CATHOLIC JUNIOR COLLEGE
General Certificate of Education Advanced Level
Higher 2
JC2 Preliminary Examination

CANDIDATE NAME	

CLASS		INDEX NUMBER	

MATHEMATICS **9758/02**

Paper 2 **17 Sep 2025**
3 hours

Additional Materials: Printed Answer Booklet
List of Formulae (MF27)

READ THESE INSTRUCTIONS FIRST

Answer **all** the questions.
Write your answers on the Printed Answer Booklet. Follow the instructions on the front cover of the answer booklet.

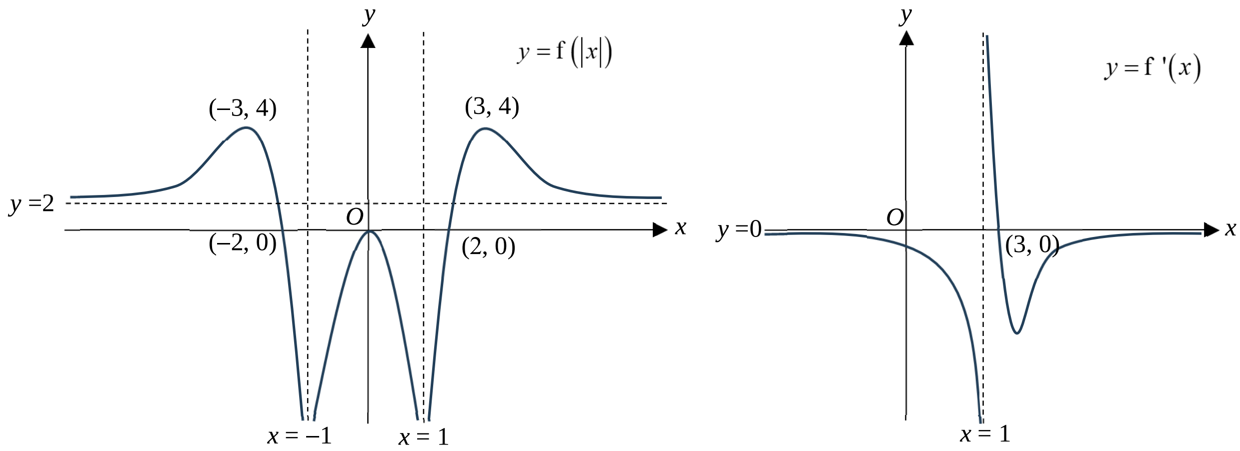
Give non-exact numerical answers correct to 3 significant figures, or 1 decimal place in the case of angles in degrees, unless a different level of accuracy is specified in the question.
You are expected to use an approved graphing calculator.
Unsupported answers from a graphing calculator are allowed unless a question specifically states otherwise.
Where unsupported answers from a graphing calculator are **not** allowed in a question, you must present the mathematical steps using mathematical notations and not calculator commands.
You must show all necessary working clearly.

The number of marks is given in brackets [ ] at the end of each question or part question.

This document consists of **8** printed pages.

Section A: Pure Mathematics [40 marks]

1 The diagram shows the graphs of $y = \mathrm{f}\left(|x|\right)$ and $y = \mathrm{f}'(x)$.



The graph of $y = \mathrm{f}\left(|x|\right)$ has turning points at $(-3, 4)$ and $(3, 4)$, crosses the x-axis at $(-2, 0)$ and $(2, 0)$, and the equations of asymptotes are $x = -1$, $x = 1$ and $y = 2$. The graph of $y = \mathrm{f}'(x)$ crosses the x-axis at $(3, 0)$, and the equations of asymptotes are $x = 1$ and $y = 0$.

On separate diagrams, sketch the graphs of

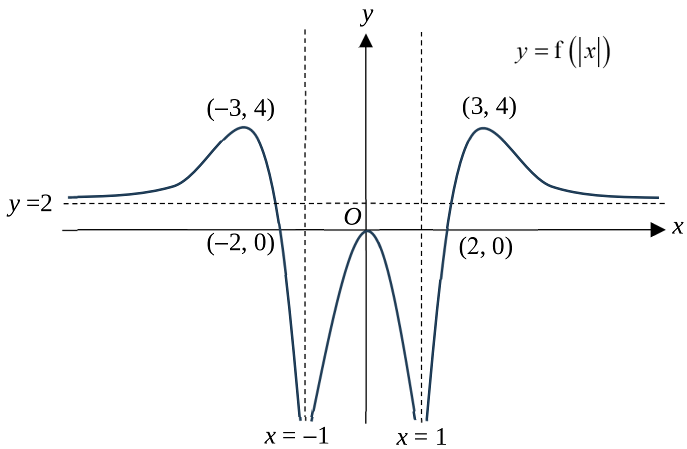
(a) $y = \mathrm{f}(x)$, [2]

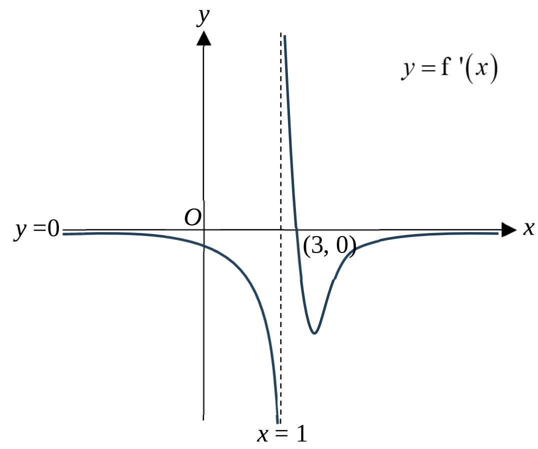
(b) $y = \dfrac{1}{\mathrm{f}'(x)}$, [3]

(c) $y = -\mathrm{f}\left(|x-1|\right)$, [3]

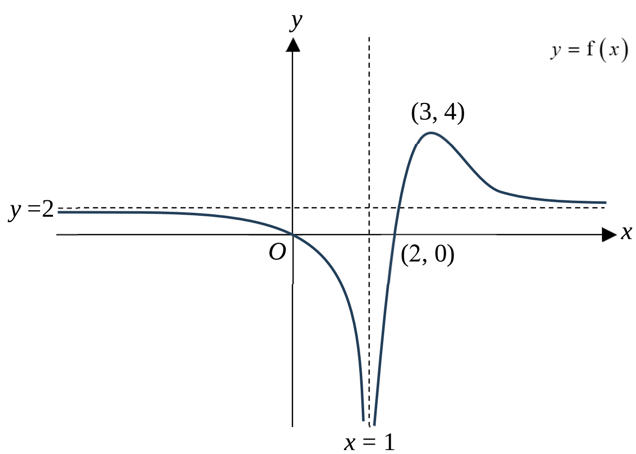
labelling clearly the equation(s) of any asymptote(s), coordinates of any axial intercept(s) and turning point(s) where applicable.

(a)



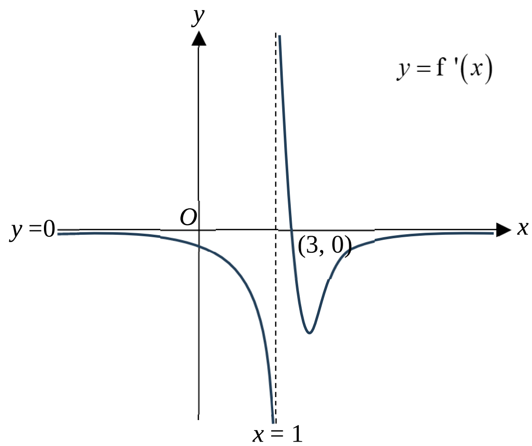


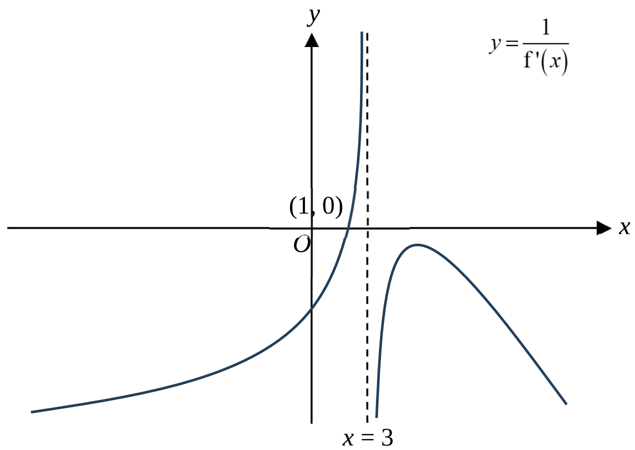
Solution:



Examiners Feedback: Well attempted. Most students were able to achieve full credit. Others were still able to get the portion of the graph, $x \geq 0$ correct.

