H2 Mathematics 2017 Preliminary Exam Paper 2 Solution

1 Given that $\sin\left[(n+1)x\right]-\sin\left[(n-1)x\right]=2\cos nx\sin x$, show that

$$\sum_{r=1}^{n}\cos rx=\frac{\sin\left(n+\frac{1}{2}\right)x-\sin\frac{1}{2}x}{2\sin\frac{1}{2}x}\quad. \qquad [4]$$

Hence express

$$\cos^2\left(\frac{x}{2}\right)+\cos^2(x)+\cos^2\left(\frac{3x}{2}\right)+\ldots+\cos^2\left(\frac{11x}{2}\right)$$

in the form $a\left(\dfrac{\sin bx}{\sin cx}+d\right)$, where a, b, c and d are real numbers. [3]

<u>Given</u>: $2\cos nx\sin x=\sin(n+1)x-\sin(n-1)x$

Thus, $\displaystyle\sum_{r=1}^{n}\cos rx=\sum_{r=1}^{n}\frac{\sin(r+1)x-\sin(r-1)x}{2\sin x}$

$$=\frac{1}{2\sin x}\begin{bmatrix}\cancel{\sin 2x}-0\\ +\cancel{\sin 3x}-\sin x\\ +\cancel{\sin 4x}-\cancel{\sin 2x}\\ \vdots\\ +\sin nx-\cancel{\sin(n-2)x}\\ +\sin(n+1)x-\cancel{\sin(n-1)x}\end{bmatrix}$$

Adding the n equations above, we have

$$\sum_{r=1}^{n}\cos rx=\frac{\sin(n+1)x+\sin nx-\sin x}{2\sin x}$$

$$=\frac{2\sin(n+\frac{1}{2})x\cos(-\frac{1}{2})x-\sin x}{2\sin x}\quad\text{using factor formula}$$

$$=\frac{2\sin(n+\frac{1}{2})x\cos\frac{1}{2}x-2\sin\frac{1}{2}x\cos\frac{1}{2}x}{4\sin\frac{1}{2}x\cos\frac{1}{2}x}$$

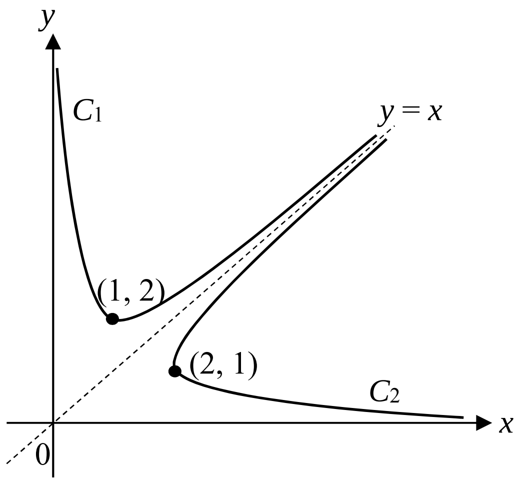
$$=\frac{\sin(n+\frac{1}{2})x-\sin\frac{1}{2}x}{2\sin\frac{1}{2}x}$$

$$\cos^2\left(\frac{x}{2}\right)+\cos^2(x)+\cos^2\left(\frac{3x}{2}\right)+\ldots+\cos^2\left(\frac{11x}{2}\right)$$

$$=\frac{1+\cos x}{2}+\frac{1+\cos 2x}{2}+\frac{1+\cos 3x}{2}+\ldots+\frac{1+\cos 11x}{2}$$

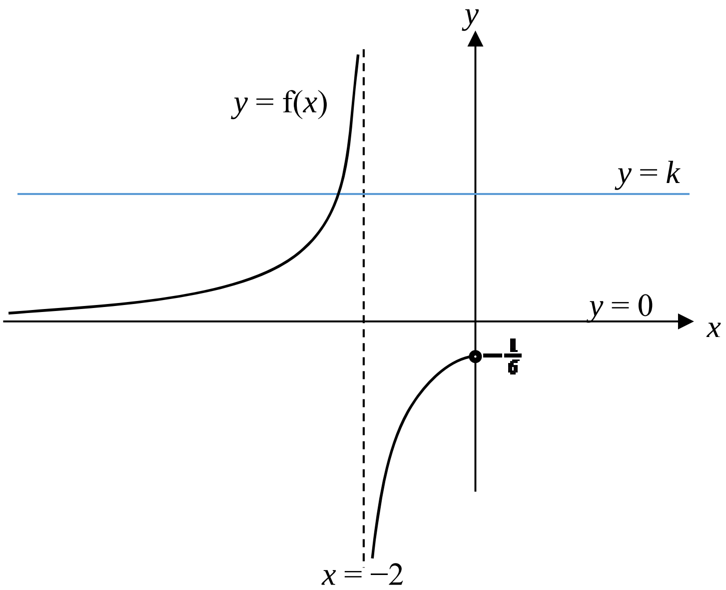
$$=\frac{1}{2}\left(11+\sum_{r=1}^{11}\cos rx\right)$$

$$=\frac{1}{2}\left[11+\frac{\sin(11+\frac{1}{2})x-\sin\frac{1}{2}x}{2\sin\frac{1}{2}x}\right]$$

