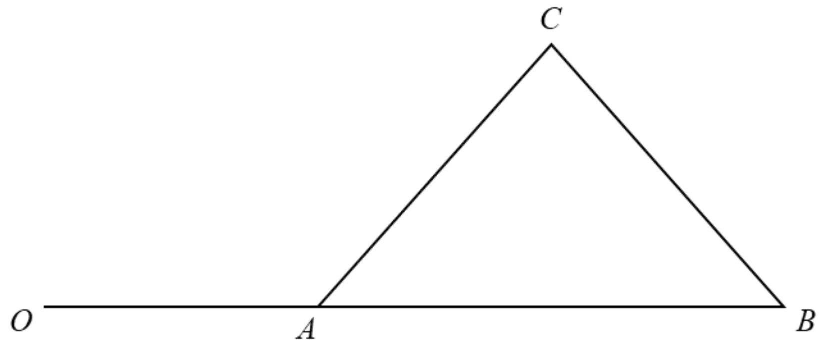
Section A: Pure Mathematics [40 marks]

1



Referred to the origin O, points A, B and C have position vectors **a**, **b** and **c** respectively. It is also given that OAB is a straight line (see diagram).

(i) Show that the area of triangle ABC can be written in the form $k\left|(\mathbf{b}-\mathbf{a})\times\mathbf{c}\right|$, where k is a constant to be determined. [3]

Area $=\frac{1}{2}\left\|\overrightarrow{AB}\times\overrightarrow{AC}\right\|$ $=\frac{1}{2}\left\|(\underset{\sim}{b}-\underset{\sim}{a})\times(\underset{\sim}{c}-\underset{\sim}{a})\right\|$ $=\frac{1}{2}\left\|(\underset{\sim}{b}-\underset{\sim}{a})\times\underset{\sim}{c}-(\underset{\sim}{b}-\underset{\sim}{a})\times\underset{\sim}{a}\right\|$ $=\frac{1}{2}\left\|(\underset{\sim}{b}-\underset{\sim}{a})\times\underset{\sim}{c}-\underset{\sim}{0}\right\|$ $\because (\underset{\sim}{b}-\underset{\sim}{a})//\underset{\sim}{a}\Rightarrow(\underset{\sim}{b}-\underset{\sim}{a})\times\underset{\sim}{a}=\underset{\sim}{0}$ $=\frac{1}{2}\left\|(\underset{\sim}{b}-\underset{\sim}{a})\times\underset{\sim}{c}\right\|$ $k=\frac{1}{2}$	Cross product gives vector, not scalar, so $\underset{\sim}{0}$, not 0. Need to write this as this is a 'show' question.
Alternative, Area $=\frac{1}{2}\left\|\overrightarrow{AB}\times\overrightarrow{CA}\right\|$ $=\frac{1}{2}\left\|(\underset{\sim}{b}-\underset{\sim}{a})\times(\underset{\sim}{a}-\underset{\sim}{c})\right\|$ $=\frac{1}{2}\left\|\underset{\sim}{b}\times\underset{\sim}{a}-\underset{\sim}{a}\times\underset{\sim}{a}-\underset{\sim}{b}\times\underset{\sim}{c}+\underset{\sim}{a}\times\underset{\sim}{c}\right\|$ $=\frac{1}{2}\left\|\underset{\sim}{0}-\underset{\sim}{0}-\underset{\sim}{b}\times\underset{\sim}{c}+\underset{\sim}{a}\times\underset{\sim}{c}\right\|$ $\because \underset{\sim}{a}//\underset{\sim}{b}\Rightarrow\underset{\sim}{b}\times\underset{\sim}{a}=\underset{\sim}{0}$ and $\underset{\sim}{a}\times\underset{\sim}{a}=\underset{\sim}{0}$ $=\frac{1}{2}\left\|\underset{\sim}{a}\times\underset{\sim}{c}-\underset{\sim}{b}\times\underset{\sim}{c}\right\|$ $=\frac{1}{2}\left\|(\underset{\sim}{a}-\underset{\sim}{b})\times\underset{\sim}{c}\right\|$ $=\frac{1}{2}\left\|-(\underset{\sim}{b}-\underset{\sim}{a})\times\underset{\sim}{c}\right\|=\frac{1}{2}\left\|(\underset{\sim}{b}-\underset{\sim}{a})\times\underset{\sim}{c}\right\|$ $k=\frac{1}{2}$	Cross product gives vector, not scalar, so $\underset{\sim}{0}$, not 0. Need to write this as this is a 'show' question. Note that: $(\underset{\sim}{b}-\underset{\sim}{a})\times\underset{\sim}{c}\neq\underset{\sim}{c}\times(\underset{\sim}{b}-\underset{\sim}{a})$ Instead, $(\underset{\sim}{b}-\underset{\sim}{a})\times\underset{\sim}{c}=-\underset{\sim}{c}\times(\underset{\sim}{b}-\underset{\sim}{a})$

It is given that $\overrightarrow{AB}$ is a unit vector and C is equidistant from A and B.

