2015 IJC H2 Maths Prelim 2 Paper 2 (Suggested Solution)

1 The parametric equations of a curve are

$$x = 1 + \mathrm{e}^{t}, \quad y = \mathrm{e}^{2t}.$$

(i) Find the equation of the normal to the curve at the point $P\left(1+\mathrm{e}^{p}\ ,\ \mathrm{e}^{2p}\right)$. [3]

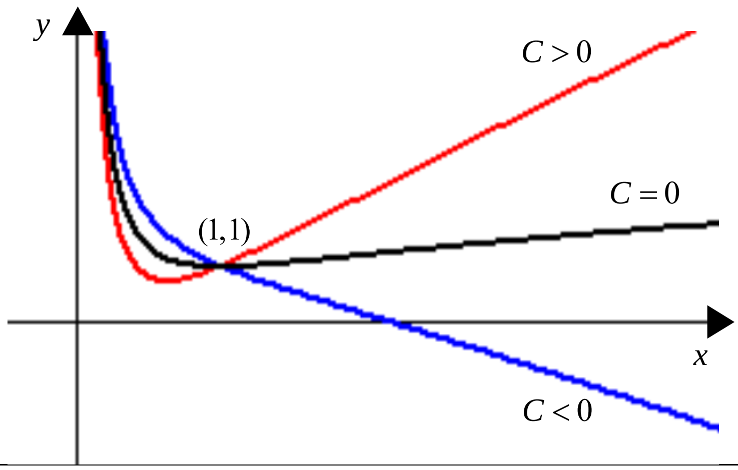
(ii) This normal meets the x-axis at the point Q. Find the cartesian equation of the locus of the mid-point of PQ as p varies. [4]

<table>
<tr><th>Q1</th><th>Solutions</th></tr>
<tr><td>(i)</td><td>
$\dfrac{\mathrm{d}x}{\mathrm{d}t} = \mathrm{e}^{t}$ and $\dfrac{\mathrm{d}y}{\mathrm{d}t} = 2\mathrm{e}^{2t}$
$\therefore \dfrac{\mathrm{d}y}{\mathrm{d}x} = 2\mathrm{e}^{t}$
OR
$x = 1 + \mathrm{e}^{t} \Rightarrow \mathrm{e}^{t} = x - 1$ ...... (1)
$y = \mathrm{e}^{2t}$ ....(2)
Sub (1) to (2),
$y = (x-1)^{2}$
$\therefore \dfrac{\mathrm{d}y}{\mathrm{d}x} = 2(x-1) = 2\mathrm{e}^{t}$
At point P, $t = p$.
Hence of gradient normal at $P = -\dfrac{1}{2\mathrm{e}^{p}}$
Equation of the normal:
$y - \mathrm{e}^{2p} = -\dfrac{1}{2\mathrm{e}^{p}}\left(x - (1+\mathrm{e}^{p})\right)$
$y = -\dfrac{1}{2\mathrm{e}^{p}}x + \dfrac{1}{2}\mathrm{e}^{-p} + \dfrac{1}{2} + \mathrm{e}^{2p}$
OR
Let the equation of normal be $y = mx + C$
</td></tr>
</table>

	$\therefore \quad e^{2p} = -\frac{1}{2e^{p}}(1+e^{p}) + C$ $C = e^{2p} + \frac{1}{2e^{p}} + \frac{1}{2}$ $\therefore$ Equation of the normal $y = -\frac{1}{2e^{p}}x + \frac{1}{2}e^{-p} + \frac{1}{2} + e^{2p}$
(ii)	At point *Q*, where $y=0$, $x = 1 + e^{p} + 2e^{3p}$. Mid-point of *PQ*: $\left(1+e^{p}+e^{3p}, \frac{1}{2}e^{2p}\right)$. Let $x = 1+e^{p}+e^{3p}$ and $y = \frac{1}{2}e^{2p}$. Eliminating *p*: Since $e^{p} = (2y)^{\frac{1}{2}}$, $x = 1 + (2y)^{\frac{1}{2}} + (2y)^{\frac{3}{2}}$. $\Rightarrow \; x - 1 = (2y)^{\frac{1}{2}}(1+2y)$ $\Rightarrow \; (x-1)^{2} = (2y)(1+2y)^{2}$

2 **(i)** Find the general solution of the differential equation $x^3\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2}=2-x$. [3]