<table>
<tr><td></td><td>$$=\frac{1}{2}\left(11+\frac{\sin\frac{23}{2}x}{2\sin\frac{1}{2}x}-\frac{1}{2}\right)$$
$$=\frac{1}{2}\left(\frac{\sin\frac{23}{2}x}{2\sin\frac{1}{2}x}+\frac{21}{2}\right)=\frac{1}{4}\left(\frac{\sin\frac{23}{2}x}{\sin\frac{1}{2}x}+21\right)$$</td></tr>
<tr><td>2</td><td>

(a) The diagram above shows two curves C_1 and C_2 which are reflections of each other about the line $y = x$. State with justification, whether the following statement is true: "If C_1 is the graph of $y=\mathrm{f}(x)$, then C_2 is the graph of $y=\mathrm{f}^{-1}(x)$."
(b) The functions f and g are defined as follows
$$\mathrm{f}: x \mapsto \frac{1}{x^2-x-6},\quad x\in\mathbb{R},\ x<0,\ x\neq -2$$
$$\mathrm{g}: x \mapsto \tan^{-1}\left(\frac{x}{2}\right),\quad x\in\mathbb{R}$$
(i) Sketch the graph of $y=\mathrm{f}(x)$. Determine whether f^2 exists. [3]
(ii) Find $\mathrm{f}^{-1}(x)$. [2]
(iii) Given that $\mathrm{gf}(a)=\frac{\pi}{4}$, find the exact value of a. [2]</td></tr>
<tr><td></td><td>(a) From the graph of $y=\mathrm{f}(x)$ which is C_1, there exists a horizontal line $y = 3$ which cuts the graph of $y=\mathrm{f}(x)$ at 2 points. f is not one to one and thus f^{-1} does not exist. Since f^{-1} does not exist, C_2 is not the graph of $\mathrm{f}^{-1}(x)$.

(b) (i)</td></tr>
</table>



Since $R_f = \left(-\infty, -\frac{1}{6}\right) \cup (0, \infty)$ and $D_f = (-\infty, 0)$ i.e. $R_f \not\subset D_f$. Therefore, f^2 does not exist.

(ii) Let $y = \dfrac{1}{x^2 - x - 6}$

$$\Rightarrow \quad yx^2 - xy - 6y - 1 = 0$$

$$\Rightarrow \quad x = \frac{y \pm \sqrt{y^2 + 4y(6y+1)}}{2y}$$

$$\Rightarrow \quad x = \frac{y \pm \sqrt{25y^2 + 4y}}{2y}$$

Since $x < 0$, $x = \dfrac{y - \sqrt{25y^2 + 4y}}{2y} = \dfrac{1}{2} - \dfrac{\sqrt{25y^2 + 4y}}{2y}$

Thus, $f^{-1}(x) = \dfrac{1}{2} - \dfrac{\sqrt{25x^2 + 4x}}{2x}$.

(iii) $gf(a) = \dfrac{\pi}{4} \quad \Rightarrow \quad \tan^{-1}\left(\dfrac{f(a)}{2}\right) = \dfrac{\pi}{4} \quad \Rightarrow \quad \dfrac{f(a)}{2} = 1 \quad \Rightarrow \quad f(a) = 2$

$$\therefore \quad a = f^{-1}(2) = \frac{1}{2} - \frac{\sqrt{25(2)^2 + 4(2)}}{2(2)} = \frac{1}{2} - \frac{3\sqrt{3}}{2}$$

3 Given that $e^y = \sqrt{e + x + \sin x}$. Show that

$$2e^{2y}\frac{d^2y}{dx^2} + 4e^{2y}\left(\frac{dy}{dx}\right)^2 + \sin x = 0.$$ [2]