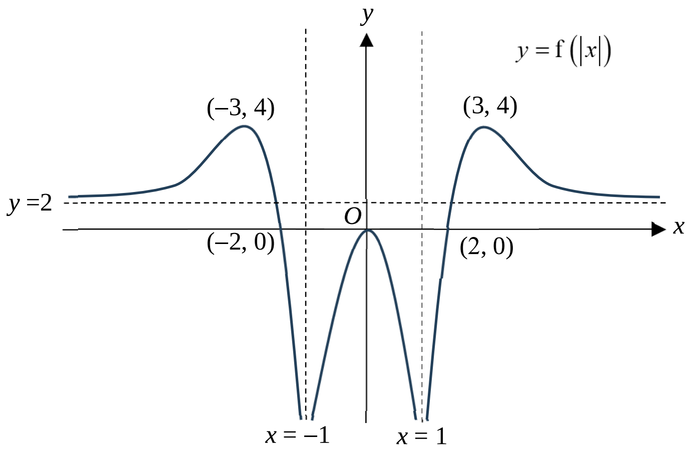
(b)





Examiners Feedback:
Poorly attempted. Students were able to identify x-intercept (1,0) and vertical asymptote, $x = 3$ but unable to sketch the curve. Others were able to go on and identify that there is a maximum point in the curve $y = \dfrac{1}{\mathrm{f}\,'(x)}$, but indicated this point in the 1st quadrant instead of the 4th quadrant.

(c)



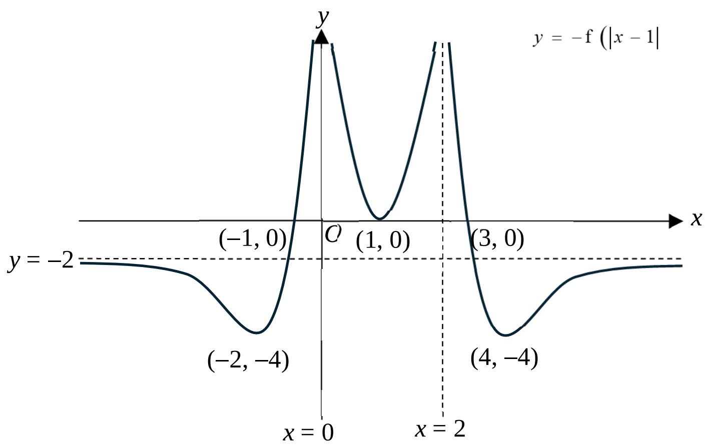
$y = \mathrm{f}\left(|x|\right)$

$\downarrow$ translate 1 unit in positive x-axis direction

$y = \mathrm{f}\left(|x-1|\right)$

$\downarrow$ reflect in x-axis

$y = -\mathrm{f}\left(|x-1|\right)$



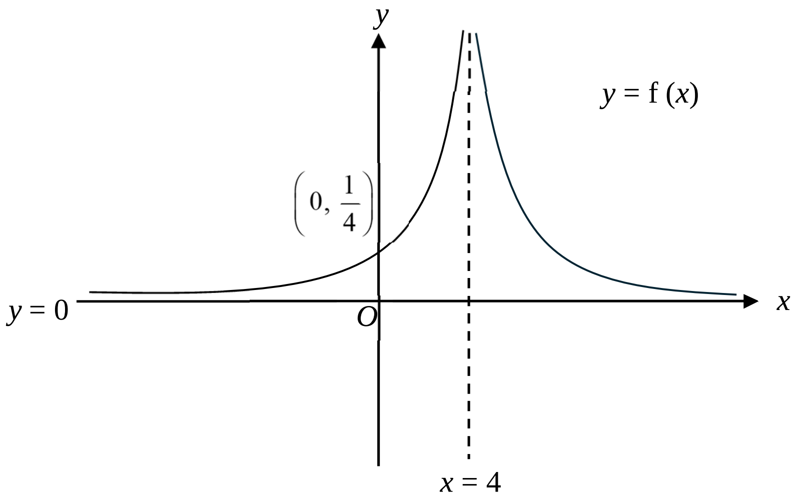
Examiners Feedback: Poorly attempted. Students made several careless mistakes including labeling of coordinates of the axial intercepts and turning points and asymptotes. Students are advised to sketch an intermediate graph if they are not confident in sketching the final graph directly. This will allow them to secure maximum marks in the exam.

2 Functions f and g are defined by

$$\mathrm{f} : x \to \frac{4}{(x-4)^2}, \quad x \in \mathbb{R},\ x \neq 4,$$

$$\mathrm{g} : x \to \ln\left(1+\frac{1}{x}\right), \quad x \in \mathbb{R},\ x > 0.$$

(a) Sketch the graph of $y = \mathrm{f}(x)$, stating the equations of any asymptotes, the coordinates of the points where it crosses the axes and the coordinates of the turning points, if any. [2]



Examiners Feedback: Majority of the candidates were able to sketch the graph well. However, many candidates could not get full credit due to lapses in the labelling of the features of the graph. <u>Common mistakes observed include:</u> • Missing vertical asymptotes. • Not labelling the horizontal asymptote when the question asked to state. • Not giving the *y*-intercept in coordinates form when required by the question. <u>Candidates should review:</u> • The techniques of curve sketching which includes the features such as asymptotes, intercepts and turning points. • Read the question carefully to identify any requirements from the question (e.g. coordinates form, state equation of asymptotes etc).

(b) Show that gf exists. Hence find the rule, domain and range of gf. [4]

For gf to exist, $\mathrm{R_f} \subseteq \mathrm{D_g}$,

$\mathrm{R_f} = (0, \infty)$

$\mathrm{D_g} = (0, \infty)$

Since $\mathrm{R_f} \subseteq \mathrm{D_g}$, gf exists (shown)

Hence, find the rule,

$\mathrm{gf}(x)$

$= \mathrm{g}\left[\dfrac{4}{(x-4)^2}\right]$

$= \ln\left[1+\left(\dfrac{1}{\frac{4}{(x-4)^2}}\right)\right]$

$= \ln\left[1+\dfrac{(x-4)^2}{4}\right]$

domain

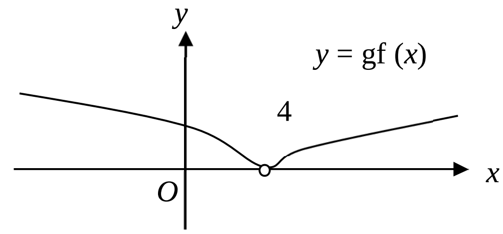
$\mathrm{D_{gf}} = \mathrm{D_f} = (-\infty, 4)\cup(4, \infty)$

OR

$\mathrm{D_{gf}} = \mathrm{D_f} = \mathbb{R}\setminus\{4\}$

and range of gf.

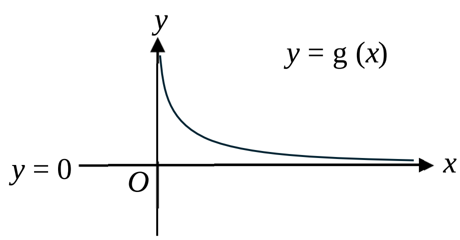
<u>Method ①:</u> Sketch $y = \mathrm{gf}(x)$ for given domain



$\mathrm{R_{gf}} = (0, \infty)$

<u>Method ②:</u> Mapping

$\mathrm{D_f} = (-\infty, 4)\cup(4, \infty) \xrightarrow{\mathrm{f}} \mathrm{R_f} = (0, \infty) \xrightarrow{\mathrm{g}} \mathrm{R_{gf}} = ($



Examiners Feedback:

This part of the question was not very well attempted. Many candidates missed out certain components (did not find rule/domain/range) in this part of the question.

<u>Common mistakes observed include:</u>

- Unable to establish the range of f and domain of g.
- Incorrect condition for gf to exist, e.g. $\mathrm{D_g} \subseteq \mathrm{R_f}$, $\mathrm{R_g} \subseteq \mathrm{D_f}$, $\mathrm{D_f} \subseteq \mathrm{R_g}$ etc
- Incorrect $\mathrm{D_{gf}}$ such as $\mathrm{D_{gf}} = \mathrm{D_g}$, or writing as $(-\infty,4)\cap(4,\infty)$.
- Incorrect method to find $\mathrm{R_{gf}}$ such as using $\mathrm{R_{gf}} = \mathrm{R_f}$ or $\mathrm{R_{gf}} = \mathrm{R_g}$ without mentioning about the restricted domain. Note: Marks were not awarded even though the answer is correct here.

<u>Candidates should review:</u>

The condition for existence of a composite function, as well as to find the rule, domain and range of the composite function (including the use of the restricted domain to find the range).

(c) If the domain of f is further restricted to $x < k$, state the largest value of k for which the function f^{-1} exists. [1]

Largest value of k is 4.

Examiners Feedback:

This part of the question was well done. Most candidates were able to recognise that the largest value of k occurs at the turning point. However, some candidates were confused by the vertical asymptote and gave $k = 3$ instead.

(d) Using the restricted domain found in part **(c)**, find f^{-1} in a similar form. [3]

$$y=\frac{4}{(x-4)^2}$$

$$(x-4)^2=\frac{4}{y}$$

$$x-4=\pm\frac{2}{\sqrt{y}}$$

$$x=4+\frac{2}{\sqrt{y}}\ (\text{rej.} \because x<4) \text{ or } x=4-\frac{2}{\sqrt{y}}$$

$$\mathrm{D}_{\mathrm{f}^{-1}}=\mathrm{R}_{\mathrm{f}}=(0,\infty)$$

$$\mathrm{f}^{-1}: x \to 4-\frac{2}{\sqrt{x}}, \quad x\in\mathbb{R},\ x>0$$

Examiners Feedback:

This part of the question was well done. Most candidates were able to let $y=\mathrm{f}(x)$ and make x the subject to find the rule.

