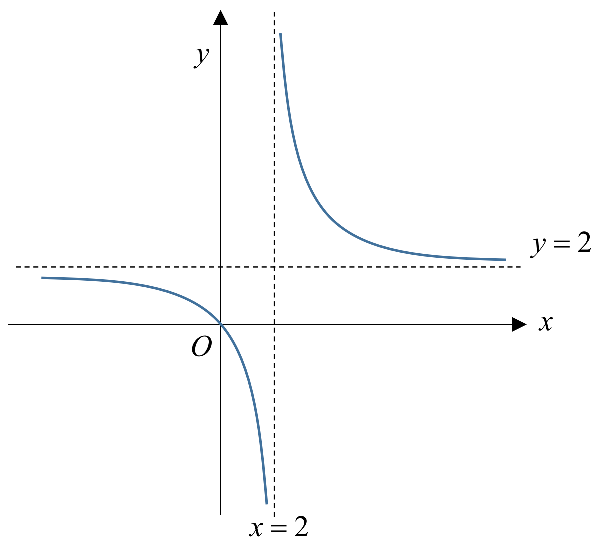
RAFFLES INSTITUTION
2024 Year 6 H2 Mathematics Prelim Exam Paper 2
Questions and Solutions with comments

1 The function f is defined by $\mathrm{f}: x \mapsto \dfrac{2x}{x-2}$, for $x \in \mathbb{R}$, $x \neq 2$.

(a) Sketch the graph of f and find its range. [3]

Another function g is defined by $\mathrm{g}: x \mapsto 3+|x+2|$, for $x \in \mathbb{R}$.

(b) Show that the composite function fg exists. Find $\mathrm{fg}(x)$ and state the domain and range of fg. [5]

(a) **[3]**	 Range of f $= (-\infty, \infty) \setminus \{2\}$	Note that the curve passes through origin. Other possible notations for range of f: • $(-\infty, \infty) \setminus \{2\}$ • $\mathbb{R} \setminus \{2\}$
(b) **[5]**	$R_g = [3, \infty)$ $D_f = (-\infty, 2) \cup (2, \infty)$ Since $R_g \subseteq D_f$, function fg exists. $\mathrm{fg}(x) = \mathrm{f}(3+\lvert x+2\rvert) = \dfrac{2(3+\lvert x+2\rvert)}{3+\lvert x+2\rvert-2} = \dfrac{6+2\lvert x+2\rvert}{1+\lvert x+2\rvert}$ $D_{fg} = \mathbb{R}$ Note that $\mathrm{f}(3) = \dfrac{2(3)}{3-2} = 6$. $D_{fg} = D_g$ R_g R_{fg} $\mathbb{R} \xrightarrow{\ \mathrm{g}\ } [3, \infty) \xrightarrow{\ \mathrm{f}\ } (2, 6]$ Hence, $R_{fg} = (2, 6]$.	Give domain and range in set notations. Some students are confused about the domain and range of a composite function.

2 The function f is defined by $\mathrm{f}(z)=z^4+Az^3+Bz^2+Cz+45$, where A, B and C are real numbers. Given that $2+\mathrm{i}$ is a root of $\mathrm{f}(z)=0$ and $(z-k)^2$ is a factor of $\mathrm{f}(z)$, where k is a positive real number, find the values of A, B, C and k. [5]

[5]	Since all coefficients of $\mathrm{f}(z)$ are real and $2+\mathrm{i}$ is a root of $\mathrm{f}(z)=0$, then $2-\mathrm{i}$ is also a root of the equation. The quadratic factor of the equation is $\left(z-(2+\mathrm{i})\right)\left(z-(2-\mathrm{i})\right)=z^2-4z+5$ Then, $z^4+Az^3+Bz^2+Cz+45=\left(z-k\right)^2(z^2-4z+5)$ $=\left(z-2kz+k^2\right)(z^2-4z+5)$ Comparing constants, $45=5k^2 \quad \Rightarrow \quad k=\pm 3$ Since $k>0$, therefore, $k=3$. So, $z^4+Az^3+Bz^2+Cz+45=\left(z-6z+9\right)(z^2-4z+5)$ $=z^4-10z^3+38z^2-66z+45$ $A=-10, B=38, C=-66$	This question is generally well done. Some students tried to substitute the root in, then compare the real and imaginary parts to obtain 2 equations. However, 2 equations are not enough to solve for 3 unknowns, without using the condition that $(z-k)^2$ is also a factor of $\mathrm{f}(z)$.

3 **(a)** The points A, B and C on the plane π have position vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ respectively. Show that a vector perpendicular to π is parallel to

$$\mathbf{b}\times\mathbf{c}+\mathbf{c}\times\mathbf{a}+\mathbf{a}\times\mathbf{b}.$$ [3]

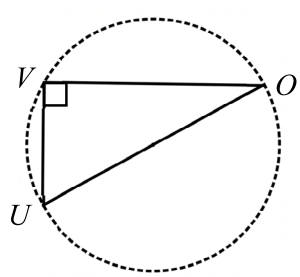
(b) $\mathbf{p}$ and $\mathbf{q}$ are non-zero vectors and $\mathbf{p}=(\mathbf{p}\cdot\mathbf{q})\mathbf{q}$.

(i) Find the relationship between $\mathbf{p}$ and $\mathbf{q}$. [1]

(ii) Find $|\mathbf{q}|$. [1]

(c) $\mathbf{u}$ is the position vector of a fixed point U relative to a fixed origin O. A variable point V has position vector $\mathbf{v}$ relative to O.