(ii) Give a geometrical interpretation of $\left|(\mathbf{a}-\mathbf{b})\times\mathbf{c}\right|$. [1]

It represents the perpendicular distance from C to the line (through) AB. OR It represents the height of the triangle ABC, with AB as the base.	$\left|(\mathbf{a}-\mathbf{b})\times\mathbf{c}\right|$ is a scalar, so think along the line of distances, magnitude, etc… A triangle has 3 heights, state the base of the triangle to distinguish the heights.

(iii) Show that OB has length $\left|\mathbf{c.b}-\mathbf{c.a}\right|+q$, where q is a constant to be determined. [3]

$OB=\left(\text{length of projection of } \overrightarrow{OC} \text{ on } \overrightarrow{AB}\right)+\dfrac{AB}{2}$ $=\left|\underset{\sim}{c}.\left(\underset{\sim}{b}-\underset{\sim}{a}\right)\right|+\dfrac{1}{2}$ $=\left|\underset{\sim}{c}.\underset{\sim}{b}-\underset{\sim}{c}.\underset{\sim}{a}\right|+\dfrac{1}{2}$	Keep in mind what's given in the question, that $\overrightarrow{AB}$ is a unit vector and C is equidistant from A and B. $\left(\underset{\sim}{b}-\underset{\sim}{a}\right)=\overrightarrow{AB}$ and hence a unit vector. In dot product, if given a unit vector, think about length of projection.

2 **(a)** Functions f and g are defined by

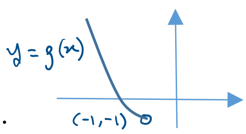
$$\mathrm{f}: x \mapsto 2-\mathrm{e}^{x+a}, \quad \text{for } x \in \mathbb{R},\ x>-2,$$
$$\mathrm{g}: x \mapsto x^2+2x, \quad \text{for } x \in \mathbb{R},\ x<-1,$$

where a is a constant.

(i) Find $\mathrm{g}^{-1}(x)$. [2]

Let $y=\mathrm{g}(x)=x^2+2x$ $x^2+2x-y=0$ $x=\dfrac{-2\pm\sqrt{4-4(-y)}}{2}=-1\pm\sqrt{1+y}=-1-\sqrt{1+y} \quad (\because x<-1)$ $\therefore \mathrm{g}^{-1}(x)=-1-\sqrt{1+x}$	

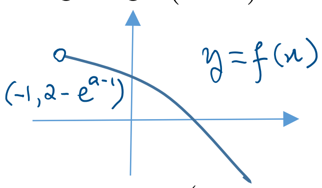
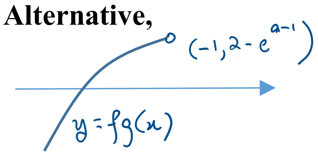
(ii) Explain why the composite function fg exists. [2]

Range of g $=(-1,\infty)$. Domain of f $=(-2,\infty)$. Since range of g $\subseteq$ domain of f, fg exists.  	Check if it should be [or (for the interval notation.

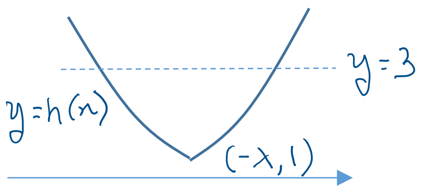
(iii) Find, in terms of a, an expression for $\mathrm{fg}(x)$ and write down the domain of fg. [2]

$\mathrm{fg}(x)=\mathrm{f}\left(x^2+2x\right)$ $=2-\mathrm{e}^{x^2+2x+a}$ $\mathrm{D_{fg}}=\mathrm{D_g}=(-\infty,-1)$.	

(iv) Find the range of fg, giving your answer in terms of e and a. [2]

Range of $\mathrm{g}=(-1,\infty)$.  Range of $\mathrm{fg}=\left(-\infty, 2-\mathrm{e}^{a-1}\right)$ **Alternative,**  Range of $\mathrm{fg}=\left(-\infty, 2-\mathrm{e}^{a-1}\right)$	State clearly which graph you are sketching. Draw the graph for the domain of the function only.

(b) The function h is defined by $\mathrm{h}: x \mapsto \mathrm{e}^{\lvert x+\lambda \rvert}, \quad x \in \mathbb{R}$, where λ is a constant.
Does h have an inverse? Justify your answer. [2]

$\mathrm{h}(-\lambda+1)=\mathrm{e}^{\lvert -\lambda+1+\lambda \rvert}=\mathrm{e}$, $\mathrm{h}(-\lambda-1)=\mathrm{e}^{\lvert -\lambda-1+\lambda \rvert}=\mathrm{e}$ Hence, h is not one-one and h does not have an inverse. **Alternative,**  The line $y=3$ cuts the graph of $y=\mathrm{h}(x)$ more than once. Hence, h is not one-one and h does not have an inverse.	Need to indicate the minimum value of y as it is a critical feature. Draw the line $y=k$ for any $k>1$.

