TEMASEK JUNIOR COLLEGE, SINGAPORE

Preliminary Examination 2014
Higher 2

MATHEMATICS

Paper 2

9740/02

15 September 2014

Additional Materials: Answer Paper
List of Formulae (MF 15)

3 hours

READ THESE INSTRUCTIONS FIRST

Write your *Civics Group* and *Name* on all the work that you hand in.
Write in dark blue or black pen on both sides of the paper.
You may use a soft pencil for any diagrams or graphs.
Do not use staples, paper clips, highlighters, glue or correction fluid.

Answer **all** the questions.

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 a graphic calculator.
Unsupported answers from a graphic calculator are allowed unless a question specifically states otherwise.

Where unsupported answers from a graphic calculator are not allowed in a question, you are required to present the mathematical steps using mathematical notations and not calculator commands.
You are reminded of the need for clear presentation in your answers.

At the end of the examination, fasten all your work securely together.
The number of marks is given in brackets [ ] at the end of each question or part question.

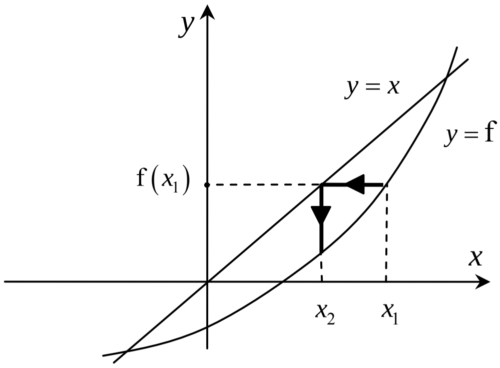
This document consists of **6** printed pages.

TEMASEK JUNIOR COLLEGE, SINGAPORE
PASSION PURPOSE DRIVE

Section A: Pure Mathematics [40 marks]

1 It is given that $\mathrm{f}(x) = \frac{1}{5}\left(3^x\right) - 1$. Find the roots α and β of the equation $\mathrm{f}(x) = x$, where $\alpha < \beta$. [2]

A sequence of real numbers $x_1, x_2, x_3, \ldots$ satisfies the recurrence relation given by $x_{n+1} = \mathrm{f}\left(x_n\right)$ for $n \geq 1$, and x_1 is a value that lies between α and β. The diagram shows how the graphs of $y = \mathrm{f}(x)$ and $y = x$ are used to obtain x_2 from x_1.



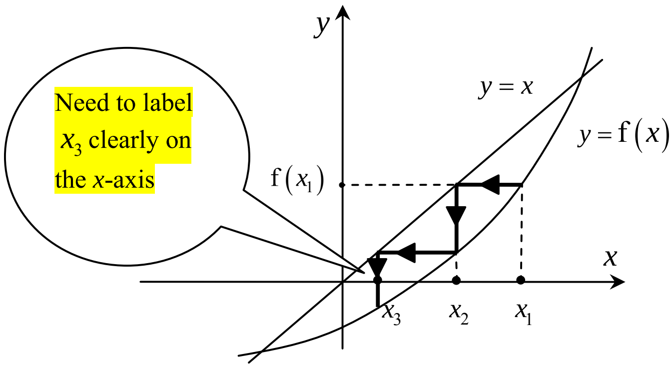
(i) Copy the above diagram, and illustrate how x_3 can be obtained. [1]

(ii) Describe the behaviour of the sequence in this case. [1]

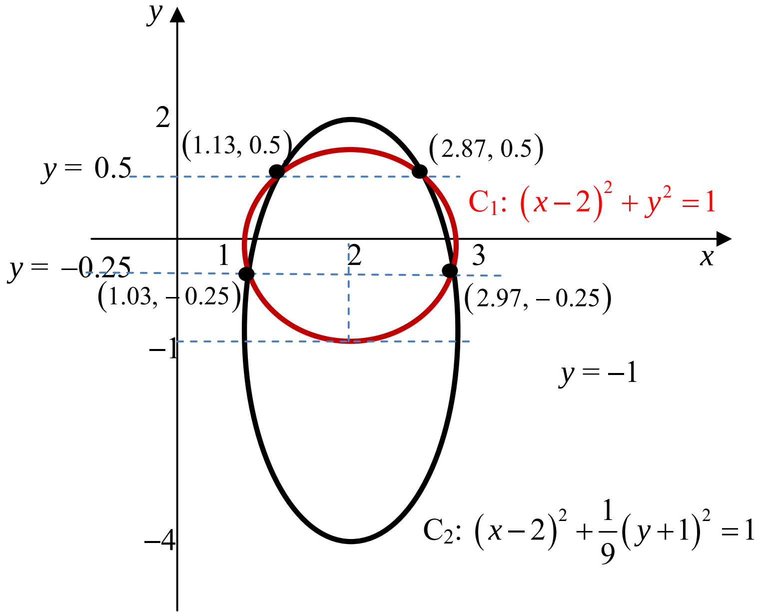
[Solution]