Given that $\mathbf{v}\cdot(\mathbf{v}-\mathbf{u})=0$, describe geometrically the set of all possible positions of the point V. [2]

(a) **[3]**	$\overrightarrow{AB}\times\overrightarrow{BC}=(\mathbf{b}-\mathbf{a})\times(\mathbf{c}-\mathbf{b})$ $=\mathbf{b}\times\mathbf{c}-\mathbf{b}\times\mathbf{b}-\mathbf{a}\times\mathbf{c}+\mathbf{a}\times\mathbf{b}$ $=\mathbf{b}\times\mathbf{c}-(-\mathbf{c}\times\mathbf{a})+\mathbf{a}\times\mathbf{b}$, since $\mathbf{b}\times\mathbf{b}=\mathbf{0}$ $=\mathbf{b}\times\mathbf{c}+\mathbf{c}\times\mathbf{a}+\mathbf{a}\times\mathbf{b}$ Since $\overrightarrow{AB}$ and $\overrightarrow{BC}$ are parallel to π but not parallel to each other, so $\overrightarrow{AB}\times\overrightarrow{BC}$ is a vector perpendicular to both $\overrightarrow{AB}$ and $\overrightarrow{BC}$ and hence perpendicular to π. Therefore a vector perpendicular to π is parallel to $\mathbf{b}\times\mathbf{c}+\mathbf{c}\times\mathbf{a}+\mathbf{a}\times\mathbf{b}$ (Shown).	Some students did not show working clearly for a "show" question. Some students started from the given expression which is to "verify" not to "show" Some students find the cross product of two position vectors which are not parallel to the plane.
(b)(i) **[1]**	$\mathbf{p}=(\mathbf{p}\cdot\mathbf{q})\mathbf{q}\Rightarrow\mathbf{p}\,//\,\mathbf{q}$ since $\mathbf{p}\cdot\mathbf{q}$ is a scalar. Hence $\mathbf{p}$ is parallel to $\mathbf{q}$.	Note that parallel vectors may not be in the same direction, it can be opposite directions, since $\mathbf{p}\cdot\mathbf{q}$ could be negative.
(b)(ii) **[1]**	$\mathbf{p}=(\mathbf{p}\cdot\mathbf{q})\mathbf{q}\Rightarrow|\mathbf{p}|=|(\mathbf{p}\cdot\mathbf{q})\mathbf{q}|$ $\Rightarrow|\mathbf{p}|=|\mathbf{p}\cdot\mathbf{q}||\mathbf{q}|$ $=|\mathbf{p}||\mathbf{q}||\mathbf{q}|$ $\Rightarrow|\mathbf{q}|^2=1$ $\Rightarrow|\mathbf{q}|=1$	$|\mathbf{q}|=\pm1$ is not accepted since the magnitude of a vector is positive.
(c) **[2]**	$\mathbf{v}\cdot(\mathbf{v}-\mathbf{u})=0\Rightarrow\mathbf{v}\perp(\mathbf{v}-\mathbf{u})$ $\Rightarrow\overrightarrow{OV}\perp\overrightarrow{UV}$  Since V is a variable point, so the set of all possible positions of the point V forms a sphere with OU as the diameter.	It would be a circle if the vectors are 2-dimensional, so it is also accepted.

4 **(a)** Given that $y=\mathrm{e}^{\sqrt{1+2x}}$, show that

$$(1+2x)\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2=y^2 \quad \text{and} \quad (1+2x)\frac{\mathrm{d}^2y}{\mathrm{d}x^2}+\frac{\mathrm{d}y}{\mathrm{d}x}=y. \qquad [3]$$

(b) By further differentiation, obtain the series expansion for y in terms of x up to and including the term in x^3. [3]

(c) Verify that the same series expansion for y in part **(b)** is obtained if the standard series expansions for e^x and $(1+x)^n$ are used. [4]

<table>
<tr>
<td>(a)
[3]</td>
<td>

$y=\mathrm{e}^{\sqrt{1+2x}}$

Method 1

$y=\mathrm{e}^{\sqrt{1+2x}}$

$\ln y=(1+2x)^{\frac{1}{2}}$

Differentiating wrt x,

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{1}{2}(1+2x)^{-\frac{1}{2}}(2)=(1+2x)^{-\frac{1}{2}}$$

$$(1+2x)^{\frac{1}{2}}\frac{\mathrm{d}y}{\mathrm{d}x}=y$$

$$(1+2x)\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2=y^2 \quad \text{(shown)}$$

Method 2

$$y=\mathrm{e}^{\sqrt{1+2x}}$$

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{1}{2\sqrt{1+2x}}(2)\mathrm{e}^{\sqrt{1+2x}}$$

$$\sqrt{1+2x}\frac{\mathrm{d}y}{\mathrm{d}x}=\mathrm{e}^{\sqrt{1+2x}}$$

$$(1+2x)\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2=y^2 \quad \text{(shown)}$$

Differentiating again wrt x,

$$(1+2x)(2)\frac{\mathrm{d}y}{\mathrm{d}x}\left(\frac{\mathrm{d}^2y}{\mathrm{d}x^2}\right)+2\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2=2y\frac{\mathrm{d}y}{\mathrm{d}x}$$

$$(1+2x)\frac{\mathrm{d}^2y}{\mathrm{d}x^2}+\frac{\mathrm{d}y}{\mathrm{d}x}=y \quad \text{since } \frac{\mathrm{d}y}{\mathrm{d}x}\neq 0 \quad \text{(shown)}$$

</td>
<td>

Most students have done well for the first part of the question.

Students must take note that $\left(\mathrm{e}^{\sqrt{1+2x}}\right)^2\neq \mathrm{e}^{1+2x}$.

Recall that $(a^m)^n=a^{mn}$

Unfortunately, many students did not realise that they could attempt this part using implicit differentiation. They attempted by finding $\frac{\mathrm{d}y}{\mathrm{d}x}$ and $\frac{\mathrm{d}^2y}{\mathrm{d}x^2}$ in terms of x first, resulting in a need for tedious simplification.

</td>
</tr>
<tr>
<td>(b)
[3]</td>
<td>

Differentiating again wrt x,

$$(1+2x)\frac{\mathrm{d}^3y}{\mathrm{d}x^3}+2\frac{\mathrm{d}^2y}{\mathrm{d}x^2}+\frac{\mathrm{d}^2y}{\mathrm{d}x^2}=\frac{\mathrm{d}y}{\mathrm{d}x}$$

$$(1+2x)\frac{\mathrm{d}^3y}{\mathrm{d}x^3}+3\frac{\mathrm{d}^2y}{\mathrm{d}x^2}=\frac{\mathrm{d}y}{\mathrm{d}x}$$

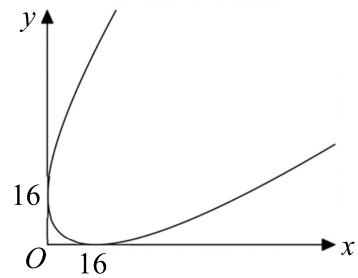
</td>
<td>

Generally, the students are able to complete part **(b)** well.

