} + \mu \begin{pmatrix} k \\ 1 \end{pmatrix}$ $\rightarrow \dfrac{x-1}{k} = y - k$ $\rightarrow y = \dfrac{1}{k}x - \dfrac{1}{k} + k$ Required sequence of transformation: 1st: Scaling by a scale factor of k^2 parallel to the $x-$axis 2nd: Translation of $k^2 - 1 - \dfrac{1}{k} + k$ units in the positive $y-$direction	This question mainly assesses on converting vector equations into cartesian equations. Admittedly, despite being within the expectations of the syllabus, two-dimensional vector equations rarely appear in A–Levels. The latter part concerning graph transformations should be relatively more routine. Note that when proposing translations involving unknown constants, **ensure that it is done (1) with a positive unit of translation, and (2) in the correct direction**. (In this case, it is graphically verifiable that $k^2 - 1 - \frac{1}{k} + k > 0$ for $k > 1$.)

2
[5]

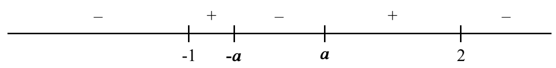
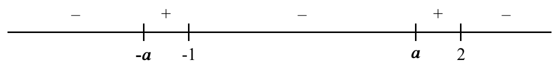
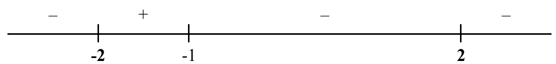
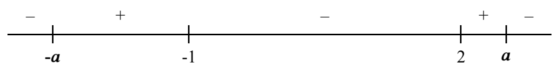
$$\frac{a^2 - x - 2}{x^2 - x - 2} \geq 1$$

$$\rightarrow \frac{a^2 - x - 2 - (x^2 - x - 2)}{x^2 - x - 2} \geq 0$$

$$\rightarrow \frac{a^2 - x^2}{x^2 - x - 2} \geq 0$$

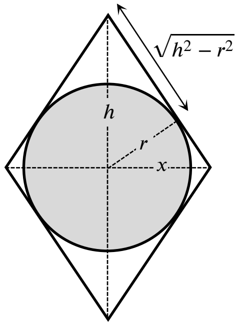
$$\rightarrow \frac{(a + x)(a - x)}{(x - 2)(x + 1)} \geq 0$$

Principal values are $x = -1, 2$ and $\pm a$. Given $a > 0$, there are 5 possible solution intervals:

Case	Value of a	Number line (deduced from GC)	Solution interval
1	$0 < a < 1$		$-1 < x \leq -a$ or $a \leq x < 2$
2	$a = 1$		$1 \leq x < 2$
3	$1 < a < 2$		$-a \leq x < 1$ or $a \leq x < 2$
4	$a = 2$		$-2 \leq x < -1$
5	$a > 2$		$-a \leq x < -1$ or $2 < x \leq a$

This question concerns graph sketching and solving inequalities. As this question does not explicitly prohibit the use of graphing calculators (GC), candidates may simply substitute different a values to observe the different cases of graphs and deduce the required solution intervals, all without sketching these graphs on paper as it is not required by the question.

Having said so, a manual attempt by hand is perfectly welcomed and encouraged. In fact, answers that rely on GC may also benefit from some manual prework which could reveal strategic choices of a values.

3 **[8]**	Consider the cross-section of the octahedron where it is tangent to the inscribed sphere.  By similar triangles, $\frac{x}{r}=\frac{h}{\sqrt{h^2-r^2}} \rightarrow x=\frac{rh}{\sqrt{h^2-r^2}}$ $\therefore$ The volume of the inscribing tetrahedron, V $=2\left(\frac{1}{3}\right)(2x)^2(h)$ $=\frac{2}{3}\left(\frac{4r^2h^2}{h^2-r^2}\right)(h)$ $=\frac{8r^2}{3}\left(\frac{h^3}{h^2-r^2}\right)$ Differentiating V with respect to h, $\frac{\mathrm{d}V}{\mathrm{d}h}=\frac{8r^2}{3}\left(\frac{3h^2\left(h^2-r^2\right)-h^3(2h)}{(h^2-r^2)^2}\right)$ $\frac{\mathrm{d}V}{\mathrm{d}h}=\frac{8r^2}{3}\left(\frac{h^2\left(3h^2-3r^2-2h^2\right)}{(h^2-r^2)^2}\right)$ $\frac{\mathrm{d}V}{\mathrm{d}h}=\frac{8}{3}r^2\left(\frac{h^2\left(h^2-3r^2\right)}{(h^2-r^2)^2}\right)$	Questions on maxima and minima and justification of the nature of stationary values are the bread and butter of A–Levels and school examinations. This question concerns packing geometrical shapes, as is the trend for most optimisation problems, and hence would call for fluency in geometry and trigonometry. When finding an expression for the volume of the octahedron in the beginning, **caution is advised when relying only on the given diagram in the question**, as there is risk of misinterpretation – in this case, potentially mistaking r as half of the side of the pyramid's square base. This could be prevented should candidates produce their own cross-sectional diagram to confirm where the sphere touches the octahedron. All lengths that are relevant to calculate the volume of the octahedron can be found in terms of h and r through geometry. When differentiating the volume, given that the letters involved in the differentiation are aplenty, **candidates should carefully distinguish the variable of differentiation from the constants** – in this case, h is the variable and r is a constant.

[Continued]

For stationary value of $V, \dfrac{\mathrm{d}V}{\mathrm{d}h} = 0$

$\rightarrow h^2\left(h^2 - 3r^2\right) = 0$

$\rightarrow h^2 = 0 \text{ or } h^2 = 3r^2$

$\rightarrow h = 0 \text{ or } h = \pm\sqrt{3}r$

Since $h > 0$, $h = \sqrt{3}r$

Considering sign test, notice that

$$\frac{\mathrm{d}V}{\mathrm{d}h} = \frac{8}{3}r^2\left(\frac{h^2\left(h^2 - 3r^2\right)}{(h^2 - r^2)^2}\right) = \frac{8}{3}\left(\frac{rh}{h^2 - r^2}\right)^2\left(h^2 - 3r^2\right)$$

h	$\left(\sqrt{3}r\right)^-$	$\sqrt{3}r$	$\left(\sqrt{3}r\right)^+$
$h^2 - 3r^2$	$-$	0	$+$
$\dfrac{\mathrm{d}V}{\mathrm{d}\theta} = \dfrac{8}{3}\left(\dfrac{rh}{h^2 - r^2}\right)^2\left(h^2 - 3r^2\right)$	$-$	0	$+$
Slope	\	—	/

$\therefore h = \sqrt{3}r$ gives the minimum volume for the inscribing tetrahedron.