3 **(i)** Show that $\cos\left(x+\frac{\pi}{6}\right)+\cos\left(x-\frac{\pi}{3}\right)=\sqrt{2}\cos\left(x-\frac{\pi}{12}\right)$. [2]

$\cos\left(x+\frac{\pi}{6}\right)+\cos\left(x-\frac{\pi}{3}\right)=2\cos\frac{1}{2}\left(x+\frac{\pi}{6}+x-\frac{\pi}{3}\right)\cos\frac{1}{2}\left(x+\frac{\pi}{6}-x+\frac{\pi}{3}\right)$ $=2\cos\left(x-\frac{\pi}{12}\right)\cos\left(\frac{\pi}{4}\right)$ $=2\left(\frac{1}{\sqrt{2}}\right)\cos\left(x-\frac{\pi}{12}\right)$ $=\sqrt{2}\cos\left(x-\frac{\pi}{12}\right)$	Refer to MF26 when need to manipulate trigo functions, and look for similar form. For sum of 2 trigo functions, factor formula could be helpful.

At time t seconds after turning on a switch, the total potential difference across two alternating current power supplies, V, is given by $\text{Re}(z_1+z_2)$, where

$$z_1=2\mathrm{e}^{\left(t+\frac{\pi}{6}\right)\mathrm{i}} \text{ and } z_2=2\mathrm{e}^{\left(t-\frac{\pi}{3}\right)\mathrm{i}}.$$

(ii) Express z_1+z_2 in the form $r\mathrm{e}^{(t-\alpha)\mathrm{i}}$, where $r>0$ and $-\pi<\alpha\le\pi$, leaving your values of r and α in exact form. [4]

z_1+z_2 $=2\left(\cos\left(t+\frac{\pi}{6}\right)+\mathrm{i}\sin\left(t+\frac{\pi}{6}\right)\right)+2\left(\cos\left(t-\frac{\pi}{3}\right)+\mathrm{i}\sin\left(t-\frac{\pi}{3}\right)\right)$ $=2\left(\cos\left(t+\frac{\pi}{6}\right)+\cos\left(t-\frac{\pi}{3}\right)\right)+2\mathrm{i}\left(\sin\left(t+\frac{\pi}{6}\right)+\sin\left(t-\frac{\pi}{3}\right)\right)$ $=2\sqrt{2}\cos\left(t-\frac{\pi}{12}\right)+2\mathrm{i}\left(2\sin\left(t-\frac{\pi}{12}\right)\cos\left(\frac{\pi}{4}\right)\right)$ $=2\sqrt{2}\cos\left(t-\frac{\pi}{12}\right)+\mathrm{i}\frac{4}{\sqrt{2}}\sin\left(t-\frac{\pi}{12}\right)$ $=2\sqrt{2}\left[\cos\left(t-\frac{\pi}{12}\right)+\mathrm{i}\sin\left(t-\frac{\pi}{12}\right)\right]$ $=2\sqrt{2}\mathrm{e}^{\left(t-\frac{\pi}{12}\right)\mathrm{i}}$	For addition of 2 complex numbers, we will consider either cartesian or trigo form. Since answer is required in exp form, we work with trigo form. Also, bear in mind what you have done in **part (i),** which involves trigo and the angles are the same.

(iii) From the time the switch is turned on, find the amount of time it takes for V to be first at half its maximum value, giving your answer in seconds, correct to 3 decimal places. [2]

$V=\mathrm{Re}(z_1+z_2)=2\sqrt{2}\cos\left(t-\frac{\pi}{12}\right)$ Maximum value of $V=2\sqrt{2}$ When V is half its maximum value, $\sqrt{2}=2\sqrt{2}\cos\left(t-\frac{\pi}{12}\right)$ $\cos\left(t-\frac{\pi}{12}\right)=\frac{1}{2}$ Hence, smallest positive value of t is such that $t-\frac{\pi}{12}=\frac{\pi}{3}\Rightarrow t=\frac{5\pi}{12}=1.309$ (to 3 d.p.). The time taken is 1.309 s.	Recall $-1\le\cos\theta\le 1$ $-A\le A\cos\theta\le A$ Max value of $A\cos\theta=A$ Min value of $A\cos\theta=-A$

(iv) The engineer modified the power supplies so that $z_1=z_2=w^{2n}e^{it}$, where $w=1+i$ and n is an integer. Show that $V=2^{n+1}\cos\left(\frac{n\pi}{2}+t\right)$. [3]