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

	When $x=0,\quad y=\mathrm{e},\ \frac{\mathrm{d}y}{\mathrm{d}x}=\mathrm{e},\ \frac{\mathrm{d}^2y}{\mathrm{d}x^2}=0,\ \frac{\mathrm{d}^3y}{\mathrm{d}x^3}=\mathrm{e}.$ $y=\mathrm{e}+\mathrm{e}x+(0)\frac{x^2}{2!}+\mathrm{e}\frac{x^3}{3!}+\cdots=\mathrm{e}\left(1+x+\frac{x^3}{6}+\cdots\right)$	
(c) **[4]**	$(1+2x)^{\frac{1}{2}}=1+\frac{1}{2}(2x)+\frac{1}{2}\left(-\frac{1}{2}\right)\frac{(2x)^2}{2}+\frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\frac{(2x)^3}{3!}+\ ...$ $=1+x-\frac{1}{2}x^2+\frac{1}{2}x^3+\ ...$ **<u>Method 1</u>** $\mathrm{e}^{\sqrt{1+2x}}=\mathrm{e}^{1+x-\frac{1}{2}x^2+\frac{1}{2}x^3+...}$ $=\mathrm{e}^1\mathrm{e}^{x-\frac{1}{2}x^2+\frac{1}{2}x^3+...}$ $=\mathrm{e}\left(1+\left(x-\frac{1}{2}x^2+\frac{1}{2}x^3+...\right)+\frac{1}{2!}\left(x-\frac{1}{2}x^2+...\right)^2+\frac{1}{3!}(x+...)^3+...\right)$ $=\mathrm{e}\left(1+x-\frac{1}{2}x^2+\frac{1}{2}x^3+\frac{1}{2}\left(x^2-x^3\right)+\frac{1}{6}x^3+...\right)$ $=\mathrm{e}\left(1+x+\frac{1}{6}x^3+...\right)$ (same results as in part **(a)**) **<u>Method 2</u>** $\mathrm{e}^{\sqrt{1+2x}}=\mathrm{e}^{1+x-\frac{1}{2}x^2+\frac{1}{2}x^3+...}$ $=\mathrm{e}^1\mathrm{e}^x\mathrm{e}^{-\frac{1}{2}x^2}\ \mathrm{e}^{\frac{1}{2}x^3}\ \cdots$ $=\mathrm{e}\left(1+x+\frac{1}{2!}x^2+\frac{1}{3!}x^3+\cdots\right)\left(1+\left(-\frac{1}{2}x^2\right)+\cdots\right)\left(1+\left(\frac{1}{2}x^3\right)+\cdots\right)\cdots$ $=\mathrm{e}\left(1+x+\frac{1}{2!}x^2-\frac{1}{2}x^2+\frac{1}{3!}x^3-\frac{1}{2}x^3+\frac{1}{2}x^3+...\right)$ $=\mathrm{e}\left(1+x+\frac{1}{6}x^3+...\right)$ (same results as in part **(a)**)	Most of the students are able to expand $(1+2x)^{\frac{1}{2}}$ correctly Many students have tried to expand $\mathrm{e}^{\sqrt{1+2x}}=\mathrm{e}^{1+x-\frac{1}{2}x^2+\frac{1}{2}x^3+...}$ as $1+\left(1+x-\frac{1}{2}x^2+\frac{1}{2}x^3+...\right)+\frac{1}{2!}\left(1+x-\frac{1}{2}x^2+\frac{1}{2}x^3+...\right)^2+\frac{1}{3!}\left(1+x-\frac{1}{2}x^2+\frac{1}{2}x^3+...\right)^3+...$ They did not realise that every subsequent term of this expansion will need to be taken in to account in order to obtain the correct constant term and the coefficients of x, x^2 and x^3 terms.

5



The diagram shows the curve C with parametric equations

$$x=(1+t)^2, \quad y=(3-t)^2.$$

The curve C meets the axes at $(16,0)$ and $(0,16)$.

(a) Show that the line $x=16$ meets C at the point P where $t=-5$. [1]

The normal to C at P is denoted by l.

(b) Find the cartesian equation of l. [3]

(c) The line l meets C again at the point Q where $x=b$. Show that the area of the region bounded by l, the lines $x=16$, $x=b$ and the x-axis is $\frac{266240}{81}$ units^2. [3]

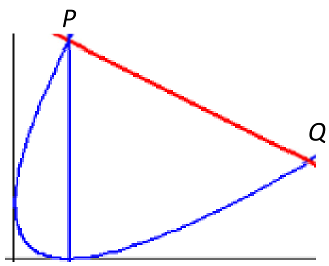
(d) Show that the area (in units^2) of the region bounded by C and l can be given by

$$\frac{266240}{81}+\int_c^d \mathrm{f}(t)\,\mathrm{d}t,$$

where $\mathrm{f}(t)$, and the constants c and d are to be determined.

Hence find the value of this area. [3]

(a) **[1]**	When the line $x=16$ meets C, $(1+t)^2=16$ $1+t=\pm 4$ $t=3$ or -5 Since $t=3$ gives the point $(16,0)$, so $t=-5$ at P.	To "show", need to <u>find</u> the answer assuming answer is not given. Substituting $t=-5$ into the equation to check the statement is true can only be used if the question says "verify".
(b) **[3]**	$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\frac{\mathrm{d}y}{\mathrm{d}t}}{\frac{\mathrm{d}x}{\mathrm{d}t}}=\frac{-2(3-t)}{2(1+t)}$ At $t=-5$, $x=16$, $y=64$, gradient of normal is $\frac{-1}{2}$ l: $y-64=\frac{-1}{2}(x-16) \Rightarrow y=\frac{-1}{2}x+72$	Need to be careful when simplifying an expression involving negative signs.