	(i) Find the values of y, $\frac{\mathrm{d}y}{\mathrm{d}x}$ and $\frac{\mathrm{d}^2 y}{\mathrm{d}x^2}$ when $x=0$. Hence, find in terms of e, the Maclaurin's series for $\ln(\mathrm{e}+x+\sin x)$, up to and including the term in x^2. [4] **(ii)** By using appropriate standard series expansions from the List of Formulae (MF26), verify the correctness of the first three terms in the series expansion for $\ln(\mathrm{e}+x+\sin x)$ found in part **(i)**. [3] **(iii)** Use your answer to part **(i)** to give an approximation for $\int_0^{\mathrm{e}^{-1}} \frac{2\mathrm{e}-4x}{\mathrm{e}^2\ln(\mathrm{e}+x+\sin x)}\,\mathrm{d}x$, giving your answer in terms of e. [3]
	$\mathrm{e}^y=\sqrt{\mathrm{e}+x+\sin x} \quad\Rightarrow\quad \mathrm{e}^{2y}=\mathrm{e}+x+\sin x$ Differentiate wrt x: $\mathrm{e}^{2y}\left(2\frac{\mathrm{d}y}{\mathrm{d}x}\right)=1+\cos x$ i.e. $2\mathrm{e}^{2y}\frac{\mathrm{d}y}{\mathrm{d}x}=1+\cos x$ Differentiate wrt x: $2\left[\mathrm{e}^{2y}\frac{\mathrm{d}^2y}{\mathrm{d}x^2}+\frac{\mathrm{d}y}{\mathrm{d}x}\mathrm{e}^{2y}\left(2\frac{\mathrm{d}y}{\mathrm{d}x}\right)\right]=-\sin x$ i.e. $2\mathrm{e}^{2y}\frac{\mathrm{d}^2y}{\mathrm{d}x^2}+4\mathrm{e}^{2y}\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2+\sin x=0$ (Shown) (i) When $x=0$, $\mathrm{e}^{2y}=\mathrm{e}+0+0 \Rightarrow y=\frac{1}{2}$ $2\mathrm{e}\frac{\mathrm{d}y}{\mathrm{d}x}=1+1 \Rightarrow \frac{\mathrm{d}y}{\mathrm{d}x}=\frac{1}{\mathrm{e}}$ $2\mathrm{e}\frac{\mathrm{d}^2y}{\mathrm{d}x^2}+4\mathrm{e}\left(\frac{1}{\mathrm{e}}\right)^2+0=0 \Rightarrow \frac{\mathrm{d}^2y}{\mathrm{d}x^2}=\frac{-2}{\mathrm{e}^2}$ $\therefore\ y=\frac{1}{2}+\frac{1}{\mathrm{e}}x-\frac{2}{\mathrm{e}^2}\left(\frac{x^2}{2!}\right)+\cdots$ $\mathrm{e}^{2y}=\mathrm{e}+x+\sin x \Rightarrow \ln\left(\mathrm{e}+x+\sin x\right)=2y=2\left(\frac{1}{2}+\frac{1}{\mathrm{e}}x-\frac{2}{\mathrm{e}^2}\frac{x^2}{2!}+\cdots\right)$ i.e., $\ln(\mathrm{e}+x+\sin x)=1+\frac{2}{\mathrm{e}}x-\frac{2}{\mathrm{e}^2}x^2+\cdots$ (ii) $\ln(\mathrm{e}+x+\sin x)$ $=\ln(\mathrm{e}+x+x+...)$ $=\ln(\mathrm{e}+2x+...)$ $=\ln\left(\mathrm{e}\left(1+\frac{2}{\mathrm{e}}x+...\right)\right)$ $=\ln\mathrm{e}+\ln\left(1+\frac{2}{\mathrm{e}}x+...\right)$ $=1+\ln\left(1+\frac{2}{\mathrm{e}}x+...\right)$

$$= 1 + \left[\left(\frac{2}{e} x \right) - \frac{1}{2} \left(\frac{2}{e} x \right)^2 + \cdots \right]$$

$$= 1 + \frac{2}{e} x - \frac{2}{e^2} x^2 + \cdots \qquad \text{(Verified)}$$

(iii) $\displaystyle\int_0^{e^{-1}} \frac{2e - 4x}{e^2 \ln(e + x + \sin x)} \, dx$

$$\approx \int_0^{e^{-1}} \frac{2e - 4x}{e^2 \left(1 + \frac{2}{e} x - \frac{2}{e^2} x^2 \right)} \, dx$$