Minimum volume

$$= \frac{8r^2}{3}\left(\frac{\left(\sqrt{3}r\right)^3}{\left(\sqrt{3}r\right)^2 - r^2}\right) = \frac{8r^2}{3}\left(\frac{3\sqrt{3}r^3}{3r^2 - r^2}\right) = \frac{8r^2}{3}\left(\frac{3\sqrt{3}r}{2}\right) = 4\sqrt{3}r^3 \text{ units}^3$$

As per routine, stationary values are found through the roots of the first derivative. In that process, **it is considered good practice to justify any root rejections or acceptance in context**, which would otherwise be a common cause of penalisation. In this context, it is obvious that non-positive solutions for the height h of the pyramid should be expressly rejected.

Proving that the volume found is indeed minimum can be done in a couple of ways: second differentiation or sign test. In this case, the expression for the first derivative lends itself to a factored form that works best with sign test. **Be advised that a proper sign test must show how all significant factors contribute to the sign change.**

If all constants values are known, justifications using derivative values around the stationary point as quoted from a calculator would also be acceptable.

4 (a) [3]	$y = e^{2x} \cos ax$ Differentiating with respect to x, $\frac{dy}{dx} = 2e^{2x} \cos ax - ae^{2x} \sin ax$ $\frac{dy}{dx} = 2y - ae^{2x} \sin ax$ Differentiating with respect to x for the second time, $\frac{d^2y}{dx^2} = 2\frac{dy}{dx} - 2ae^{2x} \sin ax - a^2e^{2x} \cos ax$ $\frac{d^2y}{dx^2} = 2\frac{dy}{dx} - 2\left(2y - \frac{dy}{dx}\right) - a^2y$ $\frac{d^2y}{dx^2} = 2\frac{dy}{dx} - 4y + 2\frac{dy}{dx} - a^2y$ $\frac{d^2y}{dx^2} = 4\frac{dy}{dx} - \left(a^2 + 4\right)y$	Like most questions on Maclaurin expansions in A–Levels, the bulk of the mark is attributed to the rigour of performing repeated differentiation. Of note in this question is the use of intermediate results as substitutions, which is a common exploit in Maclaurin expansion questions in A–Levels – in this case, substituting $ae^{2x} \sin ax = 2y - \frac{dy}{dx}$, which may not be apparent on first glance.
4 (b) [4]	Differentiating with respect to x for the third time, $\frac{d^3y}{dx^3} = 4\frac{d^2y}{dx^2} - \left(a^2 + 4\right)\frac{dy}{dx}$ When $x = 0$, $\rightarrow y = 1$, $\rightarrow \frac{dy}{dx} = 2(1) - a(1)(0) = 2$, $\rightarrow \frac{d^2y}{dx^2} = 4(2) - \left(a^2 + 4\right)(1) = 4 - a^2$ $\rightarrow \frac{d^3y}{dx^3} = 4\left(4 - a^2\right) - \left(a^2 + 4\right)(2) = 16 - 4a^2 - 2a^2 - 8 = 8 - 6a^2$ $\therefore$ Maclaurin expansion of y is given by $y = 1 + 2x + \frac{1}{2}\left(4 - a^2\right)x^2 + \frac{1}{6}\left(8 - 6a^2\right)x^3 + \cdots$ $y = 1 + 2x + \left(2 - \frac{a^2}{2}\right)x^2 + \left(\frac{4}{3} - a^2\right)x^3 + \cdots$	This second part requires as much rigour and attentiveness as the first part, especially since this Maclaurin expansion involves coefficients in terms of an unknown constant.

4 (c) [2]	From first derivative, we have: $\frac{\mathrm{d}y}{\mathrm{d}x} = 2y - a\mathrm{e}^{2x}\sin ax$ $\therefore \mathrm{e}^{2x}\sin ax$ $= \frac{2}{a}y - \frac{1}{a}\frac{\mathrm{d}y}{\mathrm{d}x}$ $= \frac{2}{a}\left[1 + 2x + \left(2 - \frac{a^2}{2}\right)x^2 + \cdots\right] - \frac{1}{a}\left[2 + 2\left(2 - \frac{a^2}{2}\right)x + 3\left(\frac{4}{3} - a^2\right)x^2 + \cdots\right]$ $= \frac{2}{a} + \frac{4}{a}x + \left(\frac{4}{a} - a\right)x^2 + \cdots - \frac{2}{a} - \left(\frac{4}{a} - a\right)x - \left(\frac{4}{a} - 3a\right)x^2 - \cdots$ $= ax + 2ax^2 + \cdots$	As this final part asks to "hence" deduce the expansion for another function, it would be only fitting to start by referring to results found in earlier parts, particularly where $\mathrm{e}^{2x}\sin ax$ may have appeared in some intermediary results. The subsequent steps call for differentiation involving unknown constants, which should lead to the required result if done carefully and correctly.

5 (a) [4]

Tabulate amounts as follows:

Month	Amount at the start of this month	Amount at the end of this month
1	\$100,000	\$100,000(1.02)
2	\$100,000(1.02) + \$1,000	$\$100{,}000(1.02)^2 + \$1{,}000(1.02)$
3	$\$100{,}000(1.02)^2 + \$1{,}000(1.02+1)$	$\$100{,}000(1.02)^3 + \$1{,}000\left(1.02^2 + 1.02\right)$
⋮	⋮	⋮
n	S_n	

The amount at the start of the nth month after January 2023 in Mr Wong's account, S_n

$= \$100{,}000(1.02)^{n-1} + \$1{,}000\left(1.02^{n-2} + 1.02^{n-3} + \cdots + 1\right)$

$= \$100{,}000(1.02)^{n-1} + \$1{,}000\left(\dfrac{1.02^{n-1} - 1}{1.02 - 1}\right)$

$= \$100{,}000(1.02)^{n-1} + \$50{,}000\left(1.02^{n-1} - 1\right)$

$= \$150{,}000(1.02)^{n-1} - \$50{,}000$

Using GC to find the least n such that $S_n > \$300{,}000$,

n	S_n
42	\$287,830.07 (< \$300,000)
43	\$294,586.67 (< \$300,000)
44	\$301,478.40 (> \$300,000)

At the start of the 44th month after January 2023, the total is \$301,478.40, which exceeds \$300,000. At the end of the 43rd month after January 2023, the total is \$300,478.40, since Mr Wong is yet to put \$1,000 more into the account. This also exceeds \$300,000.