(c) **[3]**	$(3-t)^2=\frac{-1}{2}(1+t)^2+72$ $2\left(t^2-6t+9-72\right)=-t^2-2t-1$ $3t^2-10t-125=0$ $(t+5)(3t-25)=0$ since $t=-5$ at P At Q, $t=\frac{25}{3}$, $x=(1+t)^2=\frac{784}{9}$ **Method 1** Required area is $\int_{16}^{\frac{784}{9}} \frac{-1}{2}x+72 \, \mathrm{d}x=\frac{266240}{81}$ using GC **Method 2** At Q, $t=\frac{25}{3}$, $y=(3-t)^2=\frac{256}{9}$. Required area = Area of Trapezium $=\frac{1}{2}\left(\frac{784}{9}-16\right)\left(64+\frac{256}{9}\right)$ $=\frac{266240}{81}$	To find intersection, substitute parametric equations into cartesian equation. No need to find the corresponding cartesian equation of the parametric equations. Note that the question asks for area under the normal line, not area under curve.
(d) **[3]**	 The curve meets the y-axis at $t=-1$. The required area is: Area under l as found in **(c)** + Area under C from $t=-1$ to -5 $-$ Area under C from $t=-1$ to $\frac{25}{3}$ $=\frac{266240}{81}+\int_{x=0,\,t=-1}^{x=16,\,t=-5} y\,\mathrm{d}x-\int_{x=0,\,t=-1}^{x=\frac{784}{9},\,t=\frac{25}{3}} y\,\mathrm{d}x$ $=\frac{266240}{81}+\int_{x=\frac{784}{9},\,t=\frac{25}{3}}^{x=16,\,t=-5} y\,\mathrm{d}x$ $=\frac{266240}{81}+\int_{\frac{25}{3}}^{-5}(3-t)^2\cdot 2(1+t)\,\mathrm{d}t$ $=\frac{256000}{81}$ using GC	Question asks to "show", so working must be clear. Since we need to express in one integral and the trapezium is bounded by the x-axis, we should consider the different areas under the curve bounded by the x-axis so that they can be combined into one integral.

6 Eleven cards each bears a single letter and together they can be made to spell the word COFFEEHOUSE. The 11 cards are arranged in a row.

(a) Find the number of different arrangements that can be made. [1]

(b) Find the number of different arrangements in which the 2 F's are next to each other and no E's are next to each other. [3]

Three cards are selected from the eleven cards and the order of selection is not relevant. Find the number of possible selections that can be made

(c) if the three cards all bear different letters, [1]

(d) if exactly two of the three cards bear the same letter. [2]

(a) **[1]**	3 E, 2 F, 2 O, 1 C, 1 H, 1 U, 1 S Number of arrangements $=\dfrac{11!}{3!2!2!}=1663200$	Be careful when counting the number of repeated letters. Some students counted wrongly and the entire question was affected.
(b) **[3]**	**Method 1** _ (FF) _ O _ O _ C _ H _ U _ S _ Put the 3 E aside first. Group the 2 Fs as 1 unit and arrange it together with the other 6 letters. Then slot in the 3 E. Number of different arrangements where the 2 F are together $=\dfrac{7!}{2!}=2520$ Number of ways to slot in the 3 E in the 8 possible slots $={}^{8}C_{3}=56$ $\therefore$ Required number of arrangements $=2520\times 56=141120$ **Method 2** Number of arrangements where the 2 F are together $=\dfrac{10!}{2!3!}=302400$ Number of arrangements where the 2 F are together and 2 E are together $=\dfrac{7!}{2!}\times {}^{8}C_{2}\times 2!=141120$ (note that the 2 Es that are together must be separated from the single E. So, arrange all other letters first then slot in the 2 Es and the single E) Number of arrangements where the 2 F are together and the 3 E are together $=\dfrac{8!}{2!}=20160$ Using complement method: Required number of arrangements $=302400-141120-20160=141120$	Drawing a simple diagram as shown helps in your counting process. For P&C questions, there are often numerous ways to solve the question. For this question, the first method is the fastest – employing the "grouping" and "slotting" method. All other methods are tedious and can be tricky. Some students did the alternative method with little success. Even if they succeeded, they have spent way too much time on a 3 marks question.

(c) **[1]**	There are 7 different letters to chose from. Number of selections such that the 3 cards all bear different letters $= {}^{7}C_{3} = 35$	Do read the question carefully. It was stated that **the order of selection is not relevant** for part **(c)** and **(d)**.
(d) **[2]**	Case 1 : E, E, _ (third card is not E) Case 2 : F, F, _ Case 3 : O, O, _ For each case, there are 6 choices for the third card. $\therefore$ Number of selections $= 3 \times 6 = 18$	Learn to explain your answer. In this case, list down the cases clearly. No method marks can be awarded if your answer is wrong and there was no explanation of how the string of numbers came about.

7 The probability of obtaining a head when a particular coin is tossed is p. A fair cubical die has the number '1' on one face, number '2' on two faces and number '3' on three faces.

The coin and die are thrown simultaneously. The random variable X is defined as follows.

If the coin shows a head, then X is thrice the score on the die.
If the coin shows a tail, then X is the score on the die.