(ii) It is given that $y=1$ when $x=1$. On a single diagram, sketch three members of the family of solution curves for $x>0$. [5]

Q2	**Solutions**
(i)	$\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2}=\dfrac{2}{x^3}-\dfrac{1}{x^2}$ $\dfrac{\mathrm{d}y}{\mathrm{d}x}=-\dfrac{1}{x^2}+\dfrac{1}{x}+C$ $y=\dfrac{1}{x}+\ln\lvert x\rvert+Cx+D$
(ii)	When $x=1$ and $y=1$, $1=1+0+C+D \Rightarrow D=-C$ Hence, $y=\dfrac{1}{x}+\ln\lvert x\rvert+C\left(x-1\right)$ 

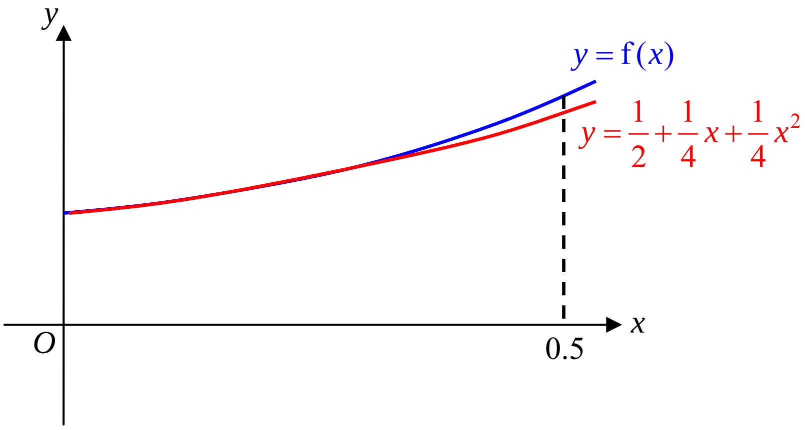
3 **(i)** Given that $f(x)=\dfrac{\sin x+2}{\cos 2x+3}$, where x is sufficiently small, find the series expansion of $f(x)$ in ascending powers of x, up to and including the term in x^2.

[4]

(ii) Use your answer to part **(i)** to give an approximation for $\int_0^n f(x)\,dx$ in terms of n. Evaluate this approximation in the case where $n=0.5$, leaving your answer in 6 decimal places. [3]

(iii) Use your calculator to find an accurate value for $\int_0^{0.5} f(x)\,dx$, correct to 6 decimal places. Explain why this value is more than the approximation obtained in part **(ii)**. [3]

Q3	**Solutions**
(i)	Since x is small, $\sin x \approx x$ and $\cos 2x \approx 1-\dfrac{(2x)^2}{2}=1-2x^2$. $f(x)=\dfrac{\sin x+2}{\cos 2x+3}\approx\dfrac{x+2}{(1-2x^2+3)}=\dfrac{x+2}{4-2x^2}$ $f(x)=(x+2)(4-2x^2)^{-1}$ $=\dfrac{1}{4}(x+2)\left(1-\dfrac{x^2}{2}\right)^{-1}$ $=\dfrac{1}{4}(x+2)\left(1+\dfrac{x^2}{2}+...\right)$ $\approx\dfrac{1}{2}+\dfrac{1}{4}x+\dfrac{1}{4}x^2$
(ii)	$\int_0^n f(x)\,dx \approx \int_0^n \dfrac{1}{2}+\dfrac{1}{4}x+\dfrac{1}{4}x^2\,dx$ $=\left[\dfrac{1}{2}x+\dfrac{1}{8}x^2+\dfrac{1}{12}x^3\right]_0^n$ $=\dfrac{1}{2}n+\dfrac{1}{8}n^2+\dfrac{1}{12}n^3$ When $n=0.5$,