$\therefore$ The total in the account will first exceed the price of the car at the end of July 2026.

Financial mathematics involving compound interest is a classic application of geometric progression in A–Levels. When approaching questions involving interest and monthly investment, it is best to tabulate the monthly amounts for clarity. This will greatly help in finding the expression for the account total in terms of the months passed n.

Just as important, **great care must be taken when writing the ratio and avoid mistaking the percentage as the ratio itself** (i.e., "$r = 0.02$" is incorrect).

The context behind financial mathematics question such as this one may vary slightly across questions, most notably the day of the month in which the monthly investment and the interest rate occurs. While it may seem no different to sum the amounts on either day, **it is recommended to find the sum on days the monthly investment occurs**. Here are some reasons:

- The newly added investment adds one (+1) to the geometric portion of the sum – that portion becomes $\left(r^{n-1} + r^{n-2} + \cdots + r + 1\right)$, which has a more direct sum expression as opposed to the incomplete $\left(r^{n-1} + r^{n-2} + \cdots + r\right)$.

- It is easier to use this sum to justify answers involving the "first day vs. last day" of the month. The question is simply reduced to "does the amount exceed with or without this '+1?'"

The final part on counting months can be done by expressing n as $12a + b$, where $a, b \in \mathbb{Z}^+$ and $b < 12$, and simply add a years and b months to the initial time.

5 (b) [5]

After deposit, the remaining amount to be paid is \$300,000 – \$75,000 = \$225,000

Tabulate amounts as follows:

Month	Outstanding at the start of this month	Outstanding at the end of this month
1	\$225,000	\$225,000(1.005)
2	\$225,000(1.005) – \$15,000	\$225,000$(1.005)^2$ – \$15,000(1.005)
3	\$225,000$(1.005)^2$ – \$15,000(1.005 + 1)	\$225,000$(1.005)^3$ – \$15,000$(1.005^2 + 1.005)$
$\vdots$	$\vdots$	$\vdots$
n	R_n	

The outstanding amount to be paid at the start of the nth payment month, R_n

$$= \$225{,}000(1.005)^{n-1} - \$15{,}000\left(1.005^{n-2} + 1.005^{n-3} + \cdots + 1\right)$$
$$= \$225{,}000(1.005)^{n-1} - \$15{,}000\left(\frac{1.005^{n-1} - 1}{1.005 - 1}\right)$$
$$= \$225{,}000(1.005)^{n-1} - \$3{,}000{,}000\left(1.005^{n-1} - 1\right)$$
$$= \$3{,}000{,}000 - \$2{,}775{,}000(1.005)^{n-1}$$

Using GC to find the least n such that $R_n < \$0$,

n	R_n
15	\$24,308.86 (> \$0)
16	\$9,430.40 (> \$0)
17	–\$5,522.44 (< \$0)

At the beginning of the seventeenth payment month, R_n starts to become negative. This means that the monthly instalments of \$15,000 first exceeds the outstanding amount to be paid at this month. Hence, he would have completely paid for the car at the beginning of the seventeenth payment month.

$\therefore$ The amount to be paid at this month = \$15,000 – \$5,522.44 = \$9,477.56.

This part onwards regarding the hire purchase scheme requires proper question interpretation, though the approach is identical to the previous part.

Great care must be taken to address several contextual information:

- **Initial payment**: After the deposit, the amount affected by the interest is \$225,000 (not \$300,000).
- **Interest**: The 0.5% interest converts to a ratio of 1.005 (not 0.5 or 1.05).
- **Instalments**: Monthly instalments are paid "on subsequent months", so apart from the deposit, there is no further deduction of \$15,000 for the first payment month.
- **Final amount**: A good initial assumption is that the final amount due is most likely less than \$15,000.

5 (c) [5]

Let the amount at this first payment month be S_k.

At the first payment month, Mr Wong pays a deposit of \$75,000 and no longer puts additional \$1,000.
$\therefore S_k = \$150{,}000(1.02)^{k-1} - \$50{,}000 - \$75{,}000 - \$1{,}000 = \$150{,}000(1.02)^{k-1} - \$126{,}000$

Using the number of month and the amount paid found in **(iii)**,
$\therefore S_{k+17} = \$527{,}328.50 + \$9{,}477.56 = \$536{,}806.06$

Tabulate amounts as follows:

Month	Amount at the start of this month	Amount at the end of this month
k	$\$150{,}000(1.02)^{k-1}$ $-\$126{,}000$	$\$150{,}000(1.02)^{k}$ $-\$126{,}000(1.02)$
$k+1$	$\$150{,}000(1.02)^{k}$ $-\$126{,}000(1.02)$ $-\$15{,}000$	$\$150{,}000(1.02)^{k+1}$ $-\$126{,}000(1.02)^2$ $-\$15{,}000(1.02)$
$k+2$	$\$150{,}000(1.02)^{k+1}$ $-\$126{,}000(1.02)^2$ $-\$15{,}000(1.02+1)$	$\$150{,}000(1.02)^{k+2}$ $-\$126{,}000(1.02)^3$ $-\$15{,}000\left(1.02^2+1.02\right)$
$\vdots$	$\vdots$	$\vdots$
$k+17$	S_{k+17}	

S_{k+17}
$= \$150{,}000(1.02)^{k+16} - \$126{,}000(1.02)^{17} - \$15{,}000\left(1.02^{16} + 1.02^{15} + \cdots + 1\right)$
$= \$150{,}000(1.02)^{k+16} - \$126{,}000(1.02)^{17} - \$15{,}000\left(\dfrac{1.02^{17}-1}{1.02-1}\right)$
$= \$150{,}000(1.02)^{k+16} - \$126{,}000(1.02)^{17} - \$750{,}000\left(1.02^{17}-1\right)$
$= \$150{,}000(1.02)^{k+16} - \$876{,}000(1.02)^{17} + \$750{,}000 = \$536{,}806.06$

Using GC, $k = 80$

$\therefore$ Mr Wong made the first payment in August 2029.