$$= \int_0^{e^{-1}} \frac{2e - 4x}{e^2 + 2ex - 2x^2} \, dx$$

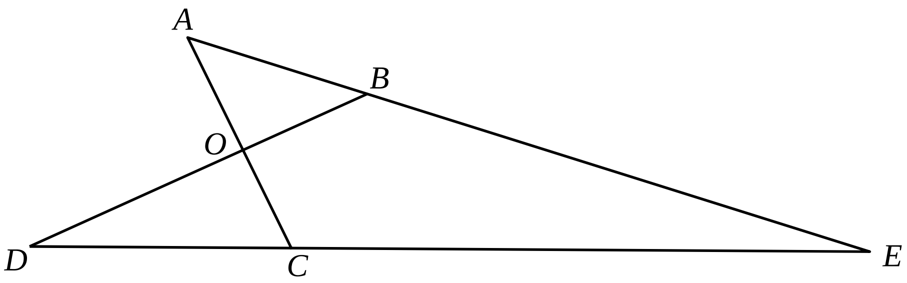
$$= \left[\ln(e^2 + 2ex - 2x^2) \right]_0^{\frac{1}{e}}$$

$$= \ln\left[e^2 + 2 - 2\left(\tfrac{1}{e}\right)^2 \right] - \ln(e^2)$$

$$= \ln\left(e^2 + 2 - \frac{2}{e^2} \right) - 2$$

$$= \ln(e^4 + 2e^2 - 2) - 4$$

4



With reference to origin O, the points A, B, C and D are such that $\overrightarrow{OA} = \mathbf{a}$, $\overrightarrow{OB} = \mathbf{b}$, $\overrightarrow{OC} = -\mathbf{a}$ and $\overrightarrow{OD} = -2\mathbf{b}$. The lines AB and DC meet at E.

Find $\overrightarrow{OE}$ in terms of $\mathbf{a}$ and $\mathbf{b}$. [4]

Hence show that $\dfrac{BE}{AB} = 3$. [1]

It is given that A and E have coordinates $(1, -4, 3)$ and $(-3, 15, -5)$ respectively.

(i) Show that the lines AC and BD are perpendicular. [4]

(ii) Find the equation of the plane p that contains E and is perpendicular to the line BD. [2]

(iii) Find the distance between the line AC and p. [2]

Equation of line AB is $\mathbf{r} = \mathbf{a} + \lambda(\mathbf{b} - \mathbf{a})$.

Equation of line DC is $\mathbf{r} = -\mathbf{a} + \mu(-2\mathbf{b} + \mathbf{a})$.

To find E, the point of intersection of lines AB and CD,

consider $\quad \mathbf{a} + \lambda(\mathbf{b} - \mathbf{a}) = -\mathbf{a} + \mu(-2\mathbf{b} + \mathbf{a})$

$$\Rightarrow \quad (1 - \lambda)\mathbf{a} + \lambda\mathbf{b} = (-1 + \mu)\mathbf{a} - 2\mu\mathbf{b}$$

$$\Rightarrow \quad (2 - \lambda - \mu)\mathbf{a} = (-2\mu - \lambda)\mathbf{b}$$

	Since **a** is not parallel to **b**, $\begin{cases} 2-\mu-\lambda=0 & \cdots(1) \\ -2\mu-\lambda=0 & \cdots(2) \end{cases}$ Solving (1) and (2), we have $\mu=-2$ and $\lambda=4$ $\therefore\ \overrightarrow{OE}=\mathbf{a}+4(\mathbf{b}-\mathbf{a})=-3\mathbf{a}+4\mathbf{b}$ $\therefore\ \overrightarrow{BE}=\overrightarrow{OE}-\overrightarrow{OB}=-3\mathbf{a}+4\mathbf{b}-\mathbf{b}=3(\mathbf{b}-\mathbf{a})=3\overrightarrow{AB}$ $\therefore\ \dfrac{BE}{AB}=3$ (i) $\overrightarrow{OE}=-3\mathbf{a}+4\mathbf{b}=\begin{pmatrix}-3\\15\\-5\end{pmatrix} \Rightarrow -3\begin{pmatrix}1\\-4\\3\end{pmatrix}+4\mathbf{b}=\begin{pmatrix}-3\\15\\-5\end{pmatrix}$ $\Rightarrow 4\mathbf{b}=\begin{pmatrix}-3\\15\\-5\end{pmatrix}+3\begin{pmatrix}1\\-4\\3\end{pmatrix} \Rightarrow \mathbf{b}=\begin{pmatrix}0\\\frac{3}{4}\\1\end{pmatrix}$ $\mathbf{a}\cdot\mathbf{b}=\begin{pmatrix}0\\-4\\3\end{pmatrix}\cdot\begin{pmatrix}0\\\frac{3}{4}\\1\end{pmatrix}=-4\left(\dfrac{3}{4}\right)+3=0$ $\Rightarrow$ *OA* and *OB* are perpendicular $\Rightarrow$ *AC* and *BD* are perpendicular (as *AC* is parallel to *OA* and *BD* is parallel to *OB*) (ii) Equation of the plane *p* is $\mathbf{r}\cdot\begin{pmatrix}0\\3\\4\end{pmatrix}=\begin{pmatrix}-3\\15\\-5\end{pmatrix}\cdot\begin{pmatrix}0\\3\\4\end{pmatrix}$, i.e. $\mathbf{r}\cdot\begin{pmatrix}0\\3\\4\end{pmatrix}=25$ (iii) Distance between the line *AC* and the plane *p* = distance of *O* from *p* $=\dfrac{\begin{pmatrix}-3\\15\\-5\end{pmatrix}\cdot\begin{pmatrix}0\\3\\4\end{pmatrix}}{\sqrt{3^2+4^2}}=5$
5	Four classes CG40, CG41, CG42 and CG43 are tasked to organise a College event. Each class sends 3 representatives for a meeting. **(i)** In how many different ways can the 12 representatives sit in a circle so that representatives from CG40 are not seated next to each other and representatives from other classes are seated with their respective classes? [3] The 12 representatives are to be split up into 3 groups for bonding activities. Each group must consist of a representative from each class. **(ii)** In how many ways can the groups be formed? [2]
	(i) Number of ways to arrange the 3 classes except CG40 $=(3!)^3(3-1)!$ Number of ways to arrange reps from CG40 for a particular arrangement of the other 3 classes = 3!