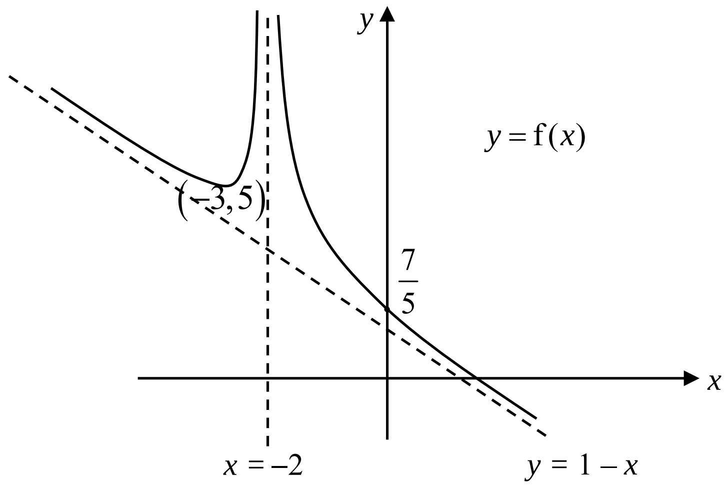
Using GC, $\alpha = -0.928$ and $\beta = 2.64$ (3sf)

(i)



(ii) The sequence decreases and converges to α (or -0.928).

2	A circle C_1 has equation given by $x^2 - 4x + y^2 + 3 = 0$. Another curve C_2 with equation $a(x+b)^2 + (y+c)^2 = a$ is obtained from circle C_1 after going through the following successive transformations: • Stretching with scale factor 3 parallel to the y-axis followed by • Translation of 1 unit in the negative y-axis direction. **(i)** Find the values of a, b and c. [3] **(ii)** Sketch, on the same diagram, the two curves, C_1 and C_2, indicating clearly the points of intersection of C_1 and C_2. [3] **(iii)** (x_1, k) and (x_2, k) are points on C_1 and C_2 respectively such that $2 < x_2 < x_1 \leq 3$, where k is a real constant. State the range of values of k. [1]
	[Solution] **(i)** $x^2 - 4x + y^2 + 3 = 0 \Leftrightarrow (x-2)^2 + y^2 = 1$ After 1st transformation: $(x-2)^2 + \left(\frac{y}{3}\right)^2 = 1$ After 2nd transformation: $(x-2)^2 + \frac{1}{9}(y+1)^2 = 1$ $\therefore 9(x-2)^2 + (y+1)^2 = 9$ $a = 9$, $b = -2$, $c = 1$
	(ii) 
	(iii) From diagram, range of k is $(-0.25, 0.5)$.

3	A curve has equation $y = \mathrm{f}(x)$, where $$\mathrm{f}(x) = \begin{cases} \dfrac{ax^2+bx+c}{x+2} & \text{for } x < -2, \\ \mathrm{h}(x) & \text{for } x > -2. \end{cases}$$ The curve has a turning point at $(-3, 5)$. The only asymptotes are $y = 1 - x$ and $x = -2$.  **(a)** Find the values of a, b and c. [3] **(b)** It is given that the tangent to the curve $y = \mathrm{f}(x)$ at the point $\left(0, \frac{7}{5}\right)$ is parallel to the line $2y + 3x = 0$. Sketch the graph of $y = \mathrm{f}'(x)$, indicating clearly all asymptotes and axial intercepts. [3] **(c)** The function g is such that $\int \mathrm{h}(x)\,\mathrm{d}x = \mathrm{g}(x) + k$ for $x > -2$, where k is an arbitrary constant. By sketching the graph of $y = \mathrm{h}(-\lvert x \rvert)$ for $-2 < x < 2$, express $$\int_{-\sqrt{2}}^{\sqrt{2}} \mathrm{h}(-\lvert x \rvert)\,\mathrm{d}x$$ in the form $p\,\mathrm{g}(u) + q\,\mathrm{g}(v)$, where p, q, u, v are constants to be determined. [3]
	[Solution] **(a)** Given $\mathrm{f}(x) = \dfrac{ax^2+bx+c}{x+2}$ for $x < 2$ Since oblique asymptote is $y = 1 - x$, let $\dfrac{ax^2+bx+c}{x+2} = 1 - x + \dfrac{d}{x+2}$

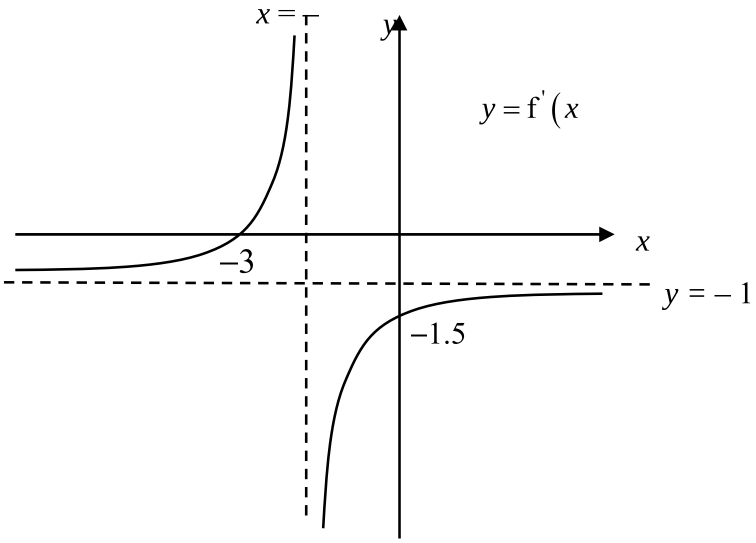
$$\frac{ax^2+bx+c}{x+2} = \frac{(1-x)(x+2)+d}{x+2}$$