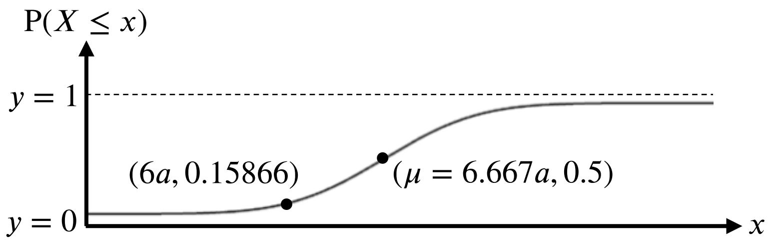
This part requires the expression for the amount in the bank account pre-purchase, which will be used accordingly as a starting point. Once the expression for the starting capital and the fixed amount to be deducted monthly are found, the approach to this question becomes identical to the previous part.

Great care must be taken to address several contextual information:

- **No more monthly investment**: Avoid using a new expression that assumes Mr Wong is still investing \$1,000 monthly during the hire purchase. Also, **make sure to deduct \$1,000 accordingly from S_n**, should it include the last monthly investment.
- **Final payment**: This must be added back to the given balance to obtain the correct total after interest.
- **Interest**: Note that bank interest still applies throughout the deduction, and as such **S_n is not as simple to find as adding back the total amount paid to the given balance, as this would ignore interest being applied while deduction is still occurring**. Also, care must be taken to not confuse the interest value from the bank (2%) with that from the hire purchase (0.5%).

This part also requires counting months, which can be done by expressing n in the form $12a + b$, as recommended in **(b)**.

Section B: Probability and Statistics [60 marks]		
6 (a) [5]	$\mathrm{P}(X \le 6a) = 0.15866$ $\to \mathrm{P}\left(Z \le \frac{6a-\mu}{\sigma}\right) = 0.15866$ $\to \frac{6a-\mu}{\sigma} = -0.99998$ $\to 6a - \mu + 0.99998\sigma = 0$ — (1) $\frac{X_1 + X_2 + X_3 + X_4}{4} = \bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{4}\right)$ $\mathrm{P}\left(6a \le \bar{X} \le 7a\right) = 0.81859$ $\to \mathrm{P}\left(\bar{X} \le 7a\right) - \mathrm{P}\left(\bar{X} \le 6a\right) = 0.81859$ $\to \mathrm{P}\left(Z \le \frac{7a-\mu}{\frac{\sigma}{2}}\right) - \mathrm{P}\left(Z \le \frac{6a-\mu}{\frac{\sigma}{2}}\right) = 0.81859$ $\to \mathrm{P}\left(Z \le \frac{7a-\mu}{\frac{\sigma}{2}}\right) - \mathrm{P}\left(Z \le 2(-0.99998)\right) = 0.81859$ $\to \mathrm{P}\left(Z \le \frac{7a-\mu}{\frac{\sigma}{2}}\right) = \mathrm{P}\left(Z \le 2(-0.99998)\right) + 0.81859 = 0.02275 + 0.81859 = 0.84134$ $\to \frac{7a-\mu}{\frac{\sigma}{2}} = 0.99998$ $\to 7a - \mu - 0.49999\sigma = 0$ — (2) Using GC to solve (1) and (2) simultaneously, $\mu = 6.667a$ and $\sigma = 0.667a$	This question mainly tests on forming relationships between μ, σ and a using the inverse of standard normal distribution. The two probability values yield a system of two equations in terms of the three variables, which can be solved using calculator to obtain the desired final expressions. A key observation here is recognising that the intermediate value obtained for $\frac{6a-\mu}{\sigma}$ is relevant in evaluating the similar looking $\mathrm{P}\left(Z \le \frac{6a-\mu}{\frac{\sigma}{2}}\right)$.

6 (b) [2]		This part is a different take on the more familiar question which asks for the sketch of a normal distribution (also known as *probability density function*), which seems to be the trend in recent A–Level papers. For the case of the graph of $\mathrm{P}(X \le x)$ against x (also known as *cumulative distribution function*), there are some important features to include: • asymptotic behaviours to indicate that cumulative probabilities can never equal 0 or 1, and • the steepest slope at the mean value.
7 (a) [1]	Let X denote the number of times that the computer decides to add 1 to the sum. $\therefore X \sim \mathrm{B}(n, p)$. For the case where $n = 9$, P(sum = 3) = P(6 additions and 3 subtractions in any order) $= \binom{9}{6} p^6 q^3$ $= 84p^6q^3$	This question tests on binomial distribution modelling given an event in context. This part also serves as a primer to subsequent parts which require similar models.
7 (b) [3]	For the case where $n = 5$, P(sum = 3) = P(4 additions and 1 subtraction in any order) $= \binom{5}{4} p^4 q$ $= 5p^4q$ Since it is equally likely that the sum is 3 after 9 actions as it is after 5 actions, $84p^6q^3 = 5p^4q$ $p^2\left(1 - p^2\right) = \dfrac{5}{84}$ $p^4 - p^2 + \dfrac{5}{84} = 0$ Using GC to find the positive roots of p, $p \approx 0.968$ or 0.252	This part follows immediately after finding the expression in **7(a)**. With further adjustment to the previously used binomial model, a second expression can be found, and equating the two expressions will yield the result.