	Required number $=(3!)^3(3-1)!3!=$ 2592 (ii) Required number $=\dfrac{(3!)^4}{3!}=216$
6	In a game at the carnival, a player rolls discs onto a board containing squares, each of which bears one of the numbers 1, 2, 5 or 10. If a disc does not land within a square, the player receives nothing. The probability that the disc does not land within the square is $\frac{3}{4}$. If a disc lands within a square, the player receives the same amount (in dollars) as the number in the square. Given that a disc falls within a square, the probabilities of landing within a square with the numbers 1, 2, 5 and 10 are 0.5, 0.3, 0.12 and 0.08 respectively. It is assumed that the rolls of the discs are independent. **(i)** A player pays \$5 to play the game and is given n discs. Find n if the game is fair. [4] **(ii)** If a player is allowed to roll 3 discs for \$2, find the probability that the player will have a profit of \$10. [4]
	(i) Let Y (in dollars) be the amount received by a player for each roll. (see table below) $\mathrm{E}(Y)=\left(0\times\frac{3}{4}\right)+\left(1\times\frac{1}{8}\right)+\left(2\times\frac{3}{40}\right)+\left(5\times\frac{3}{100}\right)+\left(10\times\frac{1}{50}\right)=0.625$ For the game to be fair, $\mathrm{E}(Y_1+Y_2+\cdots+Y_n-5)=0$ $\Rightarrow\quad n\mathrm{E}(Y)-5=0$ $\Rightarrow\quad 0.625n=5$ $\Rightarrow\quad n=8$ (ii) Pay \$2 for 3 rolls with a gain of \$10 implies that the player needs to receive \$12 from 3 rolls. Required probability $=3!\times\mathrm{P}(Y_1=0,Y_2=2,Y_3=10)+3\times\mathrm{P}(Y_1=1,Y_2=1,Y_3=10)+3\times P\left(Y_1=2,Y_2=5,Y_3=5\right)$ $=3!\times\left(\frac{3}{4}\right)\left(\frac{3}{40}\right)\left(\frac{1}{50}\right)+3\times\left(\frac{1}{8}\right)^2\left(\frac{1}{50}\right)+3\times\left(\frac{3}{40}\right)\left(\frac{3}{100}\right)^2=0.0789$ (exact)
7	A factory manufactures large number of pen refills. From past records, 3% of the refills are defective. A stationery store manager wishes to purchase pen refills from the factory. To decide whether to accept or reject a batch of refills, the manager designs a sampling

y	0	1	2	5	10
$\mathrm{P}(Y=y)$	$\frac{3}{4}$	$\frac{1}{4}\times\frac{1}{2}=\frac{1}{8}$	$\frac{1}{4}\times\frac{3}{10}=\frac{3}{40}$	$\frac{1}{4}\times\frac{3}{25}=\frac{3}{100}$	$\frac{1}{4}\times\frac{8}{100}=\frac{1}{50}$

process. He takes a random sample of 25 refills. The batch is accepted if there is no defective refill and rejected if there are more than 2 defective refills. Otherwise, a second random sample of 25 refills is taken. The batch is then accepted if the total number of defective refills in the two samples is fewer than 4 and rejected otherwise.