Common mistakes observed include:

- Not including the $\pm$ when taking square roots to both sides of the equation.
- Not knowing or reject the wrong expression.
- Used the restricted domain of f as the domain of the inverse.
- Not writing in similar form when the question requires the inverse in the similar form.

Candidates should review:

The concept of finding the rule and domain of an inverse function, especially for questions that require the decision to reject one of the expression. (Decision is based on domain of f)

3

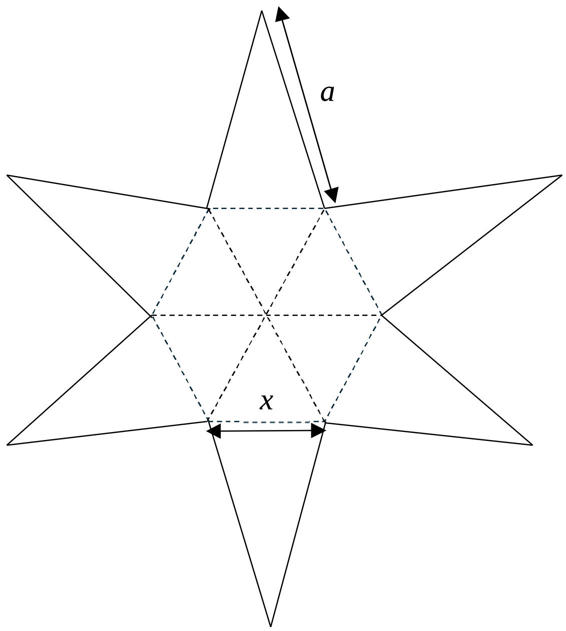


Fig. 1

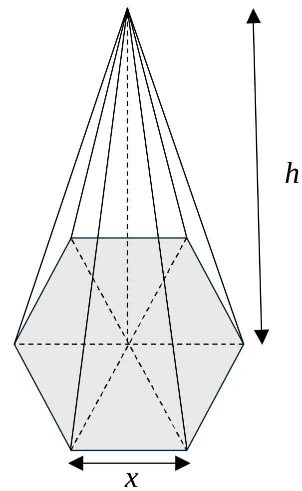


Fig. 2

Fig. 1 shows a net of a hexagonal pyramid folded from a star-shaped cardboard of equal edge length a cm. The net consists of a hexagon with equal sides of x cm and six isosceles triangles with base x cm and side a cm. The net is folded to form a right pyramid with a hexagonal base of edge length x cm and vertical height h cm, as shown in Fig. 2. The hexagonal base is made up of six equilateral triangles of side length x cm.

The volume of a right hexagonal pyramid with base edge x cm and height h cm is given by $V = \frac{\sqrt{3}}{2}x^2 h$.

(a) Show that the volume of the hexagonal pyramid, V satisfies the expression given by

$$V^2 = \frac{3}{4}\left(a^2x^4 - x^6\right)$$ [2]

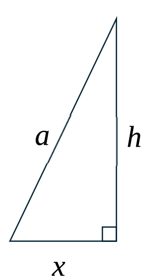
By Pythagoras Theorem,

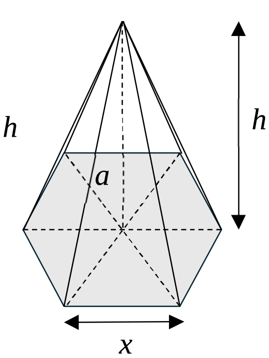
$a^2 = x^2 + h^2$

$h = \sqrt{a^2 - x^2}$

$V = \frac{\sqrt{3}}{2}x^2 h$

$= \frac{\sqrt{3}}{2}x^2\sqrt{a^2 - x^2}$





[Turn Over

$$V^2=\frac{3}{4}x^4\left(a^2-x^2\right)$$

$$V^2=\frac{3}{4}\left(a^2x^4-x^6\right) \qquad \text{(shown)}$$

Examiners Feedback:
Majority of the cohort was able to identify the correct triangle to use and apply Pythagoras Theorem correctly to obtain the required equation.

(b) Find, in terms of a, the maximum possible volume of the hexagonal pyramid. You need not show that this value is a maximum. [4]

Method ①:

$$V^2=\frac{3}{4}\left(a^2x^4-x^6\right)$$

Differentiate implicitly with respect to x:

$$2V\frac{\mathrm{d}V}{\mathrm{d}x}=\frac{3}{4}\left(4a^2x^3-6x^5\right)$$

At stationary point, $\frac{\mathrm{d}V}{\mathrm{d}x}=0$

$$4a^2x^3-6x^5=0$$

$$2x^3\left(2a^2-3x^2\right)=0$$

Since $x\neq 0$, $2a^2-3x^2=0$

$$x^2=\frac{2}{3}a^2$$

Method ②:

$$V^2=\frac{3}{4}\left(a^2x^4-x^6\right)$$

$$V=\frac{\sqrt{3}}{2}\left(a^2x^4-x^6\right)^{\frac{1}{2}}$$

$$\frac{\mathrm{d}V}{\mathrm{d}x}=\frac{\sqrt{3}}{2}\left[\frac{1}{2}\left(a^2x^4-x^6\right)^{-\frac{1}{2}}\left(4a^2x^3-6x^5\right)\right]$$

$$=\frac{\sqrt{3}}{4}\left[\left(4a^2x^3-6x^5\right)\left(a^2x^4-x^6\right)^{-\frac{1}{2}}\right]$$

At stationary point, $\frac{\mathrm{d}V}{\mathrm{d}x}=0$

$$\frac{\sqrt{3}}{4}\left[\left(4a^2x^3-6x^5\right)\left(a^2x^4-x^6\right)^{-\frac{1}{2}}\right]=0$$

$$4a^2x^3-6x^5=0$$

$$2x^3\left(2a^2-3x^2\right)=0$$

Since $x\neq 0$, $2a^2-3x^2=0$

$$x^2=\frac{2}{3}a^2$$

$$\max V=\sqrt{\frac{3}{4}\left[a^2\left(\frac{2}{3}a^2\right)^2-\left(\frac{2}{3}a^2\right)^3\right]}$$

$$=\sqrt{\frac{3}{4}\left[\frac{4}{9}a^6-\frac{8}{27}a^6\right]}$$

$$=\sqrt{\frac{3}{4}\left(\frac{4}{27}a^6\right)}$$

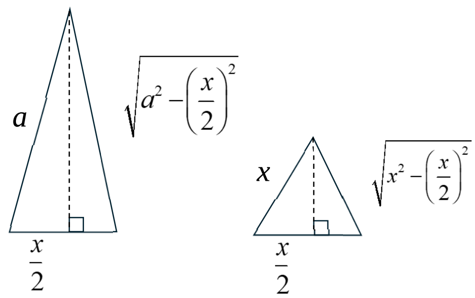
$$=\sqrt{\frac{a^6}{9}}$$

$$=\frac{a^3}{3}$$

Examiners Feedback:
Students who used the V^2 equation given in **(a)** and performed Implicit Differentiation were mostly successful.
Some students took square root of the given equation and differentiated but were less successful and made mistakes when applying Chain Rule and/or Product Rule.

(c) Find, in terms of a, the total surface area of the hexagonal pyramid when the volume is a maximum. [4]

Total surface area, $A = 6\times$ area of isosceles triangles $+$ area of hexagon



$$\begin{aligned}
A &= 6\times\frac{1}{2}\times x\times\sqrt{a^2-\left(\frac{x}{2}\right)^2}+6\times\frac{1}{2}\times x\times\sqrt{x^2-\left(\frac{x}{2}\right)^2} \\
&= 3x\sqrt{a^2-\frac{x^2}{4}}+3x\sqrt{\frac{3x^2}{4}} \\
&= 3\sqrt{\frac{2}{3}a^2}\sqrt{a^2-\frac{1}{4}\times\frac{2}{3}a^2}+3\sqrt{\frac{2}{3}a^2}\sqrt{\frac{3}{4}\times\frac{2}{3}a^2} \\
&= 3\sqrt{\frac{2}{3}a^2\times\frac{5}{6}a^2}+3\sqrt{\frac{2}{3}a^2\times\frac{1}{2}a^2} \\
&= 3\sqrt{\frac{5}{9}}a^2+\frac{3}{\sqrt{3}}a^2 \\
&= \left(\sqrt{5}+\sqrt{3}\right)a^2 \text{ units}^2
\end{aligned}$$

Examiners Feedback:
This part was badly done. The question required visualizing the surface area and finding the different area of triangles which many students were unable to perform.
There was also weakness observed generally in algebraic manipulation and the simplification of surds.

4 One of the roots of the equation

$$2z^4 - 14z^3 + 33z^2 - 26z + p = 0 \text{, where } p \text{ is a constant}$$

is $3+\mathrm{i}$.

(a) Based on the above information only, a student claims that the equation has a root $3-\mathrm{i}$. State, with a reason, why the student's claim may not be true. [1]

The student's claim may not be true as p may not be real.

Examiners Feedback:
Despite this question being similar to a discussion question (Complex Numbers Q7(i)), it was not well done.
For candidates who did not score well in the entire question, they were not familiar with the fact that the coefficients of the terms in the polynomial must be real for the other root to be $3-\mathrm{i}$. Even though candidates realise this, those that did not link to the constant *p* will not be awarded with the credit. Candidates need to mention that ***p* may not be real** to answer the question.