7 (c) [5]	The sum can only be positive if computer does addition for more than half of the times: • For $2k-1$ actions, at least k additions must occur. • For $2k+1$ actions, at least $k+1$ additions must occur. Let $X \sim \mathrm{B}(2k-1, p)$. The probability that the sum is positive after $2k-1$ actions, $\mathrm{P}(\mathrm{sum}_{2k-1} > 0)$ $= \mathrm{P}(X \geq k)$ $= \mathrm{P}(X = k) + \mathrm{P}(X \geq k+1)$ The probability that the sum is positive after $2k+1$ actions $\mathrm{P}\left(\mathrm{sum}_{2\mathrm{k}+1} > 0\right)$ $= \mathrm{P}(X = k-1) \times \mathrm{P}(\text{all 2 remaining adds}) + \mathrm{P}(X = k) \times \mathrm{P}(\text{1 of 2 remaining adds}) + \mathrm{P}(X \geq k+1)$ $= \mathrm{P}(X = k-1)p^2 + \mathrm{P}(X = k)\left(1-(1-p)^2\right) + \mathrm{P}(X \geq k+1)$ The following inequality is obtained: $\mathrm{P}\left(\mathrm{sum}_{2\mathrm{k}+1} > 0\right) > \mathrm{P}(\mathrm{sum}_{2k-1} > 0)$ $\mathrm{P}(X = k-1)p^2 + \mathrm{P}(X = k)\left(1-(1-p)^2\right) + \mathrm{P}(X \geq k+1) > \mathrm{P}(X = k) + \mathrm{P}(X \geq k+1)$ $\mathrm{P}(X = k-1)p^2 - \mathrm{P}(X = k)(1-p)^2 > 0$ $\left[\binom{2k-1}{k-1}p^{k-1}(1-p)^k\right]p^2 - \left[\binom{2k-1}{k}p^k(1-p)^{k-1}\right](1-p)^2 > 0$ $\left[\dfrac{(2k-1)!}{(k-1)!\,k!}p^k(1-p)^k\right]p - \left[\dfrac{(2k-1)!}{k!\,(k-1)!}p^k(1-p)^k\right](1-p) > 0$ $\left[\dfrac{(2k-1)!}{(k-1)!\,k!}p^k(1-p)^k\right](2p-1) > 0$ Given that $0 < p < 1$, $p^k(1-p)^k > 0$ for all possible k $\rightarrow 2p-1 > 0$ $\rightarrow p > \dfrac{1}{2}$ $\therefore \dfrac{1}{2} < p < 1$ (since $0 < p < 1$)	This part may prove challenging. To begin this part, the number of additions required for the sum to be positive for each $(2k-1)$ and $(2k+1)$ actions must first be considered. The subsequent part entails forming an inequality and requires making use of a common random variable to model both the case of $(2k-1)$ actions and $(2k+1)$ actions. In this case, both cases are modelled after $X \sim \mathrm{B}(2k-1, p)$. This model directly models the case of $(2k-1)$ actions. For $(2k+1)$ actions, X models the first $(2k-1)$ actions, and the remaining two actions are considered case-by-case. The remaining intermediate steps simplify the inequality by utilising the expression for $P(X = x)$ for binomial distributions found in the List of Formulae (MF27). With proper factorisation and justification, the required range of p is subsequently obtained. Note that the final range of values of p to be obtained must also account for the fact that $0 < p < 1$, as given by the question – while the decision to penalise may vary based on strictness of the marking scheme, it is still considered erroneous to only conclude that $p > \frac{1}{2}$.

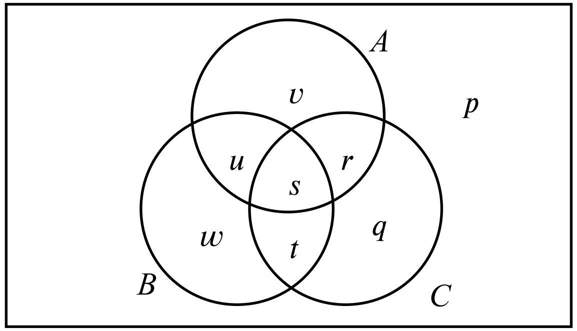
8 (a) [2]	$\mathrm{P}(X = r)$ $= \mathrm{P}(\text{pick ball } r \text{ from a box and pick ball} \le r \text{ from another box}) - \mathrm{P}(\text{pick equal ball from both})$ $= 2\left(\frac{1}{N}\right)\left(\frac{r}{N}\right) - \frac{2}{2!}\left(\frac{1}{N}\right)\left(\frac{1}{N}\right)$ $= \frac{2r}{N^2} - \frac{1}{N^2} = \frac{2r-1}{N^2}$	This question tests on formulating probability of events given a context. Note that picking the two balls can be done in any order (i.e., Box A then Box B, or Box B first then Box A) and thus the probability is doubled to give $2\left(\frac{1}{N}\right)\left(\frac{r}{N}\right)$. Equally of note is the fact that picking two balls with equal value is considered the same event, and thus the probability should be deducted accordingly.
8 (b) [2]	$\mathrm{E}(X)$ $= \sum_{r=1}^{N} r\,\mathrm{P}(X = r)$ $= \sum_{r=1}^{N} r\left(\frac{2r-1}{N^2}\right)$ $= \frac{1}{N^2}\sum_{r=1}^{N} 2r^2 - r$ $= \frac{1}{N^2}\left(2\sum_{r=1}^{N} r^2 - \sum_{r=1}^{N} r\right)$ $= \frac{1}{N^2}\left(\frac{1}{3}N(N+1)(2N+1) - \frac{1}{2}N(N+1)\right)$ $= \frac{2N(N+1)(2N+1) - 3N(N+1)}{6N^2}$ $= \frac{(N+1)(4N-1)}{6N}$ $= \frac{4N^2+3N-1}{6N}$	Finding the expectation of an event is considered routine though A–Levels tend to make the probability distribution of the event apparent, such as in a tabulated form that is either provided or to be found. This question provides a different way of representing the probability distribution without tables, which may blur the relationship $\mathrm{E}(X) = \sum_{r=1}^{N} r\,\mathrm{P}(X = r)$ and the significance of the result in **(i)**.