Compare coefficients of x^2 : $a=-1$

Compare coefficient of x: $b=-1$

Since $\mathrm{f}(-3)=5 \Rightarrow \frac{9a-3b+c}{-3+2}=5 \Rightarrow c=1$

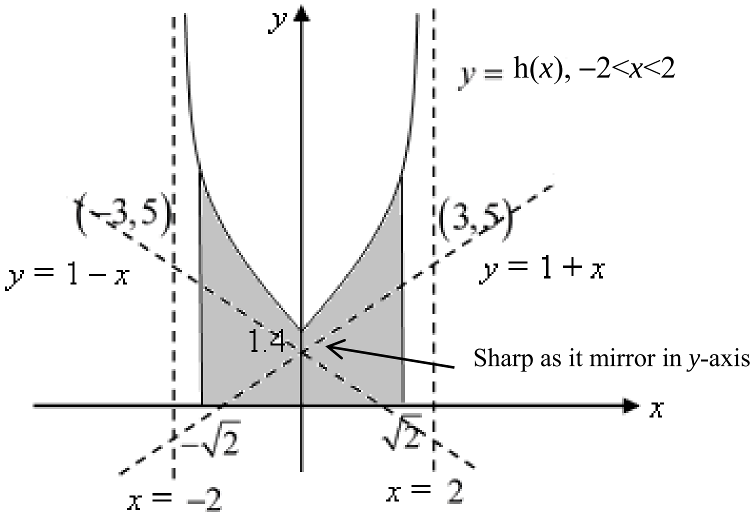
(b) Given that the tangent to the given curve at the point $\left(0,\frac{7}{5}\right)$ is parallel to the line $2y+3x=0$, this means that $\mathrm{f}'(0)=-1.5$.



(c) $y=\mathrm{h}(x) \quad \rightarrow \quad y=\mathrm{h}(-x) \quad \rightarrow \quad y=\mathrm{h}(-|x|)$

Replace x by $-x$
$\Rightarrow$ Reflect $y=\mathrm{h}(x)$
in y-axis

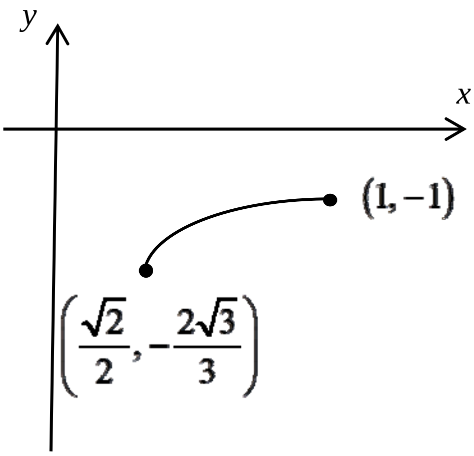
Replace x by $|x|$
$\Rightarrow$ Mirror $y=\mathrm{h}(x)$, $x \geq 0$
in the y-axis



$\int_{-\sqrt{2}}^{\sqrt{2}} \mathrm{h}\left(-|x|\right)\ \mathrm{d}x = 2\int_{-\sqrt{2}}^{0} \mathrm{h}\left(x\right)\ \mathrm{d}x$

[No labelling of asymptotes required for graph of $y = \mathrm{f}\left(-|x|\right)$]

$= 2\mathrm{g}\left(0\right) - 2\mathrm{g}\left(-\sqrt{2}\right)$

4	The parametric equations of a curve C are given by $$x=\sin\theta-\cos\theta \text{ and } y=\sec 2\theta \text{ where } \frac{5\pi}{12}\le\theta\le\frac{\pi}{2}.$$ **(i)** Sketch the curve C, indicating the coordinates of the endpoints clearly. [2] **(ii)** Show that the area of the region bounded by the curve C, x-axis and the lines $x=\frac{\sqrt{2}}{2}$ and $x=1$ can be written as $\int_{\frac{5\pi}{12}}^{\frac{\pi}{2}}\frac{1}{\sin\theta-\cos\theta}\,\mathrm{d}\theta$. [3] **(iii)** Prove that $\sin\theta-\cos\theta=\sqrt{2}\sin\left(\theta-\frac{\pi}{4}\right)$. [1] Hence show that the exact area of the region is $\frac{\sqrt{2}}{2}\ln\left(\sqrt{2}-1\right)\left(\sqrt{3}+2\right)$. [3]
	[Solution] **(i)** 
	(ii) Required area $=-\int_{\frac{\sqrt{2}}{2}}^{1} y\,\mathrm{d}x=-\int_{\frac{5\pi}{12}}^{\frac{\pi}{2}}\sec 2\theta\left(\cos\theta+\sin\theta\right)\mathrm{d}\theta$ $=-\int_{\frac{5\pi}{12}}^{\frac{\pi}{2}}\frac{\cos\theta+\sin\theta}{\cos^2\theta-\sin^2\theta}\,\mathrm{d}\theta$