$z_1=z_2=w^{2n}e^{it}\Rightarrow z_1+z_2=2w^{2n}e^{it}$ $\lvert z_1+z_2\rvert=\lvert 2w^{2n}e^{it}\rvert$ $=2\lvert 1+i\rvert^{2n}\lvert e^{it}\rvert$ $=2\left(\sqrt{2}\right)^{2n}$ $=2^{n+1}$ $\arg(z_1+z_2)=\arg\left(2w^{2n}e^{it}\right)$ $\arg(2)+2n\arg(w)+\arg\left(e^{it}\right)=0+2n\arg(1+i)+\arg\left(e^{it}\right)$ $=2n\left(\frac{\pi}{4}\right)+t$ $=\frac{n\pi}{2}+t$ Hence, $V=\mathrm{Re}(z_1+z_2)=2^{n+1}\cos\left(\frac{n\pi}{2}+t\right)$. Alternatively, $z_1+z_2=2w^{2n}e^{it}$ $=2(1+i)^{2n}e^{it}$ $=2\left(\sqrt{2}e^{\frac{i\pi}{4}}\right)^{2n}e^{it}$ $=2\left(\sqrt{2}\right)^{2n}e^{\frac{in\pi}{2}}e^{it}$ $=2^{n+1}e^{i\left(\frac{n\pi}{2}+t\right)}$ $=2^{n+1}\left(\cos\left(\frac{n\pi}{2}+t\right)+i\sin\left(\frac{n\pi}{2}+t\right)\right)$ $V=\mathrm{Re}(z_1+z_2)=2^{n+1}\cos\left(\frac{n\pi}{2}+t\right)$	$2w^{2n}e^{it}$ is multiplication of 2 complex numbers. Consider changing everything to exp form

4 It is given that $\sum_{r=1}^{N} \ln\left[\frac{(r+1)(r+3)}{r(r+2)}\right] = \ln\left[\frac{(N+1)(N+3)}{2}\right]$.

Use this result to find $\sum_{r=4}^{k+2} \ln\left[\frac{r(r+2)}{3(r-1)(r+1)}\right]$, expressing your answer in the form $\ln\left[\frac{(k+2)(k+4)}{a\left(b^{k-1}\right)}\right]$ where a and b are positive integers to be determined. [5]

$\sum_{r=4}^{k+2} \ln\left[\frac{r(r+2)}{3(r-1)(r+1)}\right]$ $= \sum_{r=4}^{k+2}\left[\ln\left[\frac{r(r+2)}{(r-1)(r+1)}\right] - \ln 3\right]$ $= \sum_{r=3}^{k+1} \ln\left[\frac{(r+1)(r+3)}{r(r+2)}\right] - \sum_{r=4}^{k+2} \ln 3$ $= \sum_{r=1}^{k+1} \ln\left[\frac{(r+1)(r+3)}{r(r+2)}\right] - \sum_{r=1}^{2} \ln\left[\frac{(r+1)(r+3)}{r(r+2)}\right] - (k-1)\ln 3$ $= \ln\left[\frac{(k+2)(k+4)}{2}\right] - \ln\left[\frac{(3)(5)}{2}\right] - (k-1)\ln 3$ $= \ln\left[\frac{(k+2)(k+4)}{15\left(3^{k-1}\right)}\right]$	Note: $\ln\left(\frac{a}{b}\right) = \ln a - \ln b$ So $\ln\left(\frac{a}{3}\right) = \ln a - \ln 3$ Do revise the logarithm properties as they are assumed O Level knowledge.

A sequence of positive real numbers $v_1, v_2, v_3, \ldots$ is given by

$$v_1 = 5 \text{ and } v_{n+1} = v_n + \sum_{r=1}^{n}\left[(2r+1) + \ln\left[\frac{(r+1)(r+3)}{r(r+2)}\right]\right].$$

Show that $v_{n+1} - v_n = n(n+2) + \ln\left[\frac{(n+1)(n+3)}{2}\right]$. [3]

$v_{n+1} - v_n = \sum_{r=1}^{n}\left[(2r+1) + \ln\left[\frac{(r+1)(r+3)}{r(r+2)}\right]\right]$ $= \sum_{r=1}^{n}(2r+1) + \sum_{r=1}^{n} \ln\left[\frac{(r+1)(r+3)}{r(r+2)}\right]$ $= \frac{n}{2}(3+2n+1) + \ln\left[\frac{(n+1)(n+3)}{2}\right]$ $= n(n+2) + \ln\left[\frac{(n+1)(n+3)}{2}\right]$	A ‘show’ question, working needs to be rigorous. Write down the formula for sum of AP with substitution of values.

By considering $\sum_{n=1}^{99}\left(v_{n+1}-v_n\right)$, find the numerical value of v_{100}, correct your answer to the nearest integer. [4]

$\sum_{n=1}^{99}\left(v_{n+1}-v_n\right)=\sum_{n=1}^{99}\left(n\left(n+2\right)+\ln\left[\frac{\left(n+1\right)\left(n+3\right)}{2}\right]\right)$ $=338916.3055$ by GC $\sum_{n=1}^{99}\left(v_{n+1}-v_n\right)=v_2-v_1$ $+v_3-v_2$ $+v_4-v_3$ $+\ldots$ $+v_{99}-v_{98}$ $+v_{100}-v_{99}$ $=v_{100}-v_1$ $=v_{100}-5$ Hence, $v_{100}-5=338916.3055$ $v_{100}=338916.3055+5=338921$ (to nearest integer)	Read question carefully for all the information provided. Need to find v_{100} using $\sum_{n=1}^{99}\left(v_{n+1}-v_n\right)$ Given $v_1=5$ and showed $v_{n+1}-v_n=n\left(n+2\right)+\ln\left[\frac{\left(n+1\right)\left(n+3\right)}{2}\right]$ in previous part. Think of how to link all these together