	$\int_0^{0.5} \mathrm{f}(x)\ \mathrm{d}x = \frac{1}{2}(0.5)+\frac{1}{8}(0.5)^2+\frac{1}{12}(0.5)^3$ $= 0.29166667$ ≈ 0.291667 (6 dec pl)
(iii)	For an accurate value, $\int_0^{0.5} \mathrm{f}(x)\ \mathrm{d}x = 0.2932183197$ ≈ 0.293218 (6 dec pl) **<u>Method 1:</u>**  From the above sketch, it can be observed that the graph of $y=\mathrm{f}(x)$ is higher than the graph of $y=\frac{1}{2}+\frac{1}{4}x+\frac{1}{4}x^2$ for $0 \le x \le 0.5$. Hence for $0 \le x \le 0.5$, the area under the curve $y=\mathrm{f}(x)$, given by 0.293218, is more than the area under the curve for $y=\frac{1}{2}+\frac{1}{4}x+\frac{1}{4}x^2$, which is given by the value of 0.291667 obtained in part **(ii)**. **<u>Method 2:</u>** The full expansion of f(x) includes terms with powers of x higher than 2, all with positive coefficients. The integration of the terms with limits 0 to 0.5 will result in the sum of positive values. Thus the answer in part (iii) is larger than the approximation obtained in part (ii).

4 **(a)** The complex number z satisfies the relations $|z-3|\le 3$ and $|z-3-3\mathrm{i}|=|z|$.

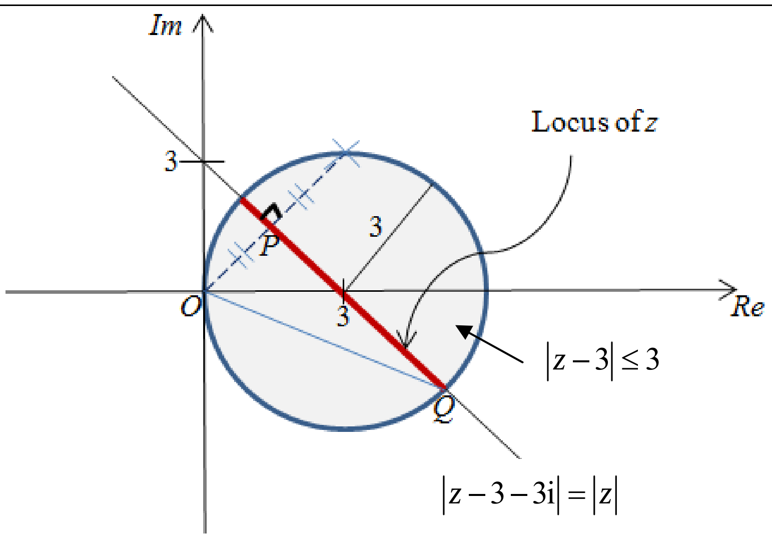
(i) Illustrate both of these relations on a single Argand diagram. [3]

(ii) Find exactly the maximum and minimum possible values of $|z|^2$. [4]

(b) The complex number w is given by $\left(\dfrac{-\sqrt{3}+\mathrm{i}}{\sqrt{2}-\mathrm{i}\sqrt{2}}\right)^2$. Without using a calculator, find

(i) $|w|$ and the exact value of $\arg w$, [4]

(ii) the set of values of n, where n is a positive integer, for which $w^n w^*$ is a real number. [4]

Q4	Solutions
(a) (i)	
(ii)	Min value of $|z|^2 = OP^2 = \left(3\cos\dfrac{\pi}{4}\right)^2 = \dfrac{9}{2}$ Max value of $|z|^2$ $= OQ^2$ $= 3^2+3^2-2(3)(3)\cos\dfrac{3\pi}{4}$ $= 18+9\sqrt{2}$