(i) Find the probability of accepting a batch. [4]

(ii) If a batch is accepted, find the probability that there are 2 defective refills found in the sampling process. [3]

The stationery store manager purchases 50 boxes of 25 refills each.

(iii) Find the probability that the mean number of defective refills in a box is less than 1. [2]

Let X be the number of defective refills in the sample of 25 refills drawn from a batch which contains 3% defective refills.
Then, $X \sim \mathrm{B}\,(25, 0.03)$

(i) P(accepting a batch)

$= \mathrm{P}(X=0) + \mathrm{P}(X=1)\mathrm{P}(X \le 2) + \mathrm{P}(X=2)\mathrm{P}(X \le 1)$

$= 0.4669747053 + 0.3473570958 + 0.1109593034$

≈ 0.9252911

$= 0.925$ (correct to 3 s.f.)

(ii) Required probability

= P(2 defective refills | batch is accepted)

$$= \frac{\mathrm{P}(X_1 = 1)\mathrm{P}(X_2 = 1) + \mathrm{P}(X_1 = 2)\mathrm{P}(X_2 = 0)}{0.9252911} = 0.209 \text{ (correct to 3 s.f.)}$$

(iii) $X \sim \mathrm{B}\,(25, 0.03)$

Since *sample size* = 50 is large, by Central Limit Theorem,

$$\bar{X} \sim \mathrm{N}\left(25(0.03), \frac{25(0.03)(0.97)}{50}\right) \text{ approximately}$$

$$\bar{X} \sim \mathrm{N}(0.75, 0.1455)$$

Required probability $= P(\bar{X} < 1) = 0.981$ (correct to 3 s.f.)

8 A study is done to find out the relationship between the age of women and the steroid levels in the blood plasma. Sample data collected from 10 females with ages ranging from 8 years old to 35 years old is as shown below.

Age (years) x	8	11	14	17	20	23	26	29	32	35
Steroid Level (mmol/litre) L	4.2	11.1	16.3	19.0	25.5	26.2	24.1	33.5	20.8	17.4

(i) Give a sketch of the scatter diagram for the data. Identify the outlier and suggest a reason, in the context of the question, why this data pair is an outlier. [3]

For the remaining part of the question, the outlier is to be removed from the calculation.

(ii) Comment on the suitability of each of the following models. Hence determine the best model for predicting the steroid level of a female based on her age.

Model A: $L = a + b\ln x$

Model B: $L = c + d\left(x-25\right)^2$

Model C: $L = e + f\left(x-25\right)^4$

where a, b, c, d, e and f are constants. [3]

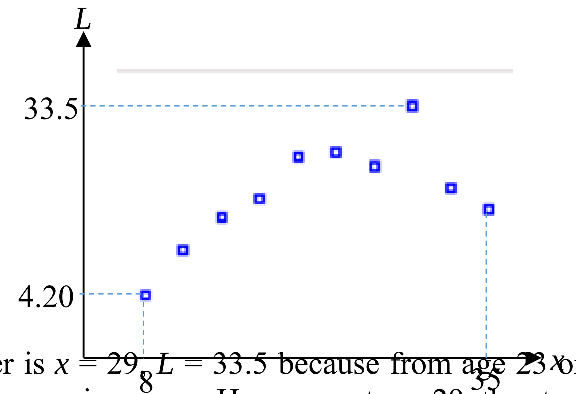
(iii) Using the best model in **(ii)**, estimate the steroid level of a woman at age 40. Comment on the reliability of your estimate. [3]

(iv) It is known that body muscle mass and steroid level has a linear correlation. The muscle mass percentage m % of the 9 females were measured. An additional female, Jane, participated in the study. Jane has her muscle mass percentage and steroid level measured. The mean muscle mass percentage of the 10 females is now found to be 26.28 %. The equation of the least squares regression line of m on L for the 10 pairs of data is

$$m = 2.22 + 1.25L.$$

Calculate Jane's steroid level. [3]

(i)



Outlier is x = 29, L = 33.5 because from age 23 onwards, there is a decreasing steroid level as age increases. However, at x = 29, the steroid level suddenly increases and this could be due to reasons such as illness/medication/pregnancy/intake of additional steroids by athlete/..etc (give any one of these reasons)

(ii) Model A is not suitable because as x increases, L is either only increasing (if $b > 0$) or only decreasing (if $b < 0$) which does not resemble the data trend of x and L whereby L increases but when it reaches about 23 years old, the steroid level decreases. Model B and C have similar trend as the given data set when $d < 0$ and $f < 0$ respectively and both are suitable models. But for model C, the r-value of L and $\left(x-25\right)^4$ is -0.929 and for model B, the r-value of L and $\left(x-25\right)^2$ is -0.987. Since for model B, the r value is closer to -1, therefore model B is a better model.