	$= -\int_{\frac{5\pi}{12}}^{\frac{\pi}{2}} \frac{\cos\theta + \sin\theta}{(\cos\theta + \sin\theta)(\cos\theta - \sin\theta)}\, \mathrm{d}\theta$ $= \int_{\frac{5\pi}{12}}^{\frac{\pi}{2}} \frac{1}{\sin\theta - \cos\theta}\, \mathrm{d}\theta$ [Shown]
	(iii) $\text{RHS} = \sqrt{2}\sin\left(\theta - \frac{\pi}{4}\right)$ $= \sqrt{2}\left(\sin\theta\cos\frac{\pi}{4} - \cos\theta\sin\frac{\pi}{4}\right)$ $= \sin\theta - \cos\theta$ $= \text{LHS}$
	(iv) Hence required area $= \int_{\frac{5\pi}{12}}^{\frac{\pi}{2}} \frac{1}{\sin\theta - \cos\theta}\, \mathrm{d}\theta$ $= \int_{\frac{5\pi}{12}}^{\frac{\pi}{2}} \frac{1}{\sqrt{2}\sin\left(\theta - \frac{\pi}{4}\right)}\, \mathrm{d}\theta$ $= \frac{\sqrt{2}}{2}\int_{\frac{5\pi}{12}}^{\frac{\pi}{2}} \operatorname{cosec}\left(\theta - \frac{\pi}{4}\right) \mathrm{d}\theta$ $= \frac{\sqrt{2}}{2}\left[-\ln\left(\operatorname{cosec}\left(\theta - \frac{\pi}{4}\right) + \cot\left(\theta - \frac{\pi}{4}\right)\right)\right]_{\frac{5\pi}{12}}^{\frac{\pi}{2}}$ $= \frac{\sqrt{2}}{2}\left[-\ln\left(\operatorname{cosec}\left(\frac{\pi}{4}\right) + \cot\left(\frac{\pi}{4}\right)\right) + \ln\left(\operatorname{cosec}\left(\frac{\pi}{6}\right) + \cot\left(\frac{\pi}{6}\right)\right)\right]$ $= \frac{\sqrt{2}}{2}\left[-\ln\left(\sqrt{2}+1\right) + \ln\left(2+\sqrt{3}\right)\right] = \frac{\sqrt{2}}{2}\ln\left[\frac{\left(2+\sqrt{3}\right)\left(\sqrt{2}-1\right)}{\left(\sqrt{2}+1\right)\left(\sqrt{2}-1\right)}\right]$ $= \frac{\sqrt{2}}{2}\ln\left(\sqrt{2}-1\right)\left(2+\sqrt{3}\right)$

5	**(i)** Show that $1+\mathrm{e}^{\mathrm{i}\theta}=\left(2\cos\frac{\theta}{2}\right)\mathrm{e}^{\mathrm{i}\frac{\theta}{2}}$. [2] **(ii)** Solve the equation $(1-z)^6-z^6=0$, giving the complex number z in the form of $re^{i\theta}$, where $r>0$ and $-\pi<\theta\le\pi$. [4] **(iii)** Two of the roots are such that $\mathrm{Im}(z)<0$. One of these roots is $z_1=\frac{\sqrt{3}}{3}\mathrm{e}^{-\mathrm{i}\frac{\pi}{6}}$. State the other root z_2. [1] **(iv)** Given that ω is the complex number such that $\arg(\omega-z_1)=-\frac{5\pi}{6}$, find the minimum value of $\lvert\omega-z_2\rvert$. [4]
	[Solution] **(i)** $\text{RHS}=\left(2\cos\frac{\theta}{2}\right)\mathrm{e}^{\mathrm{i}\frac{\theta}{2}}$ $=\left(2\cos\frac{\theta}{2}\right)\left(\cos\frac{\theta}{2}+\mathrm{i}\sin\frac{\theta}{2}\right)$ $=2\cos^2\frac{\theta}{2}+\mathrm{i}\,2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$ $=1+\cos\theta+\mathrm{i}\sin\theta$ (Using Double Angle formulae) $=1+\mathrm{e}^{\mathrm{i}\theta}$ = LHS
	(ii) $(1-z)^6-z^6=0$ $\Rightarrow \left(\frac{1-z}{z}\right)^6=1$ $\Rightarrow \left(\frac{1}{z}-1\right)^6=1$ $\Rightarrow \frac{1}{z}-1=\mathrm{e}^{i\frac{2k\pi}{6}},\quad k=0,\pm1,\pm2,3$ $\Rightarrow \frac{1}{z}=1+\mathrm{e}^{i\frac{k\pi}{3}}=\left(2\cos\frac{k\pi}{6}\right)\mathrm{e}^{\mathrm{i}\frac{k\pi}{6}}$ (from (i)) $\Rightarrow z=\left(\frac{1}{2\cos\frac{k\pi}{6}}\right)\mathrm{e}^{-\mathrm{i}\frac{k\pi}{6}},\quad k=0,\pm1,\pm2$ [Note: k = 3 is removed because $\frac{1}{z}=0$ which is not possible.]

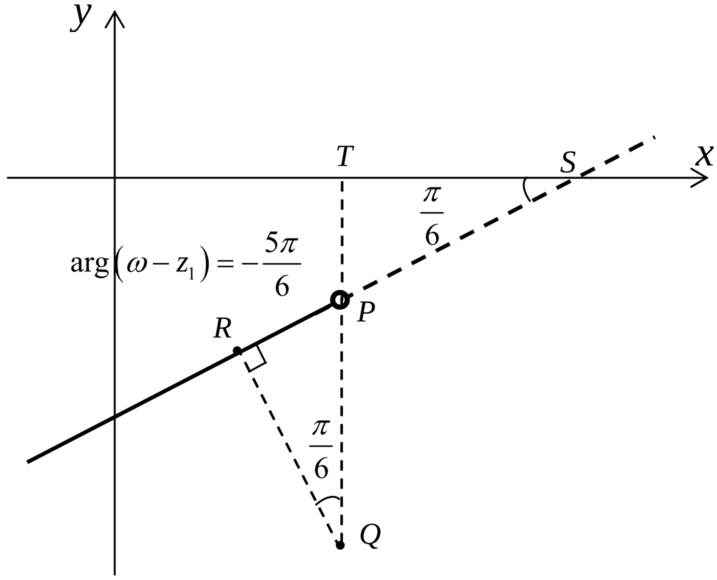
So the roots are $\frac{1}{2}\mathrm{e}^{\mathrm{i}\,0}$, $\left(\frac{\sqrt{3}}{3}\right)\mathrm{e}^{\pm\mathrm{i}\frac{\pi}{6}}$, $\mathrm{e}^{\pm\mathrm{i}\frac{\pi}{3}}$.