(a) Show that $\mathrm{P}(X=3)=\frac{1}{2}-\frac{1}{3}p$, and find the probability distribution of X. [3]

(b) Given that $\mathrm{E}(X)=5$, find the exact value of p. [2]

(c) Using the value of p found in part **(b)**, find the exact value of $\mathrm{Var}(X)$. [3]

(a) **[3]**	$\mathrm{P}(X=3)$ $=\mathrm{P}(\text{tail and face is '3'})+\mathrm{P}(\text{head and face is '1'})$ $=(1-p)\frac{1}{2}+p\left(\frac{1}{6}\right)=\frac{1}{2}-\frac{1}{3}p$ (shown) $\mathrm{P}(X=1)=\mathrm{P}(\text{tail and face is '1'})=\frac{1}{6}(1-p)$ $\mathrm{P}(X=2)=\mathrm{P}(\text{tail and face is '2'})=\frac{1}{3}(1-p)$ $\mathrm{P}(X=6)=\mathrm{P}(\text{head and face is '2'})=\frac{1}{3}p$ $\mathrm{P}(X=9)=\mathrm{P}(\text{head and face is '3'})=\frac{1}{2}p$ The probability distribution of X is: x: 1, 2, 3, 6, 9 $\mathrm{P}(X=x)$: $\frac{1}{6}(1-p)$, $\frac{1}{3}(1-p)$, $\frac{1}{2}-\frac{1}{3}p$, $\frac{1}{3}p$, $\frac{1}{2}p$	As with all word problems, do take time to read the question carefully to understand the scenario before attempting the question. Getting the correct probability distribution of X is crucial – any wrong probability will affect subsequent parts.
(b) **[2]**	$\mathrm{E}(X)=5$ $\frac{1}{6}(1-p)+2\left(\frac{1}{3}(1-p)\right)+3\left(\frac{1}{2}-\frac{1}{3}p\right)+6\left(\frac{1}{3}p\right)+9\left(\frac{1}{2}p\right)=5$ $\frac{7}{3}+\frac{14}{3}p=5$ $p=\frac{4}{7}$	Recall : $\mathrm{E}(X)=\sum x\mathrm{P}(X=x)$ Generally well done.
(c) **[3]**	x: 1, 2, 3, 6, 9 $\mathrm{P}(X=x)$: $\frac{1}{14}$, $\frac{1}{7}$, $\frac{13}{42}$, $\frac{4}{21}$, $\frac{2}{7}$ $\mathrm{E}\left(X^2\right)=\frac{1}{14}+2^2\left(\frac{1}{7}\right)+3^2\left(\frac{13}{42}\right)+6^2\left(\frac{4}{21}\right)+9^2\left(\frac{2}{7}\right)=\frac{234}{7}$ $\mathrm{Var}\left(X\right)=\mathrm{E}\left(X^2\right)-\left[\mathrm{E}(X)\right]^2=\frac{234}{7}-5^2=\frac{59}{7}$	As the exact value of $\mathrm{Var}(X)$ was required, the use of the graphing calculator is not allowed. Hence working showing how $\mathrm{E}\left(X^2\right)$ was obtained is required.

8 **(a)** It is given that X is the number of times a student is late for school in a year and Y is the student's performance in the Mathematics Examination. The product moment correlation coefficient of a bivariate sample (x_i, y_i), $i = 1, 2, \ldots, n$, for n students is r.

State, giving a reason, whether each of the following statements is true or false.

(i) When the value of r is zero, it can be implied that the variables X and Y are not related. [1]

(ii) When the value of r is -1, it can be implied that late-coming causes poor performance in the subject. [1]

(b) The table below shows the daily sales of cups of iced coffee in a week by a shop and the maximum daily temperature.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Temperature ($°C$), t	30.5	31.3	31.9	34.8	25.9	28.5	29.2
Daily sales, y	95	101	115	79	81	86	88

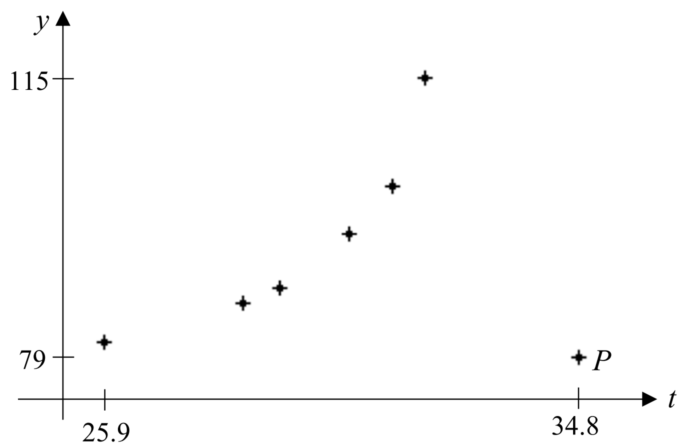
(i) Sketch a scatter diagram of y against t, labelling the axes clearly. [2]

One of the values of y appears to be incorrect.

(ii) Indicate the corresponding point on your diagram by labelling it P. Omitting P, find the equation of the least squares regression line of $\frac{1}{y}$ on t, and the value of the product moment correlation coefficient between $\frac{1}{y}$ and t. Comment on this value. [5]

(iii) Use an appropriate regression line to give an estimate of the daily sales when the temperature is 20.4 $°C$.
State, with a reason, whether the estimate is reliable. [2]

(a)(i) **[1]**	False. $r = 0$ indicates that X and Y are not linearly correlated, but they may have a non-linear relation.	Students are reminded to state whether the statement is *True* or *False*. It is important to remember that the value of r is a measurement of the **linear** correlation between two variables.
(a)(ii) **[1]**	False. The value of r measures strength of linear correlationship, it does not not indicate any causal relationship betweeen X and Y.	Similarly, the value of r does not imply any causal relationship or dependence between the two variables.

(b)(i) **[2]**		Students are reminded to label the max/min end points on both the axes. Students should also take note of the relative positions of the data points. (Note: the data points are not equally spaced along the horizontal axis)
(b)(ii) **[5]**	Label P (see diagram in **(b)(i)**). From GC, the least squares regression line $\frac{1}{y}$ on t is $\frac{1}{y} = -0.0005611311906t + 0.0273247901$ $\frac{1}{y} = -0.000561t + 0.0273$ (3 s.f.) $r = -0.934$ is very close to -1. Hence, there is a *strong*, *negative* linear correlation between $\frac{1}{y}$ and t.	Students are reminded to leave their final answers to 3 s.f. In the description of the *linear* correlation, the *strength* and *sign* must be mentioned.
(b) **(iii)** **[2]**	When $t = 20.4$, $\frac{1}{y} = -0.0005611311906(20.4) + 0.0273247901$ $y = 62.98$ Hence, the estimated daily sales is 63. Since $t = 20.4$ is outside the given temperature range $25.9 \le t \le 31.9$, extrapolation is used to estimate. Hence, the estimation is not reliable.	Students are reminded to use at least 5 s.f in their intermediate workings, otherwise the accuracy of the answer might be affected. It is important to state that it is the **value of *t*** that is outside the given data range for the estimation to be reliable.