(iii) Least squares regression line equation is

	$L = 25.238 - 0.073667(x-25)^2$ i.e. $L = 25.2 - 0.0737(x-25)^2$ When $x = 40$, $L = 25.2 - 0.0737(40-25)^2 \approx 8.7$ The prediction is unreliable because $x = 40$ is outside the data range of 8 to 35 years old. (iv) Since $\overline{m} = 2.22 + 1.25\overline{L}$, $26.28 = 2.22 + 1.25\overline{L} \Rightarrow \overline{L} = 19.248$ $\sum_{i=1}^{10} L_i = 192.48$ and since $\sum_{i=1}^{9} L_i = 164.6$, therefore Jane's steroid level is 27.88 ≈ 27.9
9	A flange beam is a steel beam with a "H"-shaped cross section, and is used as a supporting structure in construction and civil engineering. A factory manufactures both Grade *X* and Grade *Y* flange beams. The load that can be supported by a Grade *X* flange beam follows a normal distribution with mean 2.43×10^5 kN and standard deviation 4.5×10^4 kN. The load that can be supported by a Grade *Y* flange beam is 1.5 times of the load that can be supported by a Grade *X* flange beam. **(i)** Find the probability that the combined load that can be supported by two randomly chosen Grade *Y* flange beams is within 1×10^4 kN of the combined load that can be supported by three randomly chosen Grade *X* flange beams. [4] **(ii)** A construction company wants to buy 100 sets of three Grade *X* flange beams. Find the probability that fewer than 95 of these sets can support more than 6×10^5 kN. [3] The company decides to place an order with the factory for a custom-made flange beam such that the probability of being able to support a load of at least 6×10^5 kN must be at least 0.999. It is also assumed that the load that can be supported by the custom-made flange beam also follows a normal distribution. **(iii)** By taking the standard deviation of a custom-made flange beam to be 3×10^4 kN, find the smallest possible mean load in kN, giving your answer correct to the nearest thousand, for the factory to meet the company's requirements for the custom-made flange beam. [5]
	Let *A* and *B* be the load that can be supported (in kN) by a Grade *X* and Grade *Y* flange beam respectively. Then, $A \sim \mathrm{N}(2.43\times10^5, (4.5\times10^4)^2)$ and since $B = 1.5A$, then $B \sim \mathrm{N}(1.5(2.43\times10^5), 1.5^2(4.5\times10^4)^2)$ i.e. $B \sim \mathrm{N}(3.645\times10^5, (6.75\times10^4)^2)$ (i) $\mathrm{P}\left[\lvert B_1 + B_2 - (A_1 + A_2 + A_3)\rvert < 1\times10^4\right]$ $\mathrm{E}\left((B_1+B_2) - (A_1+A_2+A_3)\right) = 2\times3.645\times10^5 - 3\times2.43\times10^5 = 0$ $\mathrm{Var}\left((B_1+B_2) - (A_1+A_2+A_3)\right) = 2\times(6.75\times10^4)^2 + 3(4.5\times10^4)^2 = 1.51875\times10^{10}$ i.e. $(B_1+B_2)-(A_1+A_2+A_3) \sim \mathrm{N}(0,\ 1.51875\times10^{10})$ $\mathrm{P}\left[\lvert B_1 + B_2 - (A_1 + A_2 + A_3)\rvert < 1\times10^4\right]$ $= \mathrm{P}\left[-1\times10^4 < (B_1+B_2)-(A_1+A_2+A_3) < 1\times10^4\right] = 0.0647$ (to 3 s.f.)

(ii) $A_1 + A_2 + A_3 \sim \mathrm{N}(3\times 2.43\times 10^5,\ 3(4.5\times 10^4)^2)$

$\mathrm{P}(A_1 + A_2 + A_3 > 6\times 10^5) \approx 0.951045$

Let T be the number of sets (out of 100 sets) of three Grade X flange beams that can support more than $6\times 10^5\,\mathrm{kN}$.

$T \sim \mathrm{B}(100,\ 0.951045)$

Required probability $= \mathrm{P}(T < 95) = \mathrm{P}\left(T \le 94\right) = 0.365$ (to 3 s.f.)