(iii) For $\operatorname{Im}(z)<0$, $-\pi<\arg(z)<0$. So the two roots satisfying this are

$z=\frac{1}{2\left(\frac{\sqrt{3}}{2}\right)}\mathrm{e}^{-\mathrm{i}\frac{\pi}{6}}$ or $z=\frac{1}{2\left(\frac{1}{2}\right)}\mathrm{e}^{-\mathrm{i}\frac{\pi}{3}}$

So $z_1=\frac{\sqrt{3}}{3}\mathrm{e}^{-\mathrm{i}\frac{\pi}{6}}$ and $z_2=\mathrm{e}^{-\mathrm{i}\frac{\pi}{3}}$

(iv) Let P and Q denote the points on the Argand diagram represented by the complex numbers z_1 and z_2 respectively.
Let the foot of perpendicular from Q to the half line be R.



From diagram, the minimum value of $|\omega-z_2|$ is QR.

Method 1:
Now $\triangle PQR$ is similar to $\triangle PST$ (AA property)

So $\angle PQR=\angle PST=\frac{\pi}{6}$

Now $z_1=\frac{1}{2}-\frac{\sqrt{3}}{6}i$ and $z_2=\frac{1}{2}-\frac{\sqrt{3}}{2}i$

Also PQ = vertical distance between P and Q

$=|z_1-z_2|=\frac{\sqrt{3}}{3}$

So $QR=\frac{\sqrt{3}}{3}\sin\frac{\pi}{6}=\frac{1}{2}$.

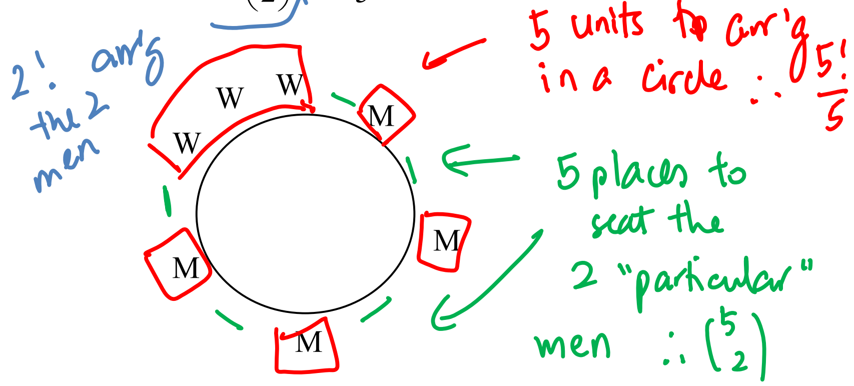
Method 2:

	Cartesian equation of half line : $y=\frac{\sqrt{3}}{3}x-\frac{\sqrt{3}}{3}$ Cartesian equation of line QR: $y=-\sqrt{3}x$ Solving simultaneously, R is $\left(\frac{1}{4},-\frac{\sqrt{3}}{4}\right)$ So $QR=\sqrt{\left(\frac{1}{2}-\frac{1}{4}\right)^2+\left(\frac{\sqrt{3}}{4}-\frac{\sqrt{3}}{2}\right)^2}=\frac{1}{2}$

Section B: Statistics [60 marks]

6	Explain what is meant by the term 'random sample'. [1] The Ministry of Education wishes to conduct a focus group discussion with teachers from primary schools, secondary schools and junior colleges to find out how they use ICT in the classroom. Give a reason why simple random sampling may not give a representative sample of teachers' feedback across different schools and levels. [1] Explain why a stratified sample is more appropriate in this context. [2]
	[Solution] Random sample refers to sample obtained from a sampling procedure whereby each member of the sample was selected from the population with an equal chance/probability.
	Simple random sampling does not guarantee representative sample of teachers' feedback from across schools and levels because there is always the possibility that a chosen random sample may exclude a certain level such as all JC teachers. A stratified sample is preferred because the use of ICT varies with different school levels (primary, secondary, JC) and by considering mutually exclusive subgroups (strata), it helps to ensure that the different feedbacks can be collected from teachers of these school levels. Furthermore, strata are selected based on their proportionality to the teacher population in the schools and levels. As such this ensures that all levels are proportionally represented.