Some mistakes in phrasing included *p* may not be "positive", "negative", "integer" which shows poor understanding of Complex Numbers. Note that some candidates also wrote *p* is a complex number which is not accurate as **real numbers are subset of complex numbers, hence a complex number can be a real number or an imaginary number.**

Candidates are encouraged to practise this topic well.

(b) Show that $p = 10$. [2]

$$2z^4 - 14z^3 + 33z^2 - 26z + p = 0$$
$$2(3+\mathrm{i})^4 - 14(3+\mathrm{i})^3 + 33(3+\mathrm{i})^2 - 26(3+\mathrm{i}) + p = 0$$
$$2(8+6\mathrm{i})(8+6\mathrm{i}) - 14(3+\mathrm{i})(8+6\mathrm{i}) + 33(8+6\mathrm{i}) - 78 - 26\mathrm{i} + p = 0$$
$$2(28+96\mathrm{i}) - 14(18+26\mathrm{i}) + 264 + 198\mathrm{i} - 78 - 26\mathrm{i} + p = 0$$
$$56 + 192\mathrm{i} - 252 - 364\mathrm{i} + 264 + 198\mathrm{i} - 78 - 26\mathrm{i} + p = 0$$
$$-10 + p = 0$$
$$p = 10$$

Examiners Feedback:
Most candidates are able to exhibit the correct method of substituting the given root $3+\mathrm{i}$ into the equation and showing that $p = 10$. As this is a "show" question, some working will be appreciated despite the fact that $(3+\mathrm{i})^2$, $(3+\mathrm{i})^3$ and $(3+\mathrm{i})^4$ can be evaluated using GC since the use of calculators are not restricted here.
There were a handful of candidates who let $p = a + b\mathrm{i}$ into the equation and compared real and imaginary coefficients which is acceptable but not required since *p* is the only unknown that can be evaluated by making *p* the subject.

Candidates who quoted the "Conjugate Root Theorem" here receive no credit since the student's claim is not proven to be correct. They should present the workings in part (c) instead. Failing to present the workings in part (c) result in zero credit too.

For the rest of this question, do not use a calculator.

(c) Find the roots of the equation $2z^4 - 14z^3 + 33z^2 - 26z + 10 = 0$ and mark them clearly on a single labelled Argand diagram. [7]

Method ①:

Since the coefficients of the polynomial are all real, $3+\mathrm{i}$ is a root, $3-\mathrm{i}$ is also a root.

$$
\begin{aligned}
&\left[z-(3+\mathrm{i})\right]\left[z-(3-\mathrm{i})\right]\\
&=\left[(z-3)-\mathrm{i}\right]\left[(z-3)+\mathrm{i}\right]\\
&=(z-3)^2-\mathrm{i}^2\\
&=z^2-6z+9+1\\
&=z^2-6z+10
\end{aligned}
$$

$$
\begin{array}{r}
2z^2-2z+1 \\
z^2-6z+10\overline{\big)\,2z^4-14z^3+33z^2-26z+10} \\
-\underline{\left(2z^4-12z^3+20z^2\right)} \\
-2z^3+13z^2-26z \\
-\underline{\left(-2z^3+12z^2-20z\right)} \\
z^2-6z+10 \\
-\underline{\left(z^2-6z+10\right)}
\end{array}
$$

$\therefore\ 2z^4-14z^3+33z^2-26z+p=\left(z^2-6z+10\right)\left(2z^2-2z+1\right)$

Method ②:

Since the coefficients of the polynomial are all real, $3+\mathrm{i}$ is a root, $3-\mathrm{i}$ is also a root.

$$
\begin{aligned}
2z^4-14z^3+33z^2-26z+10 &= \left[z-(3+\mathrm{i})\right]\left[z-(3-\mathrm{i})\right]\left(2z^2+Az+B\right)\\
&=\left[(z-3)-\mathrm{i}\right]\left[(z-3)+\mathrm{i}\right]\left(2z^2+Az+B\right)\\
&=\left[(z-3)^2-\mathrm{i}^2\right]\left(2z^2+Az+B\right)\\
&=\left(z^2-6z+9+1\right)\left(2z^2+Az+B\right)\\
&=\left(z^2-6z+10\right)\left(2z^2+Az+B\right)
\end{aligned}
$$

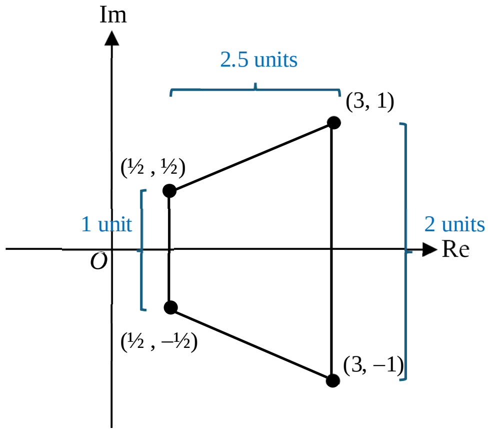
Comparing constants: $10=10B \Rightarrow B=1$

Comparing coefficients of z: $-26=-6(1)+10A \Rightarrow A=-2$

$\therefore\ 2z^4-14z^3+33z^2-26z+p=\left(z^2-6z+10\right)\left(2z^2-2z+1\right)$

$$2z^2-2z+1=0$$

$$\begin{aligned}z&=\frac{2\pm\sqrt{4-4(2)}}{4}\\&=\frac{2\pm\sqrt{-4}}{4}\\&=\frac{2\pm 2\mathrm{i}}{4}\\&=\frac{1}{2}\pm\frac{1}{2}\mathrm{i}\end{aligned}$$



Examiners Feedback:

The statement "**Since the coefficients of the polynomial are all real**, $3+\mathrm{i}$ is a root, $3-\mathrm{i}$ is also a root." MUST be written down with the 3 key points stated clearly for candidates to be awarded a mark. Despite candidates stating various names of the theorem i.e. complex-conjugate root theorem, conjugate base theorem, fundamental theorem of algebra, candidates have to explicitly state the condition as to why the theorem works which is **"all coefficients of the polynomial are real"** (and not just *p* is real). Candidates should review their discussion questions and assignments solution to see how such complex numbers questions are answered.

Candidates who were prepared for the exam generally are able to execute the solving of the other roots well, with a handful to improve on the need of having clear presentation in their answers. Candidates should note the following on solving of the polynomial:

- $3+\mathrm{i}$ and $3-\mathrm{i}$ are roots while $\left[z-(3+\mathrm{i})\right]$ and $\left[z-(3-\mathrm{i})\right]$ are factors. (Many are still mixing up these terms)
- The use of $a^2-b^2=(a+b)(a-b)$ is very useful when expanding

  $$\begin{aligned}&\left[z-(3+\mathrm{i})\right]\left[z-(3-\mathrm{i})\right]\\&=\left[(z-3)-\mathrm{i}\right]\left[(z-3)+\mathrm{i}\right]\\&=(z-3)^2-\mathrm{i}^2\\&=z^2-6z+9+1\\&=z^2-6z+10\end{aligned}$$

  Some candidates are relying on expansion of $(z-3-\mathrm{i})(z-3+\mathrm{i})$ which gives 9 terms and easily prone to algebraic slips which are not recommended.
- It also saves time by noting that the other unknown quadratic factor is $2z^2+Az+B$ (to find two coefficients) instead of Az^2+Bz+C (where you need to find 3 coefficients).
- As the use of calculators are not allowed, answers obtained through the use of GC/ calculator when solving for $2z^2-2z+1=0$ are not accepted. Candidates need to execute the use of quadratic formula here.

Candidates should note the following on the drawing of the Argand diagram:

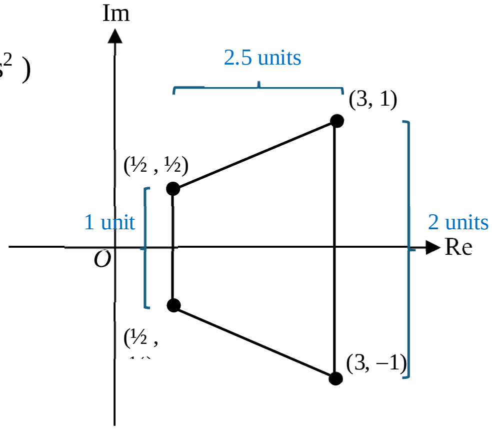
- The vertical axis must be labelled as Im while the horizontal axis must be labelled as Re. Many candidates messed this up, thereby getting the wrong points on the argand diagram.

- The numbers on both axes, are real parts i.e. $-\frac{1}{2}, \frac{1}{2}$ and 1 should be labelled on the Im-axis instead of $-\frac{1}{2}\mathrm{i}, \frac{1}{2}\mathrm{i}$ and i. (Candidates received zero credit for doing this)
- The points should be represented as coordinates i.e. $(3,1)$, $(3,-1)$, $\left(\frac{1}{2}, \frac{1}{2}\right)$ and $\left(\frac{1}{2}, -\frac{1}{2}\right)$ instead of $3+\mathrm{i}$, $3-\mathrm{i}$, $\frac{1}{2}+\frac{1}{2}\mathrm{i}$, $\frac{1}{2}-\frac{1}{2}\mathrm{i}$. (Candidates who labelled the complex number without showing the coordinates or making dotted lines to refer to the numbers on the axes received zero credit)

(d) The points of the Argand diagram in part **(c)** form the vertices of a quadrilateral. Identify the type of quadrilateral and determine its area. [2]

Trapezium

Area of qualilateral $= \frac{1}{2}(1+2)(2.5) = 3.75$ units^2 (or $\frac{15}{4}$ units^2)



Examiners Feedback:

The British English of trapezium which is Trapezoid is accepted as well.