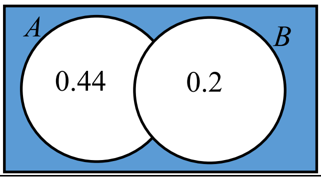
Section B: Probability and Statistics [60 marks]

5 For events A and B, it is given that $\mathrm{P}\left(A\right)=0.6$, $\mathrm{P}\left(B\right)=0.2$ and $\mathrm{P}\left(A \mid B'\right)=0.55$. Find

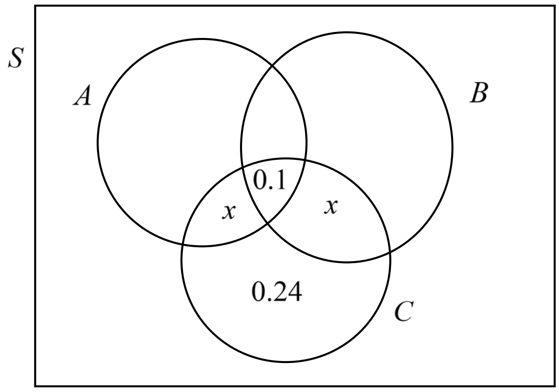
(i) $\mathrm{P}\left(A\cap B'\right)$, [1]

$\mathrm{P}\left(A\cap B'\right)=\mathrm{P}\left(A \mid B'\right)\times\mathrm{P}\left(B'\right)$ $=0.55\times\left(1-0.2\right)$ $=0.44$	

(ii) $\mathrm{P}\left(A'\cap B'\right)$. 2]

$\mathrm{P}\left(A'\cap B'\right)=1-\mathrm{P}\left(A\cup B\right)$ $=1-\left(\mathrm{P}\left(A\cap B'\right)+\mathrm{P}\left(B\right)\right)$ $=1-\left(0.44+0.2\right)$ $=0.36$  	Use a Venn diagram. Fill in the prob given and found in earlier part.

For a third event C, it is given that $\mathrm{P}(C)=0.4$, $\mathrm{P}(A\cap C)=\mathrm{P}(B\cap C)$, $\mathrm{P}(A'\cap B'\cap C)=0.24$ and $\mathrm{P}(A\cap B\cap C)=0.1$. Determine whether A and C are independent. [3]

 Since $\mathrm{P}(A\cap C)=\mathrm{P}(B\cap C)$, Let $\mathrm{P}(A\cap C\cap B')=\mathrm{P}(B\cap C\cap A')=x$ $\mathrm{P}(C)=0.1+0.24+2x=0.4 \quad \Rightarrow x=0.03$ $\mathrm{P}(A\cap C)=0.1+0.03=0.13\neq \mathrm{P}(A)\times\mathrm{P}(C)=0.6\times0.4=0.24$ $\therefore A$ and C are not independent.	Use a Venn diagram. Fill in the prob given and found in earlier part.

6 Four-figure numbers are to be formed from the digits 3, 4, 5, 6, 7 and 8. Find the number of different four-figure numbers that can be formed if

(i) no digit may appear more than once in the number, [1]

No. of numbers = ${}^6C_4\times4!={}^6P_4=360$	B1: 360

(ii) there is at least one repeated digit, but no digit appears more than twice in the number, [3]

Case 1: **AABC** no. of numbers = ${}^6C_1\times{}^5C_2\times\dfrac{4!}{2!}=720$ Case 2: **AABB** no. of numbers = ${}^6C_2\times\dfrac{4!}{2!2!}=90$ Total no. of numbers = 720+90 =810 Alternatively, Total – all different – all same – 3 same and 1 different $=1296-360-6-{}^6C_1\times{}^5C_1\times\dfrac{4!}{3!}=810$	‘at least’, ‘more than’ suggest that there a few cases to consider. List out the cases systematically. 1 repeated digit, 2 repeated digit … Be clear minded when considering complementary method. Always ask if you have considered all the cases to exclude.

(iii) no digit may appear more than once in the number and the sum of all the digits in the number is not divisible by six. [3]

No. of numbers = $360-3\times 4!=288$	Use complement with answer from (i) as there are fewer cases for number divisible by 6. To find number that are divisible by 6, be systematic so that you do not miss out cases. To get 4 digits from 3, 4, 5, 6, 7 and 8, Smallest possible sum = 3+4+5+6 = 18 (1 case only) Largest possible sum = 5+6+7+8 = 26 So the sum that is divisible by 6 can only be 18 or 24. To get 24, largest possible sum need to subtract 2, i.e 3,6,7,8 or 5,4,7,8 $\therefore$ 3 cases: 3,4,5,6 or 3,6,7,8 or 5,4,7,8. Each has 4! ways to arrange the digits.

7 An unbiased yellow cubical die has two faces labelled 10, two faces labelled 30 and two faces labelled 50. An unbiased green cubical die has four faces labelled 60, one face labelled 80 and one face labelled 100.