7	**(a) (i)** 6 men and 3 women signed up for a dance class. Find the number of ways of distributing them into 3 equally-sized groups labelled *A*, *B* and *C*. [2] **(ii)** After the groups are finalised, the participants are asked to sit in a row of 11 chairs so that they can view a dance demonstration video. Find the number of possible arrangements so that members of each group are seated together. [3] **(b)** Find the number of ways that 6 men and 3 women can sit in a circle of 9 chairs so that the women are seated together and two particular men are not seated next to each other. [3]
	[Solution] **(a) (i)** Number of ways = $\binom{9}{3}\binom{6}{3}\binom{3}{3}=1680$
	(a) (ii) There will be 11 – 3 x 3 = 2 empty seats. Therefore, the number of ways of rearranging the three groups where each family of three sit together is $\frac{5!}{2!}3!3!3! = 12960$ **Alternative Method 1:** [Take the 11 chairs to be 5 units. 5 units select 3 units to be the number of ways to seat the 3 groups of people. This is then followed by arranging the 3 groups as well as the 3 people in each group] $\binom{5}{3}\times(3!)^3\times 3! = 12960$ **Alternative Method 2:** [Selecting the spaces between the 3 groups to put the empty chairs, can be 1 chair between the 2 groups or 2 chairs between 2 groups;] $\left[\binom{4}{2}+\binom{4}{1}\right]\times(3!)^3\times 3! =12960$

	(b) 3! arr'g the 3 women in the unit Number of ways $= \binom{5}{2} 2!3!\frac{5!}{5} = 2880$ 
	Alternative Method [Elimination Method, but need to make sure that the condition of 3 women are seated together is also considered in the without restriction case] $\frac{7!}{7} \times 3! - \frac{6!}{6} \times 3! \times 2! = 2880$ Besides the 3 women seated together, the 2 particular men are also seated together.

8	Anand, Beng and Charlie patronised a restaurant that offered lucky draws to its customers depending on what they ordered. Anand and Beng ordered set meals, and each got to participate in a lucky draw from the "Blue Box", which had a 25% chance of awarding a prize. Charlie order a la carte, and got to participate in a lucky draw from the "Gold Box", which had a 40% chance of awarding a prize. Find the probability that **(i)** Anand, Beng and Charlie all won prizes from their lucky draws. [2] **(ii)** at least one of Anand, Beng and Charlie won a prize from their lucky draws. [2] **(iii)** Charlie won a prize from his lucky draw, given that at least one of Anand, Beng and Charlie won a prize from their lucky draws. [4]
	[Solution] **(i)** P(all 3 won prizes) = 0.25 × 0.25 × 0.4 $= 0.025 = \frac{1}{40}$
	(ii) P(at least 1 won a prize) = 1 – P(no one won a prize) = 1 – 0.75 × 0.75 × 0.6 $= 0.6625 = \frac{53}{80}$
	(iii) P(Charlie won a prize | at least 1 won a prize) = $\frac{\text{P(only C won)+P(only C and A won) +P(only C and B won) +P(all 3 won)}}{\text{P(at least 1 won a prize)}}$ $= \frac{(0.4\times0.75\times0.75)+2(0.4\times0.75\times0.25)+(0.4\times0.25\times0.25)}{0.6625}$ = 0.604 (3sf) **Alternative Solution 1:** P(Charlie won a prize | at least 1 won a prize) $= \frac{\text{P}\left(\text{Charlie won a prize and at least 1 won a prize}\right)}{\text{P(at least 1 won a prize)}} = \frac{0.4}{0.6625} = \frac{32}{53}$ **Alternative Solution 2:** P(Charlie won a prize | at least 1 won a prize) = $= \frac{\text{P}\left(\text{Charlie won a prize and at least 1 won a prize}\right)}{\text{P(at least 1 won a prize)}} = \frac{\frac{53}{80} - P(ABC', A'BC', ABC')}{0.6625} = \frac{32}{53}$

9	Amongst various exhibits at an art gallery, only paintings and sculptures are placed for sale. The number of paintings sold from the art gallery in a week follows a Poisson distribution with mean 1. The number of sculptures sold from the same art gallery in a week follows a Poisson distribution with mean 0.25. You can assume that the sale of the sculptures and paintings are independent. Taking a month to be 4 weeks, **(i)** find the probability that one painting is sold in a month. [1] **(ii)** find the probability that over a period of 4 weeks, the first sculpture is sold in the fourth week. [2] **(iii)** If the total sales exceed 4 pieces in a month, the gallery owner will consider that month a "Good" month. Taking a year to be 12 months, use a suitable approximation to find the probability that there are no more than 30 "Good" months in 5 years. [6]
	[Solution] **(i)** Let A be the number of paintings sold in 1 month. $A \sim P_o(4)$ Req prob $= P(A=1) = 0.0733$
	(ii) Let B be the number of sculptures sold in 1 week. $B \sim P_o(0.25)$ Req prob $= \left[P(B=0)\right]^3 P(B \geq 1)$ $= 0.104$
	(iii) Let C be the number of art pieces sold in 1 month. $C \sim P_o(5)$ Probability of a "Good" month $= P(C > 4)$ $= 1 - P(C \leq 4)$ $= 0.5595067149$ Let X be the number of "Good" months in 5 years. $X \sim B(60, 0.5595067149)$ Since $n = 60$ is large, $np = 33.57040289 > 5$ & $n(1-p) = 26.42959711 > 5$ $X \sim N(33.57040289, 14.78753706)$ approx. Req prob $= P(X \leq 30) \overset{c.c}{=} P(X < 30.5)$ $= 0.21230$