There were some candidates who did not manage to name the shape but managed to find the area of trapezium.

Area of trapezium $= \frac{1}{2} \times$ sum of parallel sides $\times$ height .

There were also candidates who made algebraic slips while finding the sum of parallel sides, and height by referring to the wrong coordinate.

A major error that candidates do, was to include i in their calculation of area. Area is scalar quantity that cannot be imaginary.

[Turn Over

Section B: Probability and Statistics [60 marks]

5 The basketball club in a college has 5 centers, 8 forwards and 7 guards. With the National School Games approaching, the coach wishes to find out the opinions of members of the club about the training programme. He gives a questionnaire to all the members of the club and receives replies from everyone.

(a) Explain whether the 20 members form a sample or a population. [1]

The 20 members form a **population** since the coach gave questionnaire to **all members in the basketball club which consists of 20 members.**

Examiners Feedback:
Candidates who understood the difference between sample and population generally did well for this question. A handful of candidates confuse this requirement with Central Limit Theorem that is irrelevant here.

The coach then decides to select teams to play in the matches for National School Games. A basketball team to play in a match consists of 1 center, 2 forwards and 2 guards.

(b) Explain an advantage for choosing a random sample in each category of members for the match. [1]

Having a random sample **reduces biasedness**, or that it would be **fair**, or that it would be **representative of the whole membership of the club**.

Examiners Feedback:
Candidates who mentioned the keywords generally were awarded the credit.
For candidates who simply define random sample i.e. each member has an equal chance/ probability of being selected and the selection process is done independently did not receive the credit as the question did not ask for definition, but "an advantage" of a random sample instead.

(c) How many different teams can be formed? [1]

No. of teams = ${}^{5}\mathrm{C}_1 \times {}^{8}\mathrm{C}_2 \times {}^{7}\mathrm{C}_2 = 2940$

Examiners Feedback:
This question was generally well-done. However, there were some students who did not know what to do as this is a ${}^{n}C_{r}$, select/choose question. Candidates do not need to arrange the members of the team.

In the club, one particular forward is the classmate of one particular guard. Both classmates are injured and cannot participate in a particular match. The coach decides that one of the remaining guards can play either as a guard or as a forward.

(d) How many different teams can now be formed? [3]

Let A be one of the remaining guards who can play either as a guard or as a forward.

5 centres	~~8 forwards~~	~~7 guards~~
5 centres	7 forwards	6 guards (A + 5 other guards)

Method 1:

Case ①: A plays as a guard

$$\text{No. of teams} = \underbrace{{}^5C_1}_{\substack{\text{choose 1}\\\text{center}}} \times \underbrace{{}^7C_2}_{\substack{\text{choose 2 forwards}\\\text{from the remaining}}} \times \underbrace{{}^5C_1}_{\substack{\text{A plays as a guard}\\\text{choose 1 guard from}\\\text{remaining 5}}} = 525$$

Case ②: A plays as a forward

$$\text{No. of teams} = \underbrace{{}^5C_1}_{\substack{\text{choose 1}\\\text{center}}} \times \underbrace{{}^7C_1}_{\substack{\text{A plays as a forward}\\\text{choose 1 forward}\\\text{from the remaining 7}}} \times \underbrace{{}^5C_2}_{\substack{\text{choose 2 guards from}\\\text{remaining 5}}} = 350$$

Case ③: A does not play at all

$$\text{No. of teams} = \underbrace{{}^5C_1}_{\substack{\text{choose 1}\\\text{center}}} \times \underbrace{{}^7C_2}_{\substack{\text{choose 2 forwards}\\\text{from the remaining 7}}} \times \underbrace{{}^5C_2}_{\substack{\text{choose 2 guards from}\\\text{remaining 5}}} = 1050$$

Total no. of teams $= 525 + 350 + 1050 = 1925$

Method 2:

Case ①: A plays as a guard, including the case that A may not be selected

$$\text{No. of teams} = \underbrace{{}^5C_1}_{\substack{\text{choose 1}\\\text{center}}} \times \underbrace{{}^7C_2}_{\substack{\text{choose 2 forwards}\\\text{from the remaining}}} \times \underbrace{{}^6C_2}_{\substack{\text{choose 2 guards from}\\\text{6 guards including A}}} = 1575$$

Case ②: A confirmed that he plays as a forward

$$\text{No. of teams} = \underbrace{{}^5C_1}_{\substack{\text{choose 1}\\\text{center}}} \times \underbrace{{}^7C_1}_{\substack{\text{A plays as a forward}\\\text{choose 1 forward}\\\text{from the remaining 7}}} \times \underbrace{{}^5C_2}_{\substack{\text{choose 2 guards from}\\\text{remaining 5}}} = 350$$

Total no. of teams $= 1575 + 350 = 1925$

Examiners Feedback:

There were several methods presented. Candidates are reminded that the case title should be written clearly with legible phrases that markers can understand.

Let A be the remaining guard who can play both roles. This A is already decided by the coach as the question mentioned "the coach decides that one of the remaining guards can play as either a guard or as a forward". There is no need to choose this person.

Majority of the candidates show the 2 cases correctly for method 2 i.e.
Case 1: A plays as a guard
Case 2: A plays as a forward
but did not realise that these two cases overlap where A does not play at all. Candidates generally included ${}^8C_2$ instead of ${}^7C_1$ (where A is intentionally left out). Hence they were unsuccessful in the question.

For method 1, candidates who show 2 cases correctly tend to leave out the last case of "A does not play at all".

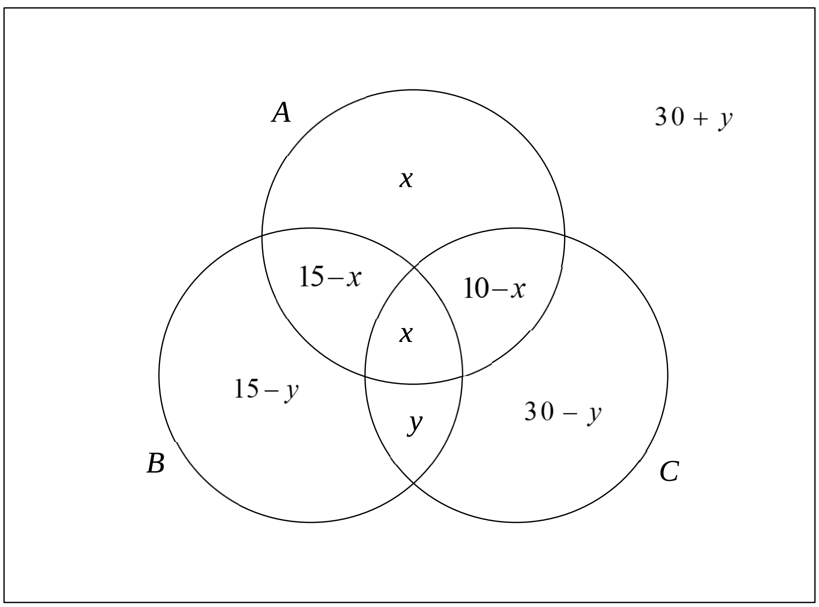
6 A group of 100 students are asked if they are student leaders, in a sports CCA, or studying in a science faculty. The number of students who are student leaders is 25, the number of students who are in a sports CCA is 30 and the number of students studying in a science faculty is 40. There are 15 student leaders who are in a sports CCA. The number of students who are student leaders, in a sports CCA and studying in a science faculty is x. The number of students who are in a sports CCA and studying in a science faculty but not a student leader is y.

One of the students is chosen at random.

A is the event that the student is a student leader.
B is the event that the student is in a sports CCA.
C is the event that the student is studying in a science faculty.

It is given that A and C are independent.

(a) Complete the Venn diagram below to represent all the above information. You are allowed to give expressions in terms of x and y. [3]



$\mathrm{P}(A) = \dfrac{25}{100}$ $\quad \mathrm{P}(B) = \dfrac{30}{100}$ $\quad \mathrm{P}(C) = \dfrac{40}{100}$

$\mathrm{P}(A \cap B) = \dfrac{15}{100}$ $\quad \mathrm{P}(A \cap B \cap C) = \dfrac{x}{100}$

Since A and C are independent, $\mathrm{P}(A \cap C) = \mathrm{P}(A) \times \mathrm{P}(C)$

$$\mathrm{P}(A \cap C) = \frac{25}{100} \times \frac{40}{100} = \frac{10}{100}$$

$$\mathrm{n}(A \cap C) = 10$$

It is further given that B and C are independent.