(b) (i)	**Method 1:** $w=\left(\dfrac{-\sqrt{3}+\mathrm{i}}{\sqrt{2}-\sqrt{2}\mathrm{i}}\right)^2=\left(\dfrac{2\mathrm{e}^{\frac{5\pi}{6}\mathrm{i}}}{2\mathrm{e}^{-\frac{\pi}{4}\mathrm{i}}}\right)^2=\mathrm{e}^{\frac{13\pi}{6}\mathrm{i}}=\mathrm{e}^{\frac{\pi}{6}\mathrm{i}}$ $\therefore\ \lvert w\rvert=1$ and $\arg w=\dfrac{\pi}{6}$ **Method 2:** $\lvert w\rvert=\left\lvert\dfrac{-\sqrt{3}+\mathrm{i}}{\sqrt{2}-\sqrt{2}\mathrm{i}}\right\rvert^2=\dfrac{\lvert-\sqrt{3}+\mathrm{i}\rvert^2}{\lvert\sqrt{2}-\sqrt{2}\mathrm{i}\rvert^2}=\dfrac{4}{4}=1$ $\arg w=\arg\left(\dfrac{-\sqrt{3}+\mathrm{i}}{\sqrt{2}-\sqrt{2}\mathrm{i}}\right)^2$ $=2\left[\arg\left(-\sqrt{3}+\mathrm{i}\right)-\arg\left(\sqrt{2}-\mathrm{i}\sqrt{2}\right)\right]$ $=2\left[\dfrac{5\pi}{6}-\left(-\dfrac{\pi}{4}\right)\right]-2\pi$ $=\dfrac{\pi}{6}$ **Method 3:** $w=\left(\dfrac{-\sqrt{3}+\mathrm{i}}{\sqrt{2}-\sqrt{2}\mathrm{i}}\right)^2$ $=\dfrac{3-2\sqrt{3}\mathrm{i}-1}{2-4\mathrm{i}-2}$ $=\dfrac{2-2\sqrt{3}\mathrm{i}}{4\mathrm{i}}$ $=\dfrac{\sqrt{3}}{2}+\dfrac{1}{2}\mathrm{i}$ $\therefore\ \lvert w\rvert=1$ and $\arg w=\dfrac{\pi}{6}$
(ii)	$\arg w^n w^*=n\arg w-\arg w=\left(n-1\right)\dfrac{\pi}{6}$ Since $w^n w^*$ is a real number, and n is a positive integer,

	$\left(n-1\right)\frac{\pi}{6}=k\pi,\quad k\in\mathbb{Z}^{+}\cup\{0\}$ $n=6k+1,\quad k\in\mathbb{Z}^{+}\cup\{0\}$ Or $n=6k-5,\quad k\in\mathbb{Z}^{+}$ $\left\{n:n\in\mathbb{Z}^{+},n=6k-5,\quad k\in\mathbb{Z}^{+}\right\}$

5 A manufacturing company has three factories that produce packets of instant noodles. The manager wants to test whether the lead content in its latest batch of instant noodles produced exceeds the legally permitted levels. The number of packets of instant noodles produced from each factory for the latest batch is shown in the table below.

Factory	Number of packets produced
A	4000
B	2000
C	1000

To carry out the test, a sample of 100 packets will be chosen from this batch of instant noodles produced.

(i) Describe how the sample could be chosen using stratified sampling. [2]

(ii) State one advantage of using stratified sampling in this context. [1]

<table>
<tr><th>Q5</th><th>Solutions</th></tr>
<tr><td>(i)</td><td>

Factory	Number of packets to be selected for the sample
A	$=\dfrac{4000}{7000}\times 100 \approx 57$
B	$=\dfrac{2000}{7000}\times 100 \approx 29$
C	$=\dfrac{1000}{7000}\times 100 \approx 14$

From each factory, select **randomly** the number of packets of instant noodles as according to the table above. The total number of packets selected form the stratified sample required.
</td></tr>
<tr><td>(ii)</td><td>Stratified sampling ensures that the sample chosen is representative of the batch of instant noodles produced by all the three factories, as it allows each factory to be represented proportionally.</td></tr>
</table>

6 A jackpot game machine at an arcade contains 4 slots where each of the first 2 slots displays any of the twelve zodiac signs and each of the next 2 slots display any of the twenty-six letters of the alphabets A–Z. The jackpot is won if the 4 slots display two identical zodiac signs and two identical letters. Find the probability that a random game played at the machine results in

(i) two different zodiac signs and two different vowels, [2]

(ii) winning the jackpot, [2]

(iii) exactly two identical zodiac signs or exactly two identical letters or both. [3]

Q6	**Solutions**
(i)	Prob required $=\frac{12}{12}\times\frac{11}{12}\times\frac{5}{26}\times\frac{4}{26}$ $=\frac{55}{2028}$
(ii)	P(Winning jackpot) $=\frac{12}{12}\times\frac{1}{12}\times\frac{26}{26}\times\frac{1}{26}$ $=\frac{1}{312}$
(iii)	Let A be the event for 2 identical Zodiac Signs. Let B be the event for 2 identical letters. $\mathrm{P}(A\cup B)=\mathrm{P}(A)+\mathrm{P}(B)-\mathrm{P}(A\cap B)$ $=\frac{12}{12}\times\frac{1}{12}+\frac{26}{26}\times\frac{1}{26}-\frac{1}{312}$ $=\frac{37}{312}$ **(Method 2)** Prob required = P(same zodiac signs and different letters) +P(different zodiac signs and same letters) +P(same zodiac signs and same letters) $=\left(\frac{12}{12}\times\frac{1}{12}\times\frac{26}{26}\times\frac{25}{26}\right)+\left(\frac{12}{12}\times\frac{11}{12}\times\frac{26}{26}\times\frac{1}{26}\right)+\frac{1}{312}$ $=\frac{37}{312}$