10	Flour is packed by a manufacturer into bags labelled 5.00 kg each. Recently, the manufacturer

	received complaints from some customers who claimed that the flour they bought weighed less. To test the claim, the manufacturer takes a random sample of 60 bags and records down the weight, t kg, of the flour in each bag. The results are summarised by $\sum t = 291.6$ and $\sum(t-4.86)^2 = 18.344$ Test, at 5% significance level, if the complaints are justified. [5] The manufacturer decides to re-label the weight of flour in each bag as μ_0 kg. Find the range of values of μ_0 so that the data previously collected provides insufficient evidence at the 10% level of significance to conclude that the mean weight of flour in a bag is not μ_0 kg. Leave your answers correct to 2 decimal places. [4]
	[Solution] **(i)** $\bar{t} = \frac{291.6}{60} = 4.86$, $s^2 = \frac{1}{59}\sum(t-4.86)^2 = \frac{18.344}{59}$, $H_0 : \mu = 5$ $H_1 : \mu < 5$ Level of significance = 5% Under H_0, $Z = \frac{\bar{T}-\mu}{s/\sqrt{n}} \sim \mathrm{N}(0,1)$ by the Central Limit Theorem $p-$value $= 0.0259$ Conclusion: Since p-value = 0.0259 < 0.05, we reject H_0. There is sufficient evidence, at 5% level of significance, to say that the manufacturer has overstated the mean weight. Alternatively, a t-test can be used [ie p-value = 0.0283, reject H_0] and the assumption that T is normally distributed must be mentioned.
	(ii) $H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$ Level of significance = 5% Test statistic: $Z = \frac{\bar{T}-\mu}{s/\sqrt{n}} \sim \mathrm{N}(0,1)$ by the Central Limit Theorem

Under H_0, $z_{cal} = \dfrac{4.86 - \mu_0}{\sqrt{\frac{18.344}{59} \Big/ 60}}$

So in order for H_0 not to be rejected at 5% level of significance, $|z_{cal}| < 1.64485$

$$\Rightarrow \left| \frac{4.86 - \mu_0}{\sqrt{\frac{18.344}{59} \Big/ 60}} \right| < 1.64485$$

$$\Rightarrow -1.64485 < \frac{4.86 - \mu_0}{\sqrt{\frac{18.344}{59} \Big/ 60}} < -1.64485$$

$\Rightarrow 4.74128 < \mu_0 < 4.9787$

So the range of values of μ_0 is $4.75 \le \mu_0 \le 4.97$ (2 d.p.)

Alternatively, students can perform a t-test, ie

$$\Rightarrow \left| \frac{4.86 - \mu_0}{\sqrt{\frac{18.344}{59} \Big/ 60}} \right| < 1.67109$$

$\Rightarrow 4.7397 < \mu_0 < 4.9803$

So the range of values of μ_0 is $4.74 \le \mu_0 \le 4.98$ (2 d.p.)

11 The following data shows the median heights of girls (in cm) for ages from each year 2 to 18.

Age, x	2	4	6	8	10	12	14	16	18

years									
Height, y cm	86	99.8	115	128	140.4	153.9	161.4	162.5	163.4

Danny computed the linear product moment correlation coefficient to be 0.970, and concludes that a linear model should be used to represent the relationship between x and y.

(i) Find the least squares estimates of a and b for the model $y = a + bx$. [1]

(ii) Draw a scatter diagram for the data. [2]

Danny used the equation in **(i)** to estimate the median height of an 11 year-old girl.
Edward, however, decides to use only the above data from ages 2 to 12 years to estimate the median height of the 11 year old girl. He calculated the least squares regression line of y on x and found it to be $y = 73.1 + 6.78x$.

(iii) Explain why Danny's estimate is not reliable as compared to Edward's. [2]

(iv) Explain why Edward should not use the regression line of x on y to estimate the median height of a girl. [1]

(v) Using Edward's regression line of y on x, estimate the median height of a girl who is 14 years and 6 months old. [1]

Explain why this estimate is not reliable, and suggest a model to obtain a reliable estimate. [2]

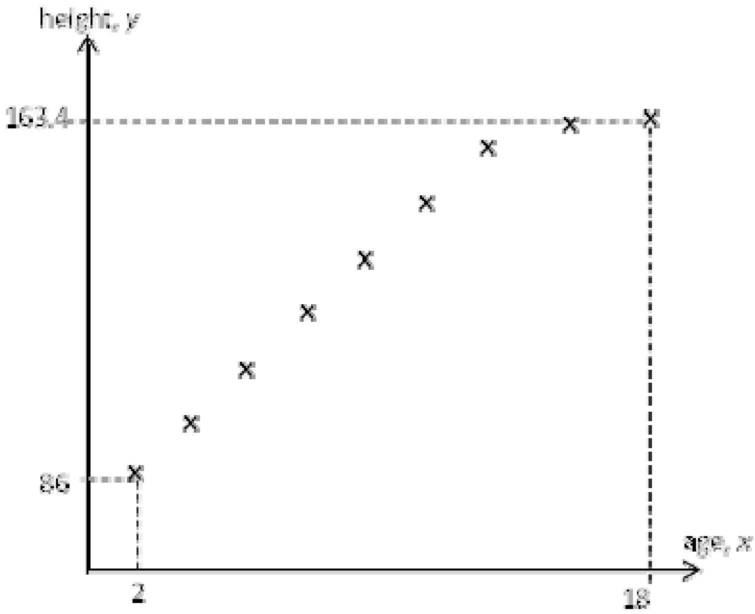
(vi) Interpret the regression coefficient of Edward's regression line in the context of the question. [1]

[Solution]

(i) From GC, the least squares regression line of y on x is $y = 83.122 + 5.1367x$

$\Rightarrow \quad y = 83.1 + 5.14x$ (to 3 sig figs)

(ii)



(iii) Edward's estimate is reliable because the six data points for 2 to 12 years old appear to lie close to a straight line suggesting that *x* and *y* have a **linear relationship for <u>this</u> age range**. Furthermore, the estimated value lies within the range of the given data.