9 **(a)** The masses of a randomly chosen bolt and a randomly chosen nut are denoted by M grams and W grams respectively. M and W are independent random variables with the distributions $N(272, 8^2)$ and $N(98, 5^2)$ respectively.

(i) Find the range of values of a for which $P(a < M < 275) > 0.2024$. [3]

(ii) Calculate the probability that twice the mass of a randomly chosen bolt differs from the total mass of 5 randomly chosen nuts by less than 80 grams. [3]

(b) Bolts are manufactured to fit into holes in steel plates. The bolts have diameters, in cm, that follow the distribution $N(2.65, 0.03^2)$ and the diameters of the holes, in cm, follow the distribution $N(2.72, 0.02^2)$. A manufacturer sells boxes of twenty pairs of these bolts and steel plates, where each pair consists of one randomly selected bolt and one randomly selected steel plate. A pair is acceptable only if the diameter of the hole in the steel plate is at least 0.02 cm larger, but no more than 0.15 cm larger, than the diameter of the bolt. Use a normal distribution to estimate the probability that the average number of acceptable pairs in 50 boxes is more than 18. [4]

(a)(i) **[3]**	$M \sim N(272, 8^2)$ $P(a < M < 275) > 0.2024$ $P(M < 275) - P(M < a) > 0.2024$ $P(M < a) < P(M < 275) - 0.2024$ $= 0.44377$ From GC, $P(M < 270.8687) = 0.44377$ Thus, $0 < a < 271$ (3sf)	You are reminded of the need for clear presentation in your answers. As a is not an integer, students should not be using GC function : [2nd][table] in solving the question. Many forget that non-exact numerical answer ought to be corrected to 3 sf. Many wrongly wrote $P(a < M < 275) = P(M < 275) - P(a < M)$.
(a)(ii) **[3]**	Let $R = 2M - (W_1 + W_2 + W_3 + W_4 + W_5)$. $E(R) = 2E(M) - 5E(W) = 2(272) - 5(98) = 54$ $Var(R) = 4Var(M) + 5E(W) = 4(64) + 5(25) = 381$ Hence, $R \sim N(54, 381)$ Required probability $= P(\lvert R \rvert < 80)$ $= P(-80 < R < 80)$ $= 0.909$ (3 sf)	It is advisable to define a new r.v. in terms of M and W, but **(i)** do not use Z, **(ii)** do not include modulus expression at this stage. $P(\lvert R \rvert < 80)$ and $P(R < 80)$ may give similar numerical answers (as the difference in this case is very small), but for correctness, only the former version is acceptable. Some students may have expanded $P(\lvert R \rvert < 80)$ wrongly in calculation, but for reason mentioned in previous para., the final answer is not affected. This too will be penalized. E.g. $P(\lvert R \rvert < 80) \neq P(R < 80) + P(R < -80)$ $P(\lvert R \rvert < 80) \neq P(R < 80) + P(R > -80)$

(b) **[4]**	Let the diameter of a bolt be X cm, and the diameter of a hole be Y cm. $X \sim \text{N}(2.65, 0.03^2)$; $Y \sim \text{N}(2.72, 0.02^2)$ $Y - X \sim \text{N}(0.07, 0.0013)$ $\text{P}(0.02 < Y - X < 0.15) = 0.903990211$ Let V be the number of acceptable pairs in a box of 20. $V \sim \text{B}(20,\ 0.903990211)$ $\text{E}(V) = 18.0798184$ $\quad$ $\text{Var}(V) = 1.73584$ **Method 1** Since sample size 50 is large, by Central Limit Theorem, $\overline{V} \sim \text{N}\left(18.0798184,\ \frac{1.73584}{50}\right)$ approximately i.e. $\overline{V} \sim \text{N}\left(18.0798184,\ 0.0347165\right)$ approximately Required probability = $\text{P}\left(\overline{V} > 18\right) = 0.666$ (3sf) **Method 2** Let $K = V_1 + V_2 + ... + V_{50}$. Since sample size 50 is large, by Central Limit Theorem, $K \sim \text{N}\left(50 \times 18.0798184,\ 50(1.73584)\right)$ approximately i.e. $K \sim \text{N}\left(903.99092,\ 86.792\right)$ approximately Required probability = $\text{P}\left(K > 900\right) = 0.666$ (3sf)	Many students seem to be not too familiar with the application of CLT. Such questions are pretty common. Do learn it well, taking care of the presentation requirements. Gentle reminder: As CLT involves either $\sum V_i$ or $\frac{1}{n}(\sum V_i)$, it is therefore better to first define V properly, as demonstrated in the suggested solution. There were instances where the students tried defining $W \sim \text{B}(1000,\ 0.90399)$, and then solve for $P(W > 900)$ directly. This is outright <u>ignoring</u> the instruction "use a normal distribution to estimate…. ". Some students try use a normal distribution to estimate $W \sim \text{B}(1000,\ 0.90399)$. However, this is only possible under stringent conditions (which need to be explicitly discussed in the presentation), which is out of syllabus.

10 **(a)** A company has a machine designed to fill bags with, on average, μ_0 kg of salt. The mass of salt in a randomly chosen bag has a normal distribution with population standard deviation denoted by σ kg. The production manager wishes to investigate if the machine is adjusted correctly. He takes a sample of n bags and carries out a hypothesis test at the 1% level of significance.