(i) When both dice are thrown, the random variable X is half of the difference between the score on the green die and the score on the yellow die. Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$. [4]

Green \ Yellow	10	30	50
60	$\frac{50}{2}$	$\frac{30}{2}$	$\frac{10}{2}$
80	$\frac{70}{2}$	$\frac{50}{2}$	$\frac{30}{2}$
100	$\frac{90}{2}$	$\frac{70}{2}$	$\frac{50}{2}$

x	5	15	25	35	45
$\mathrm{P}(X=x)$	$\frac{4}{6}\times\frac{2}{6}=\frac{2}{9}$	$\frac{5}{18}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{18}$

Always check that the probabilities sum to 1.

$$\mathrm{E}(X)=\frac{2}{9}(5)+\frac{5}{18}(15)+\frac{1}{3}(25)+\frac{1}{9}(35)+\frac{1}{18}(45)$$
$$=20$$
$$\mathrm{E}(X^2)=\frac{2}{9}(5)^2+\frac{5}{18}(15)^2+\frac{1}{3}(25)^2+\frac{1}{9}(35)^2+\frac{1}{18}(45)^2$$
$$=525$$
$$\mathrm{Var}(X)=\mathrm{E}(X^2)-\left[\mathrm{E}(X)\right]^2=525-20^2=125$$

<u>Alternative Solution</u>

Let Y and G be the score on the yellow die and green die respectively.

y	10	30	50
$\mathrm{P}(Y=y)$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{2}{6}$

g	60	80	100
$\mathrm{P}(G=g)$	$\frac{4}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\mathrm{E}(G)=\frac{2}{3}(60)+\frac{1}{6}(80)+\frac{1}{6}(100)=70$$

$$\mathrm{E}\left(G^2\right)=\frac{2}{3}(60)^2+\frac{1}{6}(80)^2+\frac{1}{6}(100)^2=\frac{15400}{3}$$

$$\mathrm{Var}(G)=\mathrm{E}\left(G^2\right)-\left[\mathrm{E}(G)\right]^2=\frac{15400}{3}-70^2=\frac{700}{3}$$

$$\mathrm{E}(Y)=30 \quad (\text{by symmetry})$$

$$\mathrm{E}\left(Y^2\right)=\frac{2}{6}(10)^2+\frac{2}{6}(30)^2+\frac{2}{6}(50)^2=\frac{3500}{3}$$

$$\mathrm{Var}(Y)=\mathrm{E}\left(Y^2\right)-\left[\mathrm{E}(Y)\right]^2=\frac{3500}{6}-30^2=\frac{800}{3}$$

$$X=\frac{1}{2}\left|G-Y\right|=\frac{1}{2}(G-Y) \qquad \because G-Y \text{ is always positive}$$

$$\mathrm{E}(X)=\mathrm{E}\left(\frac{G-Y}{2}\right)=\frac{1}{2}\left[\mathrm{E}(G)-\mathrm{E}(Y)\right]=\frac{1}{2}\left[70-30\right]=20$$

$$\mathrm{Var}(X)=\left(\frac{1}{2}\right)^2\left[\mathrm{Var}(G)+\mathrm{Var}(Y)\right]$$

$$=\left(\frac{1}{2}\right)^2\left[\frac{700}{3}+\frac{800}{3}\right]=125$$

(ii) Suppose now the yellow die is replaced by an unbiased blue cubical die with two faces labelled 15, two faces labelled 35 and two faces labelled 55. The random variable W is half of the difference between the score on the green die and the score on the blue die. Without doing further calculation, comment on $\mathrm{E}(W)$ and $\mathrm{Var}(W)$ in relation to $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ respectively . [2]

Observe that the number on blue die is obtain by adding 5 to every number on the yellow die. Hence, $W=X-2.5$.

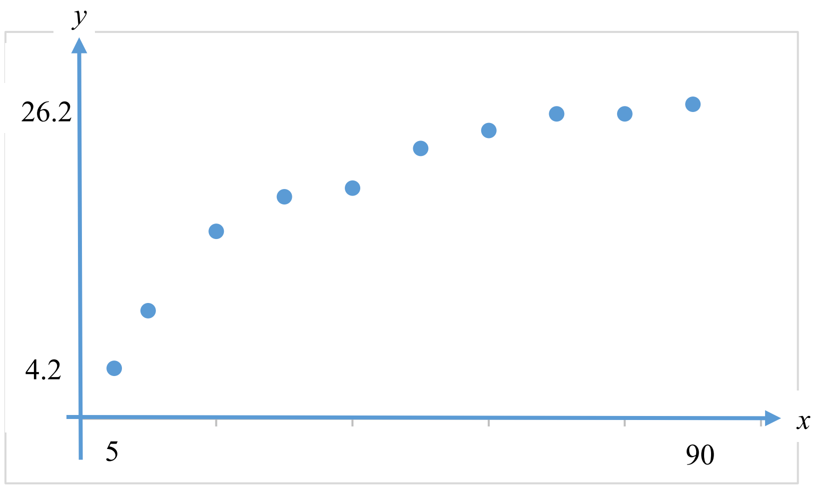
Therefore $\mathrm{E}(W)$ is 2.5 less than $\mathrm{E}(X)$.