8 (c) [5]	$\mathrm{P}(X \le m) = \sum_{r=1}^{m} \frac{2r-1}{N^2} = \frac{1}{N^2}\left(\frac{2m(m+1)}{2} - m\right) = \frac{m^2}{N^2}$ $\mathrm{P}(X \ge m) = 1 - \mathrm{P}(X \le m-1) = 1 - \frac{(m-1)^2}{N^2}$ $\mathrm{P}(X \ge m) = 1 - \frac{(m-1)^2}{N^2} \ge \frac{1}{2}$ $\rightarrow \frac{(m-1)^2}{N^2} \le \frac{1}{2}$ $\rightarrow m - 1 \le \frac{1}{\sqrt{2}}N = \frac{\sqrt{2}}{2}N$ $\rightarrow m \le \frac{\sqrt{2}}{2}N + 1$ $\mathrm{P}(X \le m) = \frac{m^2}{N^2} \ge \frac{1}{2}$ $\rightarrow m \ge \frac{1}{\sqrt{2}}N = \frac{\sqrt{2}}{2}N$ $\therefore \frac{\sqrt{2}}{2}N \le m \le \frac{\sqrt{2}}{2}N + 1$ Using the given $\mathrm{E}(X)$ value to find N and m, $\frac{4N^2 + 3N - 1}{6N} = \frac{40299}{600}$ $\rightarrow N = 100$ (using GC / by observation) $\rightarrow 50\sqrt{2} \le m \le 50\sqrt{2} + 1$ $\rightarrow 70.7 \le m \le 71.7$ $\rightarrow m = 71$	Upon a quick glance, there are three requirements in this part that can be noticed: • finding the expressions for $\mathrm{P}(X \le m)$ and $\mathrm{P}(X \ge m)$, • using these expressions to deduce the required inequality in *m*, and • finding *m* for a specific case of N. Finding an expression for $\mathrm{P}(X \le m)$ entails using the result in **(i)** as a summation of probabilities and the use of the arithmetic series formula. Finding an expression for $\mathrm{P}(X \ge m)$ entails similar working, though also of note here is the use of complementary events. Eventually, the two expressions work out to yield the maximum and minimum values of *m*, which can be combined to give the required inequality. The value of N can be easily deduced from $\mathrm{E}(X)$, and the remaining steps to deduce the integer value *m* follow.

9 (a) [3]	$\mathrm{P}(A\mid B')=\dfrac{\mathrm{P}(A\cap B')}{\mathrm{P}(B')}$ $\frac{1}{6}\left(1-\mathrm{P}(B)\right)=\mathrm{P}(A\cup B)-\mathrm{P}(B)$ $\frac{1}{6}-\frac{1}{6}\mathrm{P}(B)=\frac{1}{2}-\mathrm{P}(B)$ $\frac{5}{6}\mathrm{P}(B)=\frac{1}{2}-\frac{1}{6}=\frac{1}{3}$ $\therefore\ \mathrm{P}(B)=\frac{2}{5}$ $\mathrm{P}(A\cup B)=\mathrm{P}(A)+\mathrm{P}(B)-\mathrm{P}(A\cap B)$ $\frac{1}{2}=\frac{1}{3}+\frac{2}{5}-P(A\cap B)$ $\therefore\ \mathrm{P}(A\cap B)=\frac{1}{3}+\frac{2}{5}-\frac{1}{2}=\frac{7}{30}$	This question particularly assesses on the ability to manipulate probability values using various techniques, especially those pertaining complementary events, union events, and expressions for conditional probabilities.
9 (b) [1]	**Method 1: Using the given result** $\mathrm{P}(A\cup B\cup C)=\mathrm{P}\left(A\cup(B\cup C)\right)$ $=\mathrm{P}(A)+\mathrm{P}(B\cup C)-\mathrm{P}\left(A\cap(B\cup C)\right)$ $=\mathrm{P}(A)+\mathrm{P}(B\cup C)-\mathrm{P}\left((A\cap B)\cup(A\cap C)\right)$ $=\mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)-\mathrm{P}(B\cap C)-[\mathrm{P}(A\cap B)+\mathrm{P}(A\cap C)-\mathrm{P}(A\cap B\cap A\cap C)]$ $=\mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)-\mathrm{P}(A\cap B)-\mathrm{P}(A\cap C)-\mathrm{P}(B\cap C)+\mathrm{P}(A\cap B\cap C)$ **Method 2: Using Inclusion-Exclusion Principle (MF27)** $\mathrm{P}(A\cup B\cup C)$ $=\mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)-[\mathrm{P}(A\cap B)+\mathrm{P}(A\cap C)+\mathrm{P}(B\cap C)]+\mathrm{P}(A\cap B\cap C)$ $=\mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)-\mathrm{P}(A\cap B)-\mathrm{P}(A\cap C)-\mathrm{P}(B\cap C)+\mathrm{P}(A\cap B\cap C)$	This part makes use of the probability for the union of two events and applies it to the extended case of the union of three events. With the updated List of Formulae (MF27), there are two acceptable approaches to this question. Apart from Method 1, which was previously the only accepted method, candidates may now otherwise show the result using the Inclusion-Exclusion Principle (found under "*Mathematical Results*"). As a side, $\mathrm{P}\left(A\cap(B\cup C)\right)=\mathrm{P}\left((A\cap B)\cup(A\cap C)\right)$ is known as *De Morgan's law*, a particularly useful and fundamental result in the realm of propositional logic and Boolean algebra.
9 (c) [2]	Using the result in **(ii)**, $1-\dfrac{11}{30}=\dfrac{1}{3}+\dfrac{2}{5}+\dfrac{1}{4}-\dfrac{7}{30}-\left(\dfrac{1}{3}\right)\left(\dfrac{1}{4}\right)-\left(\dfrac{2}{5}\right)\left(\dfrac{1}{4}\right)+\mathrm{P}(A\cap B\cap C)$ $\therefore\ \mathrm{P}(A\cap B\cap C)=\dfrac{1}{15}$	This part entails using the result in **(ii)**, but not before recognising that the value of $\mathrm{P}(A'\cap B'\cap C')$ is simply the complement of $\mathrm{P}(A\cup B\cup C)$. Equally important to notice in this question is the condition of independent events.