	(Method 3) Prob required = $= 1 - \text{P(different zodiac signs and different letters)}$ $= 1 - \left(\frac{12}{12} \times \frac{11}{12} \times \frac{26}{26} \times \frac{25}{26} \right)$ $= \frac{37}{312}$

7 Bernard is carrying out an experiment with a fair tetrahedral die, which has its four triangular faces numbered from '6' to '9', and a biased 10-sided die numbered from '1' to '10'.

(i) Bernard rolls the fair die 9 times. Find the probability that the die shows a '8' between 3 and 7 times, inclusive. [2]

(ii) Bernard now rolls the fair die 65 times. Use a suitable approximate distribution, which should be stated, to find the probability that the die shows a '9' more than 12 times. [3]

The probability that the biased die shows a '9' is $\frac{1}{25}$.

(iii) Bernard rolls the biased die 65 times. Use a suitable approximate distribution, which should be stated, to find the probability that the biased die shows a '9' more than 5 times. [3]

Q7	**Solutions**
(i)	Let X be the random variable denoting the number of '8's out of 9 rolls. $X \sim \mathrm{B}\left(9,\ \frac{1}{4}\right)$ $\mathrm{P}(3 \le X \le 7) = \mathrm{P}(X \le 7) - \mathrm{P}(X \le 2)$ $= 0.3992156$ ≈ 0.399
(ii)	Let Y be the random variable denoting the number of '9's out of 65 rolls. $Y \sim \mathrm{B}\left(65,\ \frac{1}{4}\right)$ Since $n = 65$ is large, $np = 65 \times \frac{1}{4} = 16.25 > 5$ and $nq = 65 \times \left(1 - \frac{1}{4}\right) = 48.75 > 5$ $Y \sim \mathrm{N}\left(16.25,\ 12.1875\right)$ approximately $\mathrm{P}(Y > 12) = \mathrm{P}(Y \ge 12.5)$ by continuity correction $= 0.858627$ ≈ 0.859
(iii)	Let W be the random variable denoting the number of '9's out of 65 rolls

	of the biased die. $W \sim \mathrm{B}\left(65,\ \frac{1}{25}\right)$ Since $n = 65$ is large, $np = 65 \times \frac{1}{25} = 2.6 < 5$ $W \sim \mathrm{Po}(2.6)$ approximately $\mathrm{P}(W > 5) = 1 - \mathrm{P}(W \leq 5)$ ≈ 0.0490 (3 sig fig)

8 A large office building is busy during the five weekdays, Monday to Friday, and less busy during the two weekend days, Saturday and Sunday. The block is illuminated by fluorescent light tubes which frequently fail and must be replaced with new tubes. It is assumed that the number of fluorescent tubes that fail on a particular weekday has the distribution $\text{Po}(1.2)$. The number of fluorescent tubes that fail on a particular weekend day is also assumed to be an independent random variable with distribution $\text{Po}(0.5)$.

(i) Find the probability that at least 8 fluorescent light tubes fail in a period of five consecutive weekdays. [2]

A week refers to a complete seven-day week.

(ii) Given that a total of 10 fluorescent light tubes fail during a week, find the probability that at most 2 fluorescent light tubes fail during Saturday and Sunday. [3]

(iii) Using a suitable approximation, find the probability that at most 30 fluorescent light tubes fail during a period of 4 weeks. [3]