(b) Find y in terms of x. Hence, find the greatest and least possible values of y. [4]

Since B and C are independent, $\mathrm{P}(B\cap C)=\mathrm{P}(B)\times\mathrm{P}(C)$

$$\mathrm{P}(B\cap C)=\frac{30}{100}\times\frac{40}{100}=\frac{12}{100}$$
$$\mathrm{n}(B\cap C)=12$$
$$x+y=12$$
$$y=12-x$$

Hence, $x\geq 0$, $y\geq 0$

$y=12-x\geq 0$

Greatest value of $y=12$ occurs when $x=0$.

$\mathrm{n}(A\cap B'\cap C)=10-x\geq 0$

$x\leq 10$

Least value of $y=2$ occurs when $x=10$.

Examiners Feedback:

Poorly attempted. The notations used by students were incorrect, reflecting misconceptions and misunderstandings of the concepts. $\mathrm{P}(A\cap C)$ is a probability. It represents the chance that both events A and B happen at the same time. This value can only be between 0 and 1 (inclusive). $\mathrm{n}(A\cap C)$ is a number of outcomes. It tells us how many elements belong to A and B. This value is a whole number.

If A and C are independent, then $\mathrm{P}(A\cap C)=\mathrm{P}(A)\times\mathrm{P}(C)$. Many students assume that if A and C are independent, then $\mathrm{P}(A\cap C)=0$. This is not true.

7 Happie, the owner of a store selling novelty items, is organizing a publicity stunt for his store. Using a stock of ultra-rare Lubaba dolls that he acquired, he sets up a game where a player opens "mystery boxes" to try to find a Lubaba doll hidden inside one of the identical boxes used for the game.

The rules of the game are:

- The player pays an initial \$5 to start the game to open one of n boxes.
- If the opened box is empty, the player can pay an additional \$3 to open a second box.
- If the second box is empty, the player can pay \$2.50 again to open a third box.
- If the third box is still empty, the player loses the game. The doll that is not won will then be donated away and the game is reset for the next player.

As the Lubaba dolls are considered rare collectors' items, it can be assumed that in every game, the players will keep opening the boxes until they win the doll and that they have the means to pay for the maximum of 3 allowable tries at opening the boxes.

Let X be the random variable denoting the number of empty boxes a player has opened in a game.

(a) Show that $\mathrm{P}(X=2)=\dfrac{1}{n}$. [1]

$$\mathrm{P}(X=2)=\frac{n-1}{n}\times\frac{n-2}{n-1}\times\frac{1}{n-2}=\frac{1}{n}$$

Examiners Feedback:
Students have to read context carefully to understand that the number of boxes will reduce by 1 with each game. They also have to understand what X represents to answer the question correctly.

(b) Find the probability distribution of X, leaving your answers in terms of n. [2]

$$\mathrm{P}(X=0)=\frac{1}{n}$$

$$\mathrm{P}(X=1)=\frac{n-1}{n}\times\frac{1}{n-1}=\frac{1}{n}$$

$$\mathrm{P}(X=2)=\frac{n-1}{n}\times\frac{n-2}{n-1}\times\frac{1}{n-2}=\frac{1}{n}$$

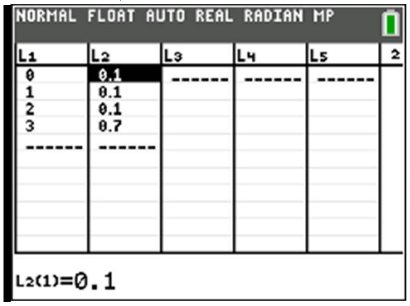
$$\mathrm{P}(X=3)=\frac{n-1}{n}\times\frac{n-2}{n-1}\times\frac{n-3}{n-2}=\frac{n-3}{n}$$

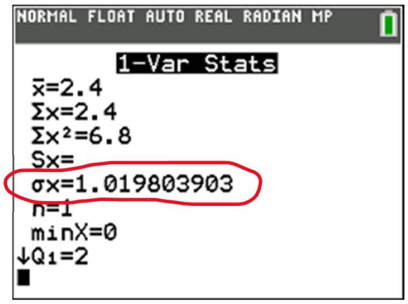
x	0	1	2	3
$\mathrm{P}(X=x)$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{n-3}{n}$

Examiners Feedback:
Similar issue here with the question interpretation. Students need to observe that $X = 0$ is a possible observation.

(c) If $n = 10$, find $\mathrm{P}(X \le \sigma)$, where σ is the standard deviation of *X*. [2]

Method 1: Using GC to find variance
From GC, σ = 1.0198





Hence,
$\mathrm{P}(X \le 1.0198) = \mathrm{P}(X = 0) + \mathrm{P}(X = 1) = 0.2$

x	0	1	2	3
$\mathrm{P}(X = x)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{7}{10}$

Method 2: Using formula $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$

$$\mathrm{E}(X) = 0 \times \frac{1}{10} + 1 \times \frac{1}{10} + 2 \times \frac{1}{10} + 3 \times \frac{7}{10} = 2.4$$

$$\mathrm{E}(X^2) = 0^2 \times \frac{1}{10} + 1^2 \times \frac{1}{10} + 2^2 \times \frac{1}{10} + 3^2 \times \frac{7}{10} = 6.8$$

$$\mathrm{Var}(X) = 6.8 - 2.4^2$$

$$\sigma^2 = 1.04$$

$$\sigma = \sqrt{1.04} = 1.0198$$

$$\mathrm{P}(X \le 1.0198) = \mathrm{P}(X = 0) + \mathrm{P}(X = 1) = 0.2$$

Examiners Feedback:
Students can use GC to obtain standard deviation. Many students mistakenly use normalcdf to calculate probability. This is a discrete random variable, refer to the probability distribution table from part (b).

(d) If the cost price of a Lubaba doll is \$10, find the number of boxes n, that Happie should prepare so that he will not make any profit nor incur any loss from organising this publicity stunt. [3]

Method ①:

Let Y (in dollars) be the amount obtained by Happie in a game.

x	0	1	2	3
y	5	8	10.5	10.5
$\mathrm{P}(Y=y)$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{n-3}{n}$

$$\mathrm{E}(Y)=\left(5\times\frac{1}{n}\right)+\left(8\times\frac{1}{n}\right)+\left(10.5\times\frac{1}{n}\right)+\left(10.5\times\frac{n-3}{n}\right)$$
$$=\frac{10.5n-8}{n}$$

For Happie to not have any profit/loss,

$$\mathrm{E}(Y)-10=0$$
$$\frac{10.5n-8}{n}=10$$
$$n=16$$

Happie should prepare 16 boxes.

Method ②:

Let Y (in dollars) be the amount of "profit" Happie makes in a game.

x	0	1	2	3
y	-5	-2	0.5	0.5
$\mathrm{P}(Y=y)$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{n-3}{n}$

If the doll is not won after 3 tries, it is donated away. Therefore it is still considered as a -\$10 profit to Happie

$$\mathrm{E}(Y)=\left(-5\times\frac{1}{n}\right)+\left(-2\times\frac{1}{n}\right)+\left(0.5\times\frac{1}{n}\right)+\left(0.5\times\frac{n-3}{n}\right)$$
$$=\frac{0.5n-8}{n}$$

For Happie to not have a profit,

$$\mathrm{E}(Y)=0$$
$$\frac{0.5n-8}{n}=0$$
$$n=16$$

Happie should prepare 16 boxes.

Examiners Feedback:

Students need to consider the cost price of Lubaba doll when finding expectation of amount/profit obtained in a game.

8 In each batch of pralines that a chocolate factory produces, they are packed into boxes of 16. Due to the new health guidelines regarding sugar content, quality control tests are implemented and show that on average, the proportion of pralines that exceed the recommended sugar content is p. The number of pralines in a box that exceed the recommended sugar content is denoted by X.

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

- The probability of pralines that exceed the recommended sugar content is constant at p for every praline.
- Whether a praline has exceeded the recommended sugar content is independent of any other pralines.

Examiners Feedback:
Students need to memorise the key words and phrases needed for this question. Do not replace probability with other words which do not have the same meaning, such as proportion, possibility, chance. Do not replace constant with other words which do not have the same meaning, such as equal, same, fixed.

Assume now that X is modelled by a binomial distribution.

(ii) It is known that the most probable number of pralines that exceed the recommended sugar content is 2. Find the exact range of values that p can take. [3]

$$X \sim \mathrm{B}(16, p)$$

Since mode = 2,

$\mathrm{P}(X=2) > \mathrm{P}(X=1)$ and $\mathrm{P}(X=2) > \mathrm{P}(X=3)$

$${}^{16}\mathrm{C}_2 p^2 (1-p)^{14} > {}^{16}\mathrm{C}_1 p (1-p)^{15} \qquad {}^{16}C_2 p^2 (1-p)^{14} > {}^{16}C_3 p^3 (1-p)^{13}$$

$$120 p^2 (1-p)^{14} > 16 p (1-p)^{15} \qquad 120 p^2 (1-p)^{14} > 56 p^3 (1-p)^{13}$$

$$15p > 2(1-p) \qquad 3(1-p) > 14p$$

$$p > \frac{2}{17} \qquad p < \frac{3}{17}$$

$$\therefore \frac{2}{17} < p < \frac{3}{17}$$

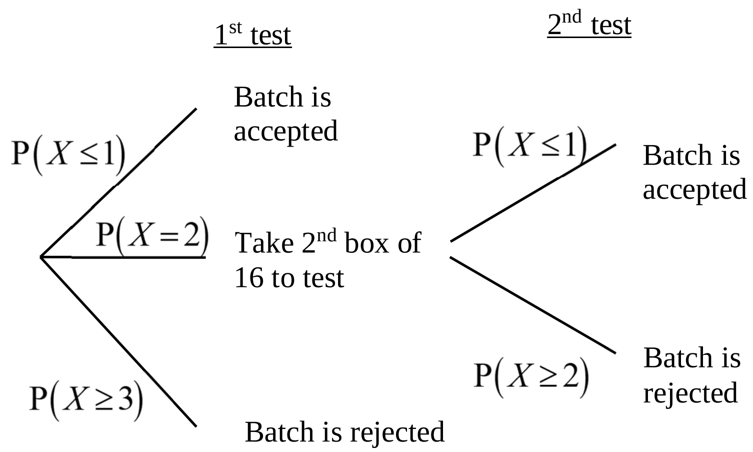
Examiners Feedback:
Answer must be exact as required in the question. Many students failed to write ‘and’ when they write out their inequalities.