On the other hand, the data points used by Danny (**from ages 2 to 12 years**) appear to have a **curvilinear relationship** and so the regression line is **not a good fit**, resulting in a less reliable estimate.

(iv) Age, *x*, is not dependent on height, *y*. In fact, height, *y*, is dependent on age, *x*, and so only the regression line of *y* on *x* should be used instead.

(v) $y = 73.1 + 6.78x$

When $x = 14.5$,

$y = 73.1 + 6.78(14.5) = 171.41$

Thus, the median height of a 14 year 6 months old girl is estimated to be 171.4 cm.

This estimate is not reliable as $x = 14$ is outside Edward's sample data range of $2 \le x \le 12$ where the linear relationship no longer holds.

<u>Or</u>

171.4cm is not reliable as this is a value larger than even the median height of a 18 year old girl (163.4cm). This large value is obtained because the linear relationship no

		longer holds beyond age 12 years. A reliable estimate can be observed by modelling the given data (for ages 2 to 12 years) using a curve that is increasing <u>and</u> concave downwards.
	(vi)	The regression coefficient of 6.78 means that a girl aged between 2 and 12 years is predicted to grow by 6.78cm each year.

12	The height of a three-year old boy from Sunshine Childcare Centre is a random variable with

mean 105 cm and standard deviation 9 cm.

(i) Find the mean and variance of the total height of a random sample of 72 three-year old boys. [2]

Hence, find the probability that their total height is greater than 76 m. [2]

The heights of three-year old girls from Sunshine Childcare Centre are modelled as having a normal distribution with mean μ cm and standard deviation σ cm.

(ii) If the probability that a randomly chosen three-year old girl has a height of 95 cm or less is $\frac{1}{6}$, find an equation satisfied by μ and σ. [2]

For the rest of the question, take μ = 100 and σ = 5.17.

(iii) 10 three-year-old girls are randomly chosen from the centre. Find the probability that the height of the shortest girl in the group is more than 95 cm. [2]

(iv) Determine if $\bar{G} = \dfrac{G_1 + G_2 + G_3}{3}$ and G_1 are independent random variables, where G_1, G_2, G_3 denote the heights of 3 randomly chosen three-year-old girls from the centre. [1]

(v) Find $\mathrm{P}\left(\bar{G} \geq G_1 + 2\right)$. [3]

[Solution]

Let B denote the height of a randomly chosen three-year old boy from Sunshine Childcare Centre. $\mathrm{E}(B) = 105$, $\mathrm{Var}(B) = 9^2$

(i) Let $S = B_1 + B_1 + \cdots + B_{72}$. Then

$\mathrm{E}(S) = 72 \times 105 = 7560$, $\mathrm{Var}(S) = 72 \times 9^2 = 5832$

Since n is large, by Central Limit Theorem,

$S \sim \mathrm{N}(7560, 5832)$ approximately

$\mathrm{P}(S > 7600) = 0.300$

(ii) Let G denote the height of a randomly chosen three-year old boy from Sunshine

		Childcare Centre. $G \sim \mathrm{N}\left(\mu, \sigma^2\right)$ Given $\mathrm{P}\left(G \le 95\right) = \frac{1}{6}$ $\mathrm{P}\left(Z \le \frac{95-\mu}{\sigma}\right) = \frac{1}{6}$ Using GC, $\frac{95-\mu}{\sigma} = -0.967421568$ $\mu - 0.967\sigma = 95$ (Shown)
	(iii)	Required Probability = $\left[\mathrm{P}\left(R > 95\right)\right]^{10} = \left(0.833333341\right)^{10} = 0.161$
	(iv)	$\overline{G}$ and G_1 are not independent of one another as G_1 is a component of $\overline{G} = \frac{G_1 + G_2 + G_3}{3}$. [Note that you are not required to prove that $\overline{G}$ and G_1 are not independent mathematically as it's beyond your scope of A-level syllabus]
	(v)	Now $\mathrm{P}\left(\overline{G} \ge G_1 + 2\right) = \mathrm{P}\left(\frac{G_1 + G_2 + G_3}{3} \ge G_1 + 2\right)$ $= \mathrm{P}\left(G_1 + G_2 + G_3 \ge 3G_1 + 6\right)$ $= \mathrm{P}\left(G_2 + G_3 - 2G_1 \ge 6\right)$ $\mathrm{E}\left(G_2 + G_3 - 2G_1\right) = 0$ and $\mathrm{Var}\left(G_2 + G_3 - 2G_1\right) = 2\left(5.17^2\right) + 4\left(5.17^2\right) = 6\left(5.17^2\right)$ So $G_2 + G_3 - 2G \sim N\left(0, 6\left(5.17^2\right)\right)$ So $\mathrm{P}\left(\overline{G} \ge G_1 + 2\right) = \mathrm{P}\left(G_2 + G_3 - 2G_1 \ge 6\right) = 0.318$