Q8	**Solutions**
(i)	Let X be the r.v. denoting the number of fluorescent light tubes that fail in a period of 5 weekdays. $X \sim \text{Po}(6)$ $\text{P}(X \geq 8) = 1 - \text{P}(X \leq 7) = 0.256$ (3 sig fig)
(ii)	Let Y be the r.v. denoting the number of fluorescent light tubes that fail in a period of 2 weekend days. $Y \sim \text{Po}(1)$ and $X + Y \sim \text{Po}(7)$ $\text{P}(Y \leq 2 \mid X + Y = 10)$ $= \dfrac{\text{P}(Y=0)\text{P}(X=10) + \text{P}(Y=1)\text{P}(X=9) + \text{P}(Y=2)\text{P}(X=8)}{\text{P}(X+Y=10)}$ $= 0.838$
	Let T be the r.v. denoting the number of fluorescent light tubes that fail in a period 4 weeks. $T \sim \text{Po}(28)$ Since the mean of T = 28 (>10), $\therefore T \sim \text{N}(28, 28)$ approximately. $\text{P}(X \leq 30) = \text{P}(X \leq 30.5)$ (with continuity corrections) $= 0.682$ (3 sig fig)

9 The continuous random variable X has the distribution $\mathrm{N}(\mu,\sigma^2)$. It is known that $\mathrm{P}\left(X > 2a\right) = 0.10$ and $\mathrm{P}\left(X < a\right) = 0.30$.

(i) Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ in terms of a. [5]

(ii) Given that X_1, X_2 and X_3 are three independent observations of X, find $\mathrm{P}\left(X_1 + X_2 - 2X_3 > a\right)$. [4]

Q9	**Solutions**
(i)	$X \sim \mathrm{N}\left(\mu,\sigma^2\right)$ $\mathrm{P}\left(X > 2a\right) = 0.10$ $\mathrm{P}\left(X < 2a\right) = 0.90$ $\mathrm{P}\left(Z < \frac{2a-\mu}{\sigma}\right) = 0.90$ $\frac{2a-\mu}{\sigma} = 1.28155$ $2a - \mu = 1.28155\sigma$ ----------- (1) $\mathrm{P}\left(X < a\right) = 0.30$ $\mathrm{P}\left(Z < \frac{a-\mu}{\sigma}\right) = 0.30$ $\frac{a-\mu}{\sigma} = -0.52440$ $a - \mu = -0.52440\sigma$ --------- (2) $(1)-(2)$: $a = 1.80595\sigma$ $\sigma = \frac{a}{1.80595} = 0.553725a$ $a - \mu = -0.52440 \times \frac{a}{1.80595}$ $= -0.29037a$ $\mu = 1.29a$ $\mathrm{E}(X) = \mu = 1.29a$ $\mathrm{Var}(X) = \sigma^2 = 0.553725^2 a^2 = 0.30661a^2 = 0.307a^2$

(ii)	$\mathrm{E}(X_1 + X_2 - 2X_3) = \mathrm{E}(X) + \mathrm{E}(X) - 2\mathrm{E}(X) = 0$ $\begin{aligned}\mathrm{Var}(X_1 + X_2 - 2X_3) &= \mathrm{Var}(X) + \mathrm{Var}(X) + 4\mathrm{Var}(X) \\ &= 6\mathrm{Var}(X) \\ &= 1.83966526a^2\end{aligned}$ $X_1 + X_2 - 2X_3 \sim \mathrm{N}(0, 1.83966526a^2)$ $\begin{aligned}\mathrm{P}(X_1 + X_2 - 2X_3 > a) &= \mathrm{P}\left(Z > \frac{a-0}{\sqrt{1.83966526a^2}}\right) \\ &= \mathrm{P}(Z > 0.737276848) \\ &\approx 0.230\end{aligned}$

10 A firm of solicitors claims that the average duration of the interviews with their clients is 45 minutes.

(a) A random sample of 12 interviews is chosen, and the time taken for each interview, x minutes, is noted. The results are shown in the following data.

53 40 61 48 51 43 50 35 42 55 65 60

(i) Calculate unbiased estimates of the population mean and variance. [2]

(ii) Stating a necessary assumption, carry out a test to determine, at the 5% significance level, whether the firm is understating the average interview time. You should define any symbols that you use. [5]

(b) Another sample of 60 interviews is chosen and the time taken for each interview, in minutes, is noted. The sample mean is found to be m minutes and the sample standard deviation is 9 minutes. A test is carried out at the 10% level of significance to determine the validity of the firm's claim. Find the set of values of m for which the firm's claim is not valid. [5]