For each batch of pralines produced, the chocolate factory implements the following quality control system by testing the first random sample of 16 pralines in a box:

- If there are at most 1 praline that exceed the recommended sugar content in a box, the batch is accepted.
- If there are 3 or more pralines that exceed the recommended sugar content in a box, the batch is rejected.
- If there are 2 pralines that exceed the recommended sugar content in a box, a second box of 16 pralines is tested. The batch will be accepted if there are at most 1 praline that exceed the recommended sugar content in the second box, otherwise the batch will be rejected.

It is given that $p = 0.15$.

(b)(i) Find the probability that the batch of pralines is accepted under this quality control system. [2]



$$\begin{aligned}
\mathrm{P}(\text{batch is accepted}) &= \mathrm{P}(X \le 1) + \mathrm{P}(X = 2)\mathrm{P}(X \le 1) \\
&= \mathrm{P}(X \le 1)\left[1 + \mathrm{P}(X = 2)\right] \\
&= 0.2839012136(1 + 0.2774781077) \\
&= 0.3626775851 \\
&\approx 0.363 \text{ (3 s.f.)}
\end{aligned}$$

Examiners Feedback:
Many students were not able to list out the correct cases. Drawing a tree diagram will help in understanding the scenario better. Some had the correct cases but made calculation errors.

(ii) Given that the batch of pralines is accepted, find the probability that a second box of pralines is tested. [2]

$$\text{P}\left(\text{2nd box is tested}\,\middle|\,\text{batch is accepted}\right) = \frac{\text{P}\left(\text{2nd box is tested} \cap \text{batch is accepted}\right)}{\text{P}\left(\text{batch is accepted}\right)}$$
$$= \frac{\text{P}\left(X = 2\right)\text{P}\left(X \le 1\right)}{0.3626775851}$$
$$= \frac{0.2774781077 \times 0.2839012136}{0.3626775851}$$
$$= 0.2172077205$$
$$\approx 0.217 \text{ (3 s.f.)}$$

Examiners Feedback:
Quite a number of students were not able to identify that this part required the use of conditional probability. Similar to part (b)(i), quite a handful of students who attempted the use of conditional probability did not get the cases correct.

9 A teacher conducts a survey on 7 students to investigate the relationship between the number of hours (h) of "screen-time" per week and the average scores (s) they obtained in a recently concluded examination She records her findings as shown in the following table.

h	8.5	14	20	27	10.5	17	23
s	68	61	44	12	α	58	31

(a) Given that the regression line of s on h is $s = -3.00799h + 100.27974$, show that $\alpha = 67.0$. [2]

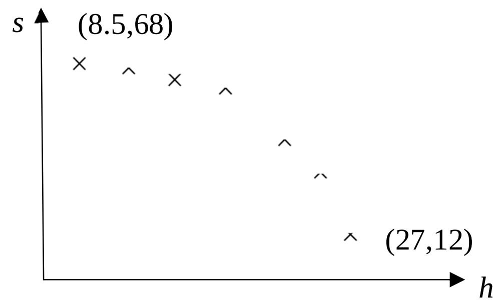
$s = -3.00799h + 100.27974$
From GC, $\bar{h} = 17.1429$

$\bar{s} = -3.00799\bar{h} + 100.27974$
$\bar{s} = -3.00799\left(17.1429\right) + 100.27974 = 48.7140$

$$\frac{68 + 61 + 44 + 12 + \alpha + 58 + 31}{7} = 48.7140$$
$$\alpha = 67.0$$

Examiners Feedback:
Many students substituted the value of h = 10.5 to show $\alpha = 67.0$ instead of using the means of s and h.

(b) Draw a scatter diagram for the data. [1]



Examiners Feedback:
Some common errors are
- did not arrange the points in ascending order and assumed that (23,31) was the last point instead of (27,12)
- labeled the *h*-axis and *s*-axis as *x*-axis and *y*-axis instead
- did not label the minimum and maximum *h* and *s* values

(c) Explain why $s = kh^2 + c$ is the better model compared to the one in part **(a),** by giving appropriate reasons. State the values of *k* and *c*. [3]

r- value between *s* and *h* is –0.961.
r- value between *s* and h^2 is -0.991 which is closer to -1 which means that there is a stronger linear correlation between s and h^2.

Also, from the scatter diagram, as *h* increases, *s* decreases at an increasing rate, which indicates that the curve does not have a linear behaviour.

From G.C., $k = -0.0873$ and $c = 77.7$. Hence $s = -0.0873h^2 + 77.7$

Examiners Feedback:
Most students were able to use the r-values of the 2 regression lines to compare the 2 models. But many did not use the scatter diagram to explain their choice of the better model. There were some calculation errors when finding the regression line of *s* on h^2.

(d) Using the better model, estimate the score a student can expect to obtain if he spends 7 hours of screen-time a week. Comment on the reliability of the estimate obtained. [2]

When $h = 7$,
$s = -0.087331(7)^2 + 77.727 = 73.4$
His predicted score is 73.4. Since $h = 7$ is outside the data range, the estimate is obtained via extrapolation and thus is not reliable.

Examiners Feedback:
Most students managed to get this part correct if they obtained the correct regression line of *s* on h^2 in **(c)**. Only a handful made calculation errors. However, some did not describe the suitability of the estimate clearly, missing out things like extrapolation or range of *h*.

(e) The teacher observes a trend from her findings and concludes with a statement: Increased screen-time will cause the exam scores to decrease. Comment on the validity of the statement. [1]

Correlation is not causation. There may be other factors that contribute to this trend.

Examiners Feedback:
Most students could not comment on the validity of the statement by stating that correlation does not mean causation.

10 Fishing Company A has two types of fishing vessels, the "Standard" vessel and the "Large" vessel. The amount of fish caught by the respective vessels on a typical fishing trip are normally distributed with means and standard deviations as shown in the table.

	Mean (kg)	Standard deviation (kg)
Standard	300	σ
Large	540	110

(a) If a Standard vessel returns with more than 400kg of fish, it is called a Bumper catch. Bumper catches occur 8% of the time. Show that $\sigma \approx 71.2$. [2]

Let S and L be the random variable denoting the amount of fish caught by the Standard and Large vessels respectively.

$$S \sim \mathrm{N}(300,\ \sigma^2)$$
$$L \sim \mathrm{N}(540,\ 110^2)$$

$$\mathrm{P}(S > 400) = 0.08$$
$$\mathrm{P}\left(Z > \frac{400-300}{\sigma}\right) = 0.08$$
$$\frac{400-300}{\sigma} = 1.40507$$
$$\sigma = 71.2\ \ (3\text{ s.f.})(\text{shown})$$

Examiners Feedback:
Most students are able to use standardization to answer this question. If the graphical method is used, then students should present the sketch and indicate the intersection point as the solution.
The table method is not accepted as it is not rigorous enough to show the required value of σ, since σ is not an integer.

(b) On a particular fishing trip, Company A sends out 3 Standard vessels and 2 Large vessels, with a target to catch at least 2100kg of fish. Find the probability that the fishing vessels are able to meet their target. [2]

Let $A = S_1 + S_2 + S_3 + L_1 + L_2$

$$A \sim \mathrm{N}\left(3\times300 + 2\times540,\ 3\times71.171^2 + 2\times110^2\right)$$
$$\sim \mathrm{N}\left(1980,\ 198.484^2\right)$$

$\mathrm{P}(A \geq 2100) = 0.273$

Examiners Feedback:
Most students are able to get this correct. However, one point students should be aware of is that the combined random variable should be "$S_1 + S_2 + S_3 + L_1 + L_2$", instead of "$3S + 2L$" as some students have presented.

(c) State an assumption needed in your calculation in part **(b)**. [1]

The amount of fish caught by all the vessels are independent from one another.

Examiners Feedback:
This is poorly attempted. There are a lot of assumptions given involving Binomial modelling, Random sampling, and Central Limit Theorem.
Students should be aware that since we are combining individual random variables, those individual variables should all be independent from one another, i.e S_1 is independent of S_2 and L_1 alike etc.
Their descriptions should convey this meaning, and therefore these variations listed below are not accepted:
- Catch from small vessel is independent of catch from large vessel.
- Fish caught is independent of other fish.
- Vessels are independent of one another.