(i) State null and alternative hypotheses for the manager's test, defining any parameters you use. [2]

(ii) Find in terms of μ_0, σ and n, the critical region(s) for this test. [4]

(b) The company has a different machine which fills bags with, on average, 25 kg of low sodium salt. One of the company's production supervisor has reported that some of the workers suspect the machine is no longer set correctly, and the average mass of low sodium salt in the bags may in fact be more than 25 kg. The production supervisor decides to carry out a hypothesis test at the 0.5% level of significance with a random sample of 80 bags of low sodium salt. Summary data for the mass, y kg, of low sodium salt in these bags is as follows.

$$n = 80 \qquad \sum(y-25) = 27.2 \qquad \sum(y-25)^2 = 85.1$$

(i) Carry out the test and state the conclusion of the test in the context of the question. [5]

(ii) State, giving a reason, whether any assumptions about the population are needed in order for the test to be valid. [1]

(a)(i) **[2]**	Let μ kg be the population mean mass of salt in a bag. Null hypothesis, $H_0 : \mu = \mu_0$ Alternative hypothesis, $H_1 : \mu \neq \mu_0$	Most students can state the hypotheses correctly. However, a high number of students missed out the word "population" when defining μ, which students are strongly reminded not to forget.
(ii) **[4]**	Let X kg be the mass of salt in a bag. $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$ Perform a 2-tail test at 1% significance level. Under H_0, $\bar{X} \sim N\left(\mu_0, \dfrac{\sigma^2}{n}\right)$. $p\text{-value} = \begin{cases} 2P(\bar{X} \le \bar{x}) & \text{if } \bar{x} \le \mu_0 \\ 2P(\bar{X} \ge \bar{x}) & \text{if } \bar{x} \ge \mu_0 \end{cases}$ To reject H_0 at the 1% significance level, p-value $\le$ 0.01	Quite a few students struggled with this part with the following common mistakes: 1. Notations issues like $P(\bar{X} \le \mu)$, $P(X \le \mu_0)$, $P(X \neq \mu)$, etc. 2. Missing out "2" for the p-value. 3. Use CLT when it is not required. 4. Gave values as the final answer instead of regions. 5. Standardize wrongly as $\dfrac{\bar{x}-\mu_0}{\sigma}$, $\dfrac{\bar{x}-\mu_0}{\sigma^2/n}$ 6. Use $\bar{X} \sim N\left(\mu_0, \dfrac{n}{n-1}\sigma^2\right)$ 7. Write σ as "6" and so 2.5758σ became 2.58. Students are reminded that in order to end up with regions, they need to use inequality in their working or indicate clearly on the diagram if they are using the critical values.

$\Rightarrow 2\text{P}\left(\bar{X} \le \bar{x}\right) \le 0.01$ or $2\text{P}\left(\bar{X} \ge \bar{x}\right) \le 0.01$

$$\Rightarrow \text{P}\left(Z \le \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right) \le 0.005 \text{ or } \text{P}\left(Z \ge \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right) \le 0.005$$

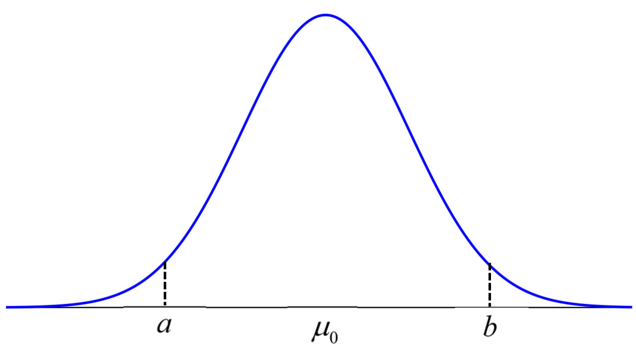
$$\Rightarrow \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \le -2.57583 \text{ or } \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \ge 2.57583$$

$$\Rightarrow \bar{x} \le \mu_0 - \frac{2.576\sigma}{\sqrt{n}} \text{ or } \bar{x} \ge \mu_0 + \frac{2.576\sigma}{\sqrt{n}}$$

The critical regions for this test such that H_0 is rejected in favour of H_1 at the 1% level of significance is $\left(0, \mu_0 - \frac{2.58\sigma}{\sqrt{n}}\right] \cup \left[\mu_0 + \frac{2.58\sigma}{\sqrt{n}}, \infty\right)$

(3 s.f. for the coeff).

Alternatively, let the critical regions be $\left(0, a\right] \cup \left[b, \infty\right)$, where a and b are as shown below.



Then we have

$2\text{P}\left(\bar{X} \le a\right) = 0.01$ or $2\text{P}\left(\bar{X} \ge b\right) = 0.01$

$$\Rightarrow \text{P}\left(Z \le \frac{a-\mu_0}{\sigma/\sqrt{n}}\right) = 0.005 \text{ or } \text{P}\left(Z \ge \frac{b-\mu_0}{\sigma/\sqrt{n}}\right) = 0.005$$

$$\Rightarrow \frac{a-\mu_0}{\sigma/\sqrt{n}} = -2.57583 \text{ or } \frac{b-\mu_0}{\sigma/\sqrt{n}} = 2.57583$$

$$\Rightarrow a = \mu_0 - \frac{2.576\sigma}{\sqrt{n}} \text{ or } b = \mu_0 + \frac{2.576\sigma}{\sqrt{n}}$$

Hence, the critical regions for this test is

$$\left(0, \mu_0 - \frac{2.58\sigma}{\sqrt{n}}\right] \cup \left[\mu_0 + \frac{2.58\sigma}{\sqrt{n}}, \infty\right)$$

(3 s.f. for the coeff).