Q10	Solutions
(a) (i)	Let X minutes be the r.v. denoting the duration of an interview with a randomly chosen client. Using GC, an unbiased estimate of the population mean $= \bar{x} = 50.25$ an unbiased estimate of the population variance $= s^2 = 9.156468156^2 = 83.8$
(ii)	$\mathrm{H_o} : \mu = 45$ $\mathrm{H_1} : \mu > 45$ where μ represents the population mean duration of an interview measured in minutes. Level of significance: 5 % Assumption: We need to assume that X is normally distributed, so that a t-test can be used. Under $\mathrm{H_o}$, $T = \dfrac{\bar{X} - 45}{S/\sqrt{12}} \sim t_{(11)}$. From GC, p-value = 0.0362 < 0.05 $\therefore$ Reject $\mathrm{H_o}$ and conclude that there is significant evidence, at 5% significance level, that the mean duration of an interview at the firm is more than 45 minutes, i.e. the firm is understating average interview time.

(b) $H_o : \mu = 45$ (claim)

$H_1 : \mu \neq 45$

Level of significance: 10 %

Under H_o , $Z = \dfrac{\bar{X} - 45}{s/\sqrt{60}} \sim N(0,1)$ approximately by Central Limit Theorem, since $n = 60$ is large.

$$s^2 = \frac{60}{59} \times 9^2 = 82.37288136$$

For the firm's claim to be invalid, the null hypothesis is to be rejected.

$\Rightarrow$ z-value falls inside critical region.

$z\text{-value} < -1.644853$ or $z\text{-value} > 1.644853$

$$\frac{m-45}{\sqrt{82.3728/60}} < -1.644853 \text{ or } \frac{m-45}{\sqrt{82.3728/60}} > 1.644853$$

$m < 43.072$ or $m > 46.927$

$m < 43.1$ or $m > 46.9$ (3 sig fig)

Required set is $\{m : m \in \mathbb{R} ,\ m < 43.1 \text{ or } m > 46.9\}$

<u>Alternative solution:</u>

$s^2 = 9^2 = 81$

For the firm's claim to be invalid, the null hypothesis is to be rejected.

$\Rightarrow$ z-value falls inside critical region.

$z\text{-value} < -1.644853$ or $z\text{-value} > 1.644853$

$$\frac{m-45}{\sqrt{81/60}} < -1.644853 \text{ or } \frac{m-45}{\sqrt{81/60}} > 1.644853$$

$m < 43.089$ or $m > 46.911$

$m < 43.1$ or $m > 46.9$ (3 sig fig)

Required set is $\{m : m \in \mathbb{R}^+,\ m < 43.1 \text{ or } m > 46.9\}$

11 **(a)** A random sample of six students is taken from those who sat for Mathematics and General Paper examinations, and their marks, x and y, each out of 100, are given in the table.

Mathematics (x)	34	41	70	86	62	56
General Paper (y)	32	34	63	55	45	58

(i) Sketch a scatter diagram for the data. [1]

(ii) Find the equations of the regression line of

(a) y on x , **(b)** x on y. [2]

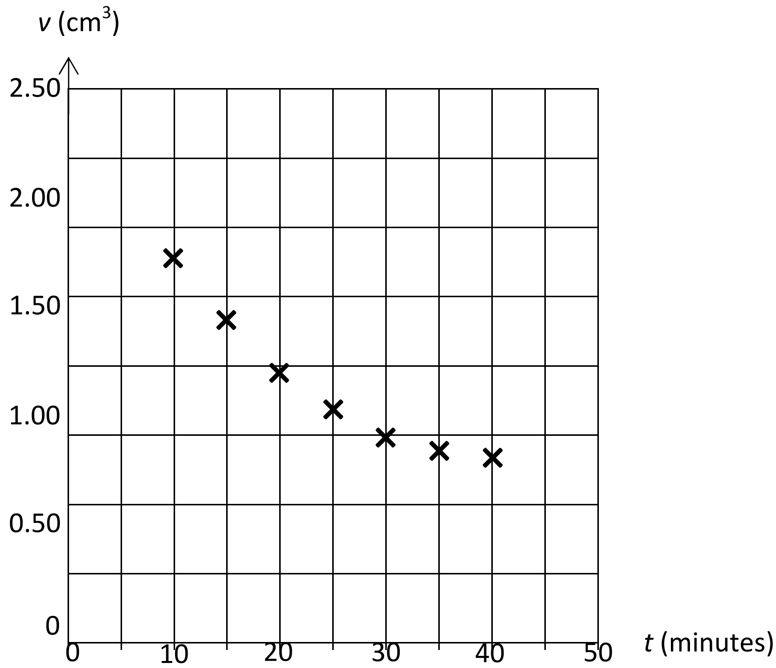
(iii) Sketch the above two regression lines on the scatter diagram in part **(i)** and mark the point $(\overline{x}, \overline{y})$ on the diagram. [3]