Fishing Company B operates in the same seas as Fishing Company A.
However, Company B has the improved versions of the two types of fishing vessels that Company A has: the "Premium-Standard" vessel and the "Premium-Large" vessel. In a fishing trip, the Premium-Standard vessel is capable of catching 3 times as much fish as the Standard vessel and the Premium-Large vessel is capable of catching 2 times as much fish as Large vessel.

(d) Company B sends out 1 Premium-Standard vessel and 1 Premium-Large vessel, with the same target to catch at least 2100kg of fish. Find the probability that these two fishing vessels are able to meet their target. [2]

Let $B = 3S + 2L$

$$B \sim \mathrm{N}\left(3\times300 + 2\times540,\ 3^2\times71.171^2 + 2^2\times110^2\right)$$
$$\sim \mathrm{N}\left(1980,\ 306.574^2\right)$$

$\mathrm{P}(B \geq 2100) = 0.348$

Examiners Feedback:
Most students are able to get this correct. Those who did not get the full credit had either not calculated the variance correctly or did not input the GC correctly.

(e) It is found that, in a sample of n randomly chosen fishing trips, the probability that the average catch of a Large vessel being less than 552kg is at least 0.7.
Find the least possible value of n. [3]

$$\overline{L} \sim \mathrm{N}\left(540,\ \frac{110^2}{n}\right)$$

$\mathrm{P}(\overline{L} < 552) \geq 0.7$

<u>Method ①:</u>

n	$\mathrm{P}\left(\overline{L} < 552\right)$
23	0.6996
24	0.7035
25	0.7073

From GC, least value of n is 24

<u>Method ②:</u>
Standardizing,

$$\mathrm{P}\left(Z < \frac{552-540}{\sqrt{110^2/n}}\right) \geq 0.7$$

$$\mathrm{P}\left(Z < \frac{6\sqrt{n}}{55}\right) \geq 0.7$$

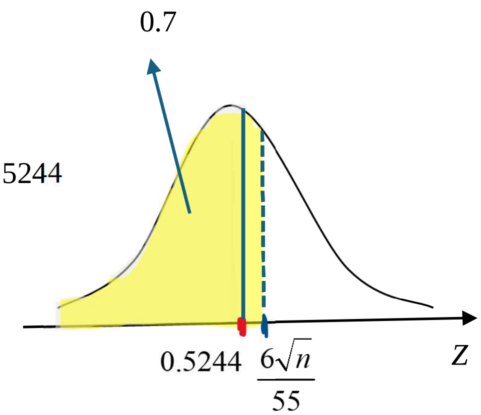
Consider $\mathrm{P}(Z < k) = 0.7 \quad \Rightarrow \quad k = 0.5244$

Hence $\dfrac{6\sqrt{n}}{55} \geq 0.5244$

$6\sqrt{n} \geq 28.842$

$n \geq 23.1$

$\therefore$ least value of n is 24.



Examiners Feedback:
Not well attempted. Some students misinterpret the question as a Binomial distribution question when it should be a Sampling distribution question.
For students who did not involve inequalities in their calculations for the least value of n, they should provide justification why the answer is obtained by rounding up instead of rounding off (by using diagram or otherwise).

11 EastHam is a football club in a competitive league. It is found that the average time taken by EastHam to score their first goal in a match is 50 minutes. The coach trialled a new play tactic and the time taken, x minutes, to score the first goal is recorded for a random sample of 35 matches in the season. The total time taken is found to be 1650 minutes and the variance of the time is 250 minutes^2.

The coach wants to test if the average time taken to score the first goal has decreased with the implementation of the new play tactic.

(a) Explain why the coach is able to carry out a hypothesis test without any assumption about the distribution of the time taken to score the first goal. [1]

The sample mean time for 35 matches is approximated to have a normal distribution by the Central Limit Theorem, since sample size is sufficiently large. Hence there is no need to make any assumptions about the distribution of the time taken to score the first goal.

Examiners Feedback:
Not well attempted by students in general. Significant number of students missed out on at least one of the following phrases: (a) sample size is sufficiently large, (b) approximated to have a normal distribution (c) by Central Limit Theorem.

(b) Find unbiased estimates of the population mean and variance. Give your answers correct to 3 decimal places. [2]

Unbiased estimate for population mean $=\dfrac{1650}{35}=47.143$ (to 3d.p.)

Unbiased estimate for population variance is $s^2=\dfrac{35}{34}\left(250\right)=257.353$ (to 3d.p.)

Examiners Feedback:
Unbiased estimate for population mean is correctly done by most students. However, students are not sensitive to the requirement of the question to leave the answer to 3 decimal places and hence got no marks for this. Unbiased estimate for population variance is poorly attempted. Students did not realise that the variance value of 250 that was given in the question was actually sample variance. Hence the unbiased estimate for population variance = $\dfrac{n}{n-1}\left(\text{sample variance}\right)$

(c) Carry out the test at 10% level of significance. You should state your hypotheses and define any symbols you use. [4]

Let X be the random variable denoting time taken by EastHam to score the first goal in a match.

$H_0 : \mu = 50$
$H_1 : \mu < 50$, where μ is the population mean time taken by EastHam to score the first goal in a match

Under H_0,
Since n is large,

$$\bar{X} \sim N\left(50, \frac{257.353}{35}\right) \text{ approximately by CLT}$$

$$Z = \frac{\bar{X} - 50}{\sqrt{s^2/n}} \sim N(0,1)$$

Method ①:
By GC, p-value $= 0.146\,(3\text{ s.f.})$.

Since p-value > 0.1, we do not reject H_0 and conclude that there is insufficient evidence at 10% significance level to say that average time taken by EastHam to score the first goal has decreased.

Method ②:

$$z_{\text{test}} = \frac{\left(\frac{1650}{35}\right) - 50}{\sqrt{\frac{4375/17}{35}}} = -1.05366, \; z_{\text{test}} = -1.28155$$

Since $z_{\text{test}} > z_{\text{critical}}$, we do not reject H_0 and conclude that there is insufficient evidence at 10% significance level to say that average time taken by EastHam to score the first goal has decreased.

Examiners Feedback:
Question was not well-attempted by students as they did not memorise the fixed template when answering Hypothesis Testing question. These include:
(a) Defining variable for the question, defining μ ,
(b) Stating the null and alternative hypothesis,
(c) Under Ho, stating the distribution of $\bar{X}$ as well as the test statistics, $Z = \frac{\bar{X} - \mu}{\sqrt{s^2/n}} \sim N(0,1)$,
(d) Finding p-value,
(e) Stating whether to reject/do not reject H_o and giving appropriate conclusion.

- When stating the distribution of $\bar{X}$, students should note that $\bar{X} \sim N(\mu, \frac{s^2}{n})$. Many students wrote the value of $\bar{x}$ instead.

- The distribution of X was not given in the question. Students need to state explicitly the conditions for $\overline{X}$ to be **<u>approximated</u>** to normal distribution by Central Limit Theorem.
- Conclusion for the hypothesis testing is mixed up.
  (a) **Do not** reject H_o. There is **insufficient** evidence at … **% level of significance** to conclude (H_1 statement in the context of the question).
  (b) Reject H_o. There is **sufficient** evidence at … **% level of significance** to conclude (H_1 statement in the context of the question).

A fan of another football club in the same league, NewPalace, claims that the average time taken by NewPalace to score their first goal in a match is 40 minutes. The time taken by NewPalace to score the first goal in a match can be assumed to be normally distributed with variance 280 minutes^2.

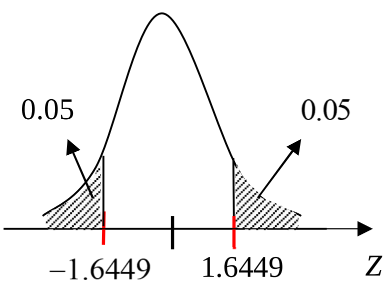
(d) A random sample of 50 matches is taken and a hypothesis test is carried out on whether the fan's claim is valid. Find the range of values of the mean time of this sample for which the fan's claim would be rejected at 10% level of significance. [4]

Let Y be the random variable denoting time taken by NewPalace to score the first goal in a match.

$H_0 : \mu = 40$

$H_1 : \mu \neq 40$

Under H_0, $\overline{Y} \sim N\left(40, \frac{280}{50}\right)$



Since H_0 is rejected,

$$\frac{\overline{y}-40}{\sqrt{280/50}} < -1.6449 \quad \text{or} \quad \frac{\overline{y}-40}{\sqrt{280/50}} > 1.6449$$

$$\overline{y} < 36.1\,(3\text{ s.f.}) \quad \text{or} \quad \overline{y} > 43.9\,(3\text{ s.f.})$$

Hence, range of values of $\overline{y}$ is: $0 < \overline{y} < 36.1$ or $\overline{y} > 43.9$

Examiners Feedback:
Poorly attempted by students.
Few things to note:
- when considering two tailed z-test, probability for each tail can be found by dividing the level of significance by two. Hence for this question, the probability for each tail is 0.05.
- Since probability is 0.05 for each tail, we can use inverse norm to calculate the value of $z_{\text{critical}} = -1.6449, 1.6449$.

Conclusion is to reject H_o. Thus $z_{\text{test}} \leq -1.6449$ or $z_{\text{test}} \geq 1.6449$ so that the test statistics falls inside the rejection region.