(b)(i) **[5]**	Let Y kg be the mass of low sodium salt in a bag. $\bar{y}=\frac{27.2}{80}+25=25.34$ $s^2=\frac{1}{79}\left(85.1-\frac{27.2^2}{80}\right)=\frac{75.852}{79}\approx 0.960152$ $H_0: \mu=25$ vs $H_1: \mu>25$ Perform a 1-tail test at 0.5% significance level. Under H_0, $\bar{Y}\sim N\left(25, \frac{s^2}{n}\right)$ approx. by Central Limit Theorem since $n=80$ is large. From the sample, $\bar{y}=25.34$. Using a z-test, p-value $=P\left(\bar{Y}\geq 25.34\right)\approx 0.000956$ Since p-value = 0.000956 (3 sf) < 0.005, we reject H_0 and conclude there is sufficient evidence, at the 0.5% level of significance, that the mean mass of low sodium salt in the bags is more than 25 kg.	This part is pretty well done with many students demonstrated that they had put in effort to remember the 5 steps in carrying out a hypothesis test. However, a handful of students still committed some of these common mistakes: 1. $s^2=\frac{80}{80-1}\left(\frac{85.1}{80}\right)$ 2. $s^2=\frac{1}{80-1}\left(27.2^2-\frac{85.1^2}{80}\right)$ 3. $s^2=\frac{1}{80-1}\left(85.1-\frac{25.34^2}{80}\right)$ 4. $0.000956>0.0005$ 5. $\bar{Y}\sim N(25, 0.960152)$ 6. $\bar{Y}\sim N\left(25.34, \frac{0.960152}{80}\right)$
(ii) **[1]**	No assumption about the population is needed. Since sample size 80 is large, by Central Limit Theorem, $\bar{Y}$ follows a normal distribution approximately.	Not well done as many students committed the following common mistakes: 1. Used assumptions meant for binomial distribution, like mass of salt in the bags are independent of each other, constant probability of success. 2. Tried to mention sample being random or being a good representative of the population. 3. Missing out the words "sample" or "sample mean". 4. Mentioned that population is large to apply CLT. 5. Assume the sample size = 80 is large when it is in fact large.

11 Two brothers Kai and Leo are duathlon (running and cycling) athletes who train regularly. For each day, the probability that Kai cycles is $\frac{3}{4}$, the probability that he runs is $\frac{3}{5}$ and the probability he does both is p.

(a) Write down, in terms of p, the probability that, on one day, Kai either runs or cycles but not both. [1]

(b) Find the range of possible values of p. [2]

On average, Leo cycles 5 out of 7 days in a week. The probability that Leo cycles when Kai cycles is 0.9.

(c) Find the probability that Leo cycles when Kai does not. [3]

(d) State, in context, two assumptions needed for the number of days Leo cycles over a period of 5 weeks to be well modelled by a binomial distribution. [2]

Assume now that the number of days Leo cycles in 5 weeks has a binomial distribution.

(e) Find the probability that, in the 5 weeks, Leo cycles on at least 20 days but fewer than 30 days. [2]

(f) Find the probability that, in the 5 weeks, there are 5 days in which both brothers do not cycle and they both cycle on the other days. [2]

(a) **[1]**	$\text{P}(\text{Kai cycles only})+\text{P}(\text{Kai runs only})=\left(\frac{3}{4}-p\right)+\left(\frac{3}{5}-p\right)$ $=\frac{27}{20}-2p$	This is generally well done.
(b) **[2]**	$\text{P}(\text{Kai cycles or runs})=\frac{3}{4}+\frac{3}{5}-p$ When p is minimum, $\frac{3}{4}+\frac{3}{5}-p=1 \quad \Rightarrow \quad p=\frac{7}{20}$ Hence $\frac{7}{20}\le p\le \min\left\{\frac{3}{5},\frac{3}{4}\right\}$, i.e. $\frac{7}{20}\le p\le\frac{3}{5}$	Many students found the minimum value of p and wrote down 1 as the upper bound. This is clearly wrong because p must be smaller than both $\frac{3}{5}$ and $\frac{3}{4}$.
(c) **[3]**	Let K be the event that Kai cycles, and L the event that Leo cycles. It's given $\text{P}(L\mid K)=\frac{9}{10}$, and here we want to find $\text{P}(L\mid K')$: $\text{P}(L\mid K')=\frac{\text{P}(L\cap K')}{\text{P}(K')}$ $=\frac{\text{P}(L)-\text{P}(L\cap K)}{\text{P}(K')}$ $=\frac{\text{P}(L)-\text{P}(K)\cdot\text{P}(L\mid K)}{\text{P}(K')}=\frac{\frac{5}{7}-\frac{3}{4}\cdot\frac{9}{10}}{\frac{1}{4}}=\frac{11}{70}$	A lot of students stopped at $\text{P}(L\cap K')$. Take note that the required probability is actually the conditional probability $\text{P}(L\mid K')$ instead.

	Or from tree diagram, $$\mathrm{P}(L)=\mathrm{P}(K)\cdot\mathrm{P}(L\mid K)+\mathrm{P}(K')\cdot\mathrm{P}(L\mid K')$$ $$\frac{5}{7}=\frac{3}{4}\cdot\frac{9}{10}+\frac{1}{4}\mathrm{P}(L\mid K')$$ $$\mathrm{P}(L\mid K')=\frac{\frac{5}{7}-\frac{3}{4}\cdot\frac{9}{10}}{\frac{1}{4}}=\frac{11}{70}$$	
(d) **[2]**	Assume that: 1. The probability that Leo cycles on a day is constant at $\frac{5}{7}$ for all days over the 5 weeks. 2. The event that Leo cycles on a day is independent of the event that he cycles on another day.	1. A handful of students lost mark for not specifying the probability that Leo cycles refers to probability he cycles **on a day**. 2. Some made the mistakes in stating that the **probability** is independent rather than the **event** is independent.
(e) **[2]**	Let X be the number of days in 5 weeks that Leo cycles. $$X\sim \mathrm{B}\left(35,\frac{5}{7}\right)$$ $$\mathrm{P}(20\le X<30)=\mathrm{P}(X\le 29)-\mathrm{P}(X\le 19)$$ $$=0.937\quad(3\text{sf})$$	This is generally well done.
(f) **[2]**	$$\mathrm{P}(K'\cap L')=\mathrm{P}(K')\cdot\mathrm{P}(L'\mid K')=\frac{1}{4}\left(1-\frac{11}{70}\right)=\frac{59}{280}$$ $$\mathrm{P}(K\cap L)=\mathrm{P}(K)\cdot\mathrm{P}(L\mid K)=\frac{3}{4}\cdot\frac{9}{10}=\frac{27}{40}$$ Required probability is $${}^{35}C_5\left(\frac{59}{280}\right)^5\left(\frac{27}{40}\right)^{30}=0.00102\quad(3\text{sf})$$	This part is poorly done as many students wrongly assume that the random variable is modelled by a binomial distribution.