Adding 5 to all the numbers on the yellow die does not affect the spread of the data, hence $\mathrm{Var}(W)$ is the same as $\mathrm{Var}(X)$

8 An agricultural experiment was carried out to study the effect of a certain fertilizer on the growth of seedlings. The fertilizer is applied at various concentrations to a random sample of ten plots of land. Seeds are sown and two weeks later, the mean height of the seedlings on each plot of land is measured. The results are shown in the table.

Concentration of fertilizer (x grams/m^2)	5	10	20	30	40	50	60	70	80	90
Mean height of seedlings (y cm)	4.2	9.0	15.6	18.5	19.2	22.5	24.0	25.4	25.4	26.2

(i) Draw the scatter diagram for these values, labelling the axes clearly. [1]

	Label minimum and maximum x and y values. Check that the 8th and 9th points are level as they have the same y-value.

It is thought that the mean height of seedlings y can be modelled by one of the formulae

$$y = a + bx \quad \text{or} \quad y = c + d\ln x,$$

where a, b, c and d are constants.

(ii) Find, correct to 4 decimal places, the value of the product moment correlation coefficient between

(a) x and y,

(b) $\ln x$ and y. [2]

(a) $r = 0.925588 \approx 0.9256$ **(b)** $r = 0.996583 \approx 0.9966$	

(iii) Use your answers to parts **(i)** and **(ii)** to explain which of $y = a + bx$ or $y = c + d\ln x$ is the better model. [1]

From the scatter diagram, it is observed that as x increases, y increases by decreasing amounts, and the product moment correlation coefficient between $\ln x$ and y is closer to 1 than that of x and y. Hence $y = c + d\ln x$ is the better model.	Use both scatter diagram and correlation coefficient to compare.

It is required to estimate the value of x for which $y = 20.0$.

(iv) Explain why neither the regression line of x on y nor the regression line of $\ln x$ on y should be used. [1]

Since the fertilizer is applied at various concentrations, x (and hence $\ln x$) is the predetermined/controlled/independent variable. Thus, neither the regression line of x on y nor the regression line of $\ln x$ on y should be used. To estimate x given $y = 20.0$, the regression line of y on $\ln x$ should be used.	

(v) Find the equation of a suitable regression line and use it to find the required estimate. [2]

From GC, the equation of the regression line of y on $\ln x$ is $y = -8.4492 + 7.8126\ln x$ $\therefore y = -8.45 + 7.81\ln x$ When $y = 20.0$, $x = 38.147 \approx 38$	

(vi) Re-write your equation from part **(v)** so that it can be used when y, the mean height of seedlings, is given in mm. [1]

Replace y by $0.1y$, $0.1y = -8.4492 + 7.8126\ln x$ $\therefore y = -84.5 + 78.1\ln x$	1 cm = 10 mm

9 In each batch of manufactured articles, 5% of the articles are found to be defective. A quality inspection is carried out by checking samples of 20 articles.

(i) If 2 or fewer defective articles are found in the sample of 20, the batch is accepted. Find the probability that the batch is accepted. [1]

Let Y be the number of defective articles out of 20. $Y \sim \mathrm{B}(20, 0.05)$ Required probability = $\mathrm{P}(Y \le 2) \approx 0.925$	

(ii) Find the least value of n such that the probability of having less than n defective articles in a sample of 20 articles is greater than 0.99. [2]

$\mathrm{P}(Y < n) > 0.99$ $\mathrm{P}(Y \le n-1) > 0.99$ From GC, $\mathrm{P}(Y \le 3) = 0.98410 < 0.99$ $\mathrm{P}(Y \le 4) = 0.99743 > 0.99$ $\therefore$ least value of $n = 5$	For your working, check that $\mathrm{P}(Y \le 3) < 0.99$ and $\mathrm{P}(Y \le 4) > 0.99$ to check that least value of $n-1$ is 4 (or draw a table to check).

(iii) Fifty random samples of 20 articles each were taken. Find the probability that the average number of defective articles per sample is at most 1.25. [3]

Let $\bar{Y} = \dfrac{Y_1 + Y_2 + \ldots + Y_{50}}{50}$ $\mathrm{E}(\bar{Y}) = \mathrm{E}(Y) = 20 \times 0.05 = 1$ $\mathrm{Var}(\bar{Y}) = \dfrac{\mathrm{Var}(Y)}{50} = \dfrac{20 \times 0.05 \times 0.95}{50} = \dfrac{0.95}{50}$ By CLT, $\bar{Y} \sim \mathrm{N}\left(1, \dfrac{0.95}{50}\right)$ approx $\mathrm{P}(\bar{Y} \le 1.25) = 0.96514 \approx 0.965$	

(iv) It is proposed that a smaller sample be taken for inspection. Find the largest value of k such that the probability of having at least 1 defective article in a sample of k articles is to be less than 0.4? [3]

Let W be the number of defective articles out of k. $W \sim \mathrm{B}(k, 0.05)$ $\mathrm{P}(W \ge 1) < 0.4$ $1 - \mathrm{P}(W = 0) < 0.4$ $\mathrm{P}(W = 0) > 0.6$ $(0.95)^k > 0.6$ $k < \dfrac{\ln 0.6}{\ln 0.95} = 9.9589$ $\therefore$ largest value of $k = 9$	

10 In this question, you should state clearly the values of the parameters of any normal distribution you use.

A meat supplier imports frozen chickens and frozen ducks which are priced by weight. The masses, in kg, of frozen chickens and frozen ducks are modelled as having normal distributions with means and standard deviations as shown in the table.