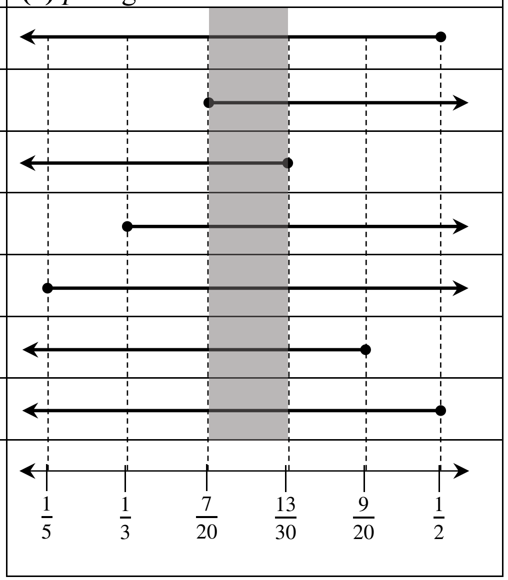
9 (d) [4]

Let $\mathrm{P}(A' \cap B' \cap C') = p$, where $p \geq 0$. Labelling the Venn diagram as follows:



Region	**(1)** Probability of region ≥ 0	**(2)** p range
q	$1 - \mathrm{P}(A \cup B) - p = \frac{1}{2} - p \geq 0$	$p \leq \frac{1}{2}$
r	$\mathrm{P}(C) - q - \mathrm{P}(B \cap C) = p - \frac{7}{20} \geq 0$	$p \geq \frac{7}{20}$
s	$\mathrm{P}(A \cap C) - r = \frac{13}{30} - p \geq 0$	$p \leq \frac{13}{30}$
t	$\mathrm{P}(B \cap C) - s = p - \frac{1}{3} \geq 0$	$p \geq \frac{1}{3}$
u	$\mathrm{P}(A \cap B) - s = p - \frac{1}{5} \geq 0$	$p \geq \frac{1}{5}$
v	$\mathrm{P}(A) - r - s - u = \frac{9}{20} - p \geq 0$	$p \leq \frac{9}{20}$
w	$\mathrm{P}(B) - s - t - u = \frac{1}{2} - p \geq 0$	$p \leq \frac{1}{2}$

(3) p range in a number line



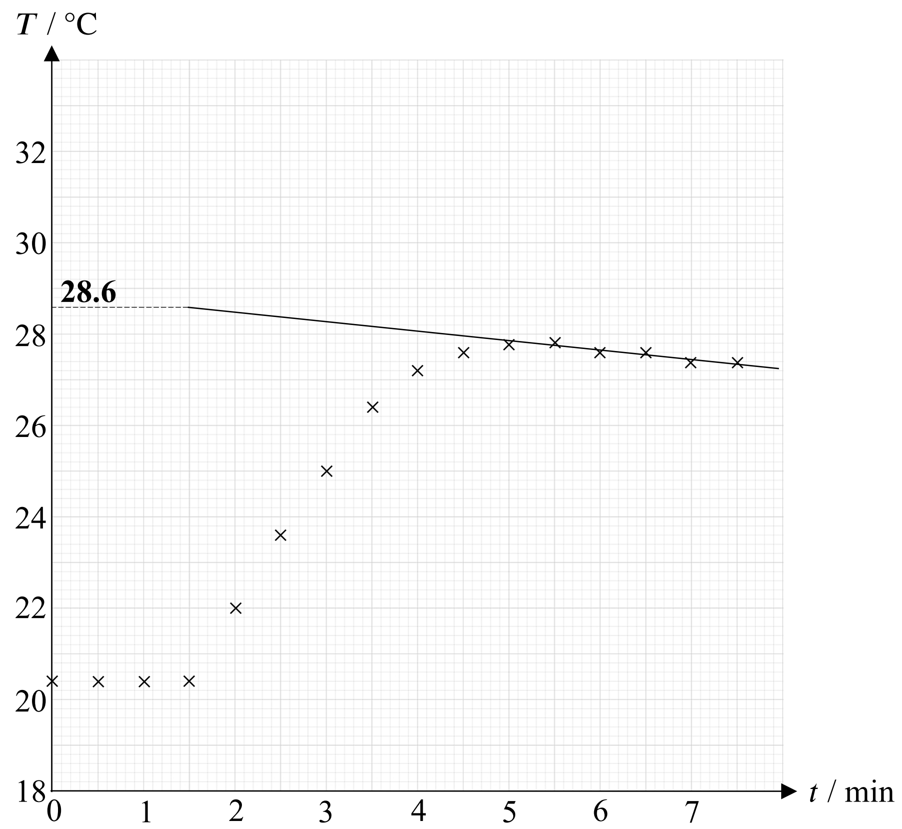
$$\therefore \frac{7}{20} \leq \mathrm{P}(A' \cap B' \cap C') \leq \frac{13}{30}$$

Questions on maximum and minimum probabilities involving Venn Diagrams have currently witnessed a rising trend in both JC papers and A–Levels. As such, it would be wise to get familiar with methods to tackle such questions.

This suggested solution offers a possible strategy. The key idea in this approach is maintaining positive (or non-negative) values on regions in the Venn diagram that are not explicitly known to be zero.

10 (a) [3]	• The sample is taken from the population of 1 000 chips of the newest model. • The sample size (or number of chips) n is greater than 30. • The n chips are obtained randomly (or independently of one another with equal probability).	This question tests on sampling methods. Of note in this part are the significant factors of consideration when obtaining a sample that is appropriate for z-test: • **Sample vs. Population**: this is subtle, but it must be acknowledged that the sample should **only** be obtained from the new 1 000 chips. • **Central Limit Theorem**: given that the sample is taken from a population with **an unknown distribution**, and yet it is already appropriate for use in z-test, it must be that Central Limit Theorem applies here and thus its size must be greater than 30. • **Randomness**: random selection must be done to remove bias (e.g., each selection should **not** depend on the performance of the previously selected chip).
10 (b) [2]	$\mathrm{E}(t)$ $= \dfrac{15(4)+25(14)+35(37)+45(33)+55(12)}{100}$ $= 38.5$ $\mathrm{E}\left(t^2\right)$ $= \dfrac{15^2(4)+25^2(14)+35^2(37)+45^2(33)+55^2(12)}{100}$ $= 1581$ $\mathrm{Var}(t)$ $= \mathrm{E}\left(t^2\right) - \mathrm{E}(t)^2$ $= 1581 - 38.5^2$ $= 98.75$	When given a tiered data summary, the calculations for expectation values makes use of the midpoint of each tier. The remainder of this question is standard practice in A–Levels, dealing with expectations and variance of a sample.