(b) An experiment is being conducted to measure how the volume of a substance, v (in cm^3), varies with time t (in minutes) in a particular chemical reaction. The results are given in the table.

t (minutes)	10	15	20	25	30	35	40
v (cm^3)	1.75	1.47	1.22	1.05	0.94	0.88	0.85

(i) Calculate, correct to 4 decimal places, the value of the product moment correlation coefficient betweeen v and t. Explain whether your answer suggests that a linear model is appropriate. [2]

The scatter diagram for the data is shown below.



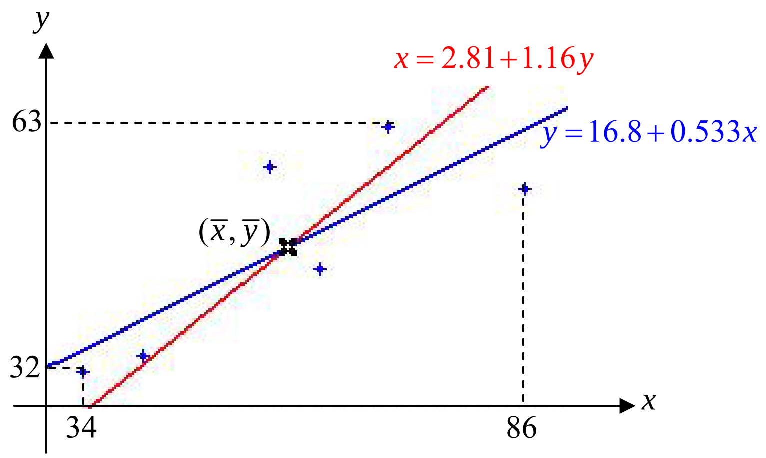
$\rightarrow$

It is proposed that the volume of the substance v can be modelled by the formula $v = a + \frac{b}{t}$, where a and b are constants, and $b > 0$.

(ii) Explain why the scatter diagram for the data is consistent with this proposed model. [1]

(iii) Calculate, correct to 4 decimal places, the value of the product moment correlation coefficient betweeen v and $\frac{1}{t}$. Comment on the value obtained. [2]

(iv) Use a suitable regression line to give the best estimate that you can of the volume of the substance at the 28th minute of the chemical reaction. [2]

Q11	**Solutions**
(a) **(i)** **(iii)**	 $(\bar{x}, \bar{y}) = (58.2, 47.8)$
(ii)	Regression line y on x: $y = 16.8328 + 0.5329594x$ $\Rightarrow$ $y = 16.8 + 0.533x$ Regression line x on y: $x = 2.80834 + 1.15731y$ $\Rightarrow$ $x = 2.81 + 1.16y$
(b) **(i)**	$r = -0.9516$ Acceptable answers: • The value of r is close to -1 indicating a very strong negative linear correlation between v and t. Therefore, a linear model is

	appropriate. • It is not appropriate because a linear model may result in volume of substance becomes negative. • The scatter diagram shows a curvilinear correlation between v and t, so a linear model is not appropriate.
(ii)	From the scatter diagram, it can be seen that as t increases, v decreases at a decreasing rate. This is consistent with the model $v = a + \frac{b}{t}$. (Or equivalently: from the scatter diagram, it can be seen that the data points follow a pattern that is decreasing and is concave upwards.)
(iii)	$r = 0.9900$ (4 d.p) The absolute value of r (0.9900) between v and $\frac{1}{t}$ is **<u>closer</u>** to 1 as compared to the absolute value of r (0.9516) for the linear model, hence indicating a **<u>stronger</u>** linear correlation between v and $\frac{1}{t}$. Therefore, $v = a + \frac{b}{t}$ is a better model.
(iv)	$v = 0.548749 + \frac{12.57018}{t}$ When $t = 28$, $v = 0.548749 + \frac{12.57018}{28}$ $= 0.997684$ $= 0.998$ (3 sig fig)