(iii) Let W be the load that can be supported (in kN) by a custom-made flange beam

Given: $W \sim \mathrm{N}(\mu, (3\times 10^4)^2)$

$$\mathrm{P}(W \ge 6\times 10^5) \ge 0.999$$

$$1 - \mathrm{P}(W < 6\times 10^5) \ge 0.999$$

$$\mathrm{P}(W < 6\times 10^5) \le 0.001$$

$$\mathrm{P}\left(Z \le \frac{6\times 10^5 - \mu}{3\times 10^4}\right) \le 0.001$$

By GC,

$$\frac{6\times 10^5 - \mu}{3\times 10^4} \le -3.0902$$

$$6\times 10^5 - \mu \le -92760$$

$$-\mu \le -92760 - 6\times 10^5$$

$$\mu \ge 692760$$

Thus, smallest mean = 693 kN (to nearest thousand)

10 **(a)** College students intending to further their studies overseas have to sit for a mandatory Overseas Universities Test (OUT). Researcher Mr Anand wishes to find out if male college students tend to score higher for OUT compared to female college students. Mr Anand's colleague randomly selects 150 male and 150 female students from the combined student population of three particular colleges near his home to form a sample of 300 college students for the research. Explain whether this sample is a random sample. [2]

(b) The mean OUT score for all college students in 2016 is 66.
Mr Anand randomly selects 240 college students taking OUT in 2017 and their scores, x, are summarised in the following table:

Score, x	60	65	68	70	75	80
Frequency, f	40	90	63	27	18	2

(i) Write down the unbiased estimates of the population mean and variance of the OUT scores for the college students in 2017. [1]

(ii) Test, at the 10% level of significance, whether the mean OUT score for all college students in 2017 is higher than the mean score attained in 2016. [4]

(iii) Explain what is meant by the phrase "10% level of significance" in this context. [1]

(iv) Mr Anand draws a new sample of 240 male college students. Using the unbiased estimate for the population variance computed in **(i)**, find the range of values for the sample mean $\overline{x}$ that is required for this new sample to achieve a different conclusion from that in **(ii)**. [4]

(c) The 2017 OUT scores of the male and female college students are independent and assumed to be normally distributed with means and standard deviations as shown in the following table:

	Mean	Standard deviation
Male College Students	64	5.5
Female College Students	66	3.5

Mr Beng and Miss Charlene both scored 70. Explain who performed better relative to their respective gender cohort. [2]

(a) Sample is non-random/biased since students from other colleges do not have any chance of being selected.

(b) (i) Using GC,

unbiased estimate of population mean, $\overline{x} = 66.391 = 66.4$ (to 3 s.f.)

and unbiased estimate of population variance, $s^2 = 4.1048^2 = 16.8$ (to 3 s.f.)

(ii) Let μ be the population mean OUT score of students in 2017 .

$\mathrm{H}_0 : \mu = 66$

$\mathrm{H}_1 : \mu > 66$

Level of significance: 10%

Test Statistic: $\dfrac{\overline{X} - \mu}{s/\sqrt{n}} \sim \mathrm{N}(0,1)$ by Central Limit Theorem since $n = 240$ is large

Under H_0, with $\overline{x} = 66.391$, $s = 4.1048$, $n = 240$, we have $p = 0.0697$

Since p-value < 0.1, we reject H_0. There is sufficient evidence at the 10% level of significance to conclude that the mean OUT score of male college students is higher than 66.

(iii) There is a probability of 0.1 of wrongly concluding that the mean OUT score of male college students is higher than 66.

(iv) $\mathrm{H}_0 : \mu = 66$

$\mathrm{H}_1 : \mu > 66$

Level of significance: 10%

Do not reject H_0, $p > 0.10$

$$\frac{\overline{x} - 66}{4.1048/\sqrt{240}} < 1.28155 \qquad \Rightarrow \qquad \overline{x} > 66.3396$$

$\therefore \quad \overline{x} > 66.3$ (to 3 s.f.)

(c) Let M and F be the OUT scores of male and female college students respectively

$M \sim \mathrm{N}(64, 5.5^2)$ and $F \sim \mathrm{N}(66, 3.5^2)$

$\mathrm{P}(M \le 70) = 0.86234 \Rightarrow$ Mr Beng is in the 86th percentile of male students

$\mathrm{P}(F \le 70) = 0.87345 \Rightarrow$ Miss Charlene is in the 87th percentile of female students

$\therefore$ Miss Charlene performed better relative to her gender