10 (c) [7]	Unbiased estimate of the population mean $= \mathrm{E}(t) = 38.5$ Unbiased estimate of the population variance $= \dfrac{n}{n-1}\mathrm{Var}(t) = \dfrac{100}{99}(98.75)$ Let T be the time taken, in minutes, for the device to carry out the tasks with a random new chip. H_0: $\mu = 40$ H_1: (to be found) Test at α% significance level. Under H_0, since $n = 100$ is large, by Central Limit Theorem, $\bar{T} \sim \mathrm{N}\left(40, \dfrac{98.75}{99}\right)$ approximately. Test statistics is given by $Z = \dfrac{\bar{T} - 40}{\sqrt{\dfrac{98.75}{99}}} \sim \mathrm{N}(0,1)$ With $\bar{t} = 38.5$, z–value $= -1.501897534$ For rejection, z–value must lie in the critical region. The unique alternative hypothesis condition requires that, given a value of α, z–value must lie in **only one** out of three critical regions possible. Consider right-tailed test or two-tailed test, H_1: $\mu > 40$ or H_1: $\mu \neq 40$. → If there is α such that the right-tailed critical region or two-tailed critical region includes the z–value respectively, that α also yields a left-tailed critical region big enough to include the z–value. → ∴ The alternative hypothesis is not H_1: $\mu > 40$ or H_1: $\mu \neq 40$. Consider left-tailed test, H_1: $\mu < 40$. → The left-tailed critical value should be more than the z–value. → $\therefore \alpha\% \geq \mathrm{P}(Z \leq -1.501897534) = 0.0666$ → The α must also be such that the two-tailed critical region excludes the z–value. → $\therefore \alpha\% < 2(0.0666) = 0.133$ $\therefore H_1: \mu < 40, \quad 6.66 \leq \alpha < 13.3$	As hinted by the question, a feasible approach to this question is first finding the test statistic: the z–value. Care must be taken during the following: • finding unbiased estimate of the population variance • stating Central Limit Theorem • writing down the parameters for $\bar{T}$, especially its variance This test statistic is subsequently used to find H_1 by considering possible critical regions and proceeding with elimination. A step-by-step thought process follows: 1. **List down the possible cases for H_1**: since A–Levels only tests on two-tailed tests and one-tailed tests, there are effectively only three cases of H_1. 2. **Consider the respective critical region for each H_1 that "just nice" rejects H_0:** how does the critical region look like for each case of H_1 that "just nice" contains the z–value? What should the α–value be in this case? 3. **Deduce whether there is only one H_1 that cause rejection under this "just nice" α–value**: if the "just nice" critical region is large, the required α–value will also be large, making it more likely for other critical regions with the same area to contain the z–value and thus also cause rejection. After deducing the correct alternative hypothesis, the range of α can be deduced from considering the "just nice" critical regions.

11 (a) [1]	[The following diagram also bears suggested sketches for **11(ii)** and **11(iii)**.] 	A strategic grid sketch considers scale of axes beforehand. Upon preliminary reading of the subsequent question parts, it can be noticed that in **(c)** the sketch will need to be used to find an extrapolated value of T. An appropriate scale would consider the case that the extrapolated value lies beyond the data range for T.
11 (b) [2]	To ensure that the line is best-fit, the square of the residual **horizontal** distances between the relevant data points and the best-fit line is minimised. For this data set, this line passes through midpoints of pairs of data that has the same value of T (i.e., passes through midpoint of (5, 27.8) and (5.5, 27.8), etc.).	This question mainly aims to test on awareness of the type of residual distances used, horizontal or vertical, for a given regression line.
11 (c) [3]	Required value = 28.6 − 20.4 = 8.2 degree Celsius This value is unreliable. The temperature obtained at $t = 1.5$ from L is an extrapolation as the t value lies outside the data range $5 \le t \le 7.5$ from which L was drawn.	Note that the temperature difference using the maximum T data is less than using the extrapolated T value. This shows that while extrapolation is mathematically unreliable, it is still serviceable in certain applications.

<table>
<tr>
<td>11
(d)
[4]</td>
<td>
The following data is used:
<table>
<tr><td>t</td><td>5</td><td>5.5</td><td>6</td><td>6.5</td><td>7</td><td>7.5</td></tr>
<tr><td>T</td><td>27.8</td><td>27.8</td><td>27.6</td><td>27.6</td><td>27.4</td><td>27.4</td></tr>
</table>
For the given data, regression line of t on T:
$t = 144.25 - 5T$

If T were given in degrees Rankine, convert to Celsius first:
$t = 144.25 - 5\left(\frac{5(T - 491.67)}{9}\right)$
$t = 417.40 - \frac{25}{9}T$

$r = -0.9561828875 \approx\ 0.956$
This applies to both cases where T is given in degrees Celsius or degrees Rankine.
</td>
<td>
This question tests on three concepts:

calculation of regression line using calculator,
conversion of regression line in the context of transformed data, and
understanding that product moment correlation coefficient of a regression line remains unaffected after translation and scaling

Great care must be taken during the Celsius–Rankine conversion. It would be remiss to just substitute T in the regression line equation with $\frac{9}{5}T + 491.67$.
</td>
</tr>
<tr>
<td>11
(e)
[3]</td>
<td>
The following data is used:
<table>
<tr><td>t</td><td>1.5</td><td>2</td><td>2.5</td><td>3</td><td>3.5</td></tr>
<tr><td>T</td><td>20.4</td><td>22.0</td><td>23.6</td><td>25.0</td><td>26.4</td></tr>
</table>
For the given data, regression line of T on t:
$T = 15.98 + 3t$
$\rightarrow k = 3$

The data is selected in the range of t where T is strictly increasing and such that the product moment correlation coefficient is as close to 1 as possible. For the regression line used above, $r = 0.99938 \approx 0.999$, which indicates a very strong positive linear relationship between t and T, and thus its gradient accurately represents the rate at which T increases with respect to t.
</td>
<td>
This question, atop testing on calculating the regression line using calculator, also tests on the ability to justify the reliability of the behaviour of the data based on the r–value of its regression line.

To obtain an accurate value of k in this question, the regression line must first be suitably obtained from an appropriate range which indicates a constant rate of increase of T (since k is constant). This constant rate of increase can be observed from the scatter diagram in (a) in the range $1.5 \le t \le 3.5$. Any data outside this range not only reduces the r–value (and hence affecting the reliability of the answer) but also suggests that T does not increase at a constant rate.
</td>
</tr>
</table>