	Mean mass	Standard deviation
Frozen chickens	1.5	0.3
Frozen ducks	2.6	0.5

(i) Find the probability that a randomly chosen frozen duck has a mass which is more than twice that of a randomly chosen frozen chicken. [3]

Let X kg and Y kg be the mass of a frozen chicken and a frozen duck respectively. $X \sim \mathrm{N}(1.5, 0.3^2)$ and $Y \sim \mathrm{N}(2.6, 0.5^2)$ $Y - 2X \sim \mathrm{N}(-0.4, 0.61)$ $\mathrm{P}(Y > 2X) = \mathrm{P}(Y - 2X > 0)$ $= 0.30427$ ≈ 0.304	

The frozen chickens are imported at a cost price of \$6 per kg and the frozen ducks at \$8 per kg.

(ii) The meat supplier orders 100 frozen chickens and 50 frozen ducks in a particular consignment. Find the probability that the meat supplier paid no more than \$2000 for this consignment. State an assumption needed for your calculation. [5]

Let $T=6\left(X_1+X_2+...+X_{100}\right)+8\left(Y_1+Y_2+...+Y_{50}\right)$ $\mathrm{E}(T)=6(100)(1.5)+8(50)(2.6)=1940$ $\mathrm{Var}(T)=6^2(100)(0.3^2)+8^2(50)(0.5^2)=1124$ $\therefore T\sim \mathrm{N}(1940,1124)$ $\mathrm{P}(T\le 2000)=0.96324$ ≈ 0.963 The masses of **all** frozen chickens and frozen ducks are independent of one another.	Assumption: For the calculations of $\mathrm{E}(T)$ and $\mathrm{Var}(T)$ to be valid, the random variables $X_1,...,X_{100},Y_1,...,Y_{50}$ must be independent of one another. Then write this in context.

(iii) A restaurant owner buys frozen ducks from this supplier, who sells the frozen ducks at a profit of 25%. The restaurant owner does not wish to purchase frozen ducks that are too big or too small. If there is a probability of 0.7 that the restaurant owner paid between \$20.00 and \$$a$ for a randomly chosen frozen duck, find the value of a. [4]

Let \$$W$ be the selling price of a randomly chosen frozen duck. $W=1.25(8)Y=10Y$ $W\sim \mathrm{N}\left(10\times 2.6,10^2\times 0.5^2\right)$ i.e. $W\sim \mathrm{N}(26,25)$ $\mathrm{P}(20<W<a)=0.7$ $\mathrm{P}(W<a)-\mathrm{P}(W<20)=0.7$ $\mathrm{P}(W<a)=0.7+\mathrm{P}(W<20)$ $=0.81507$ $\therefore a=30.48$ **Alternatively,** Let \$$W$ be the selling price of a randomly chosen frozen duck. $W=1.25(8)Y=10Y$ $\mathrm{P}(20<10Y<a)=0.7$ $\mathrm{P}\left(2<Y<\frac{a}{10}\right)=0.7$ $\mathrm{P}\left(Y<\frac{a}{10}\right)-\mathrm{P}(Y<2)=0.7$ $\mathrm{P}\left(Y<\frac{a}{10}\right)=0.7+\mathrm{P}(Y<2)=0.81507$	Leave final answer for money to 2 decimal places.

$\therefore \frac{a}{10} = 3.04837$ $\Rightarrow a = 30.48$	

11 **(a)** The average time taken by George to swim 100 m freestyle is 120.05 seconds. He bought a new pair of special goggles from a salesperson who claimed that the goggles will help George swim faster. After wearing the goggles, the time, t seconds, for George to swim 100 m freestyle of each of 50 randomly chosen timings is recorded. The results are summarised as follows.

$$\sum(t-120.05) = -66.4, \quad \sum(t-120.05)^2 = 1831.945.$$

(i) Test, at the 10% level of significance, whether the goggles helped George to swim faster. You should state your hypotheses and define any symbols you use. [6]

Let T seconds be the time taken for George to swim the 100m freestyle and let μ be the population mean of T. $H_0: \mu = 120.05$ $H_1: \mu < 120.05$ Level of significance: 10% Test Statistic: Since n = 50 is sufficiently large, by Central Limit Theorem, $\overline{T}$ is approximately normally distributed. When H_0 is true, $Z = \frac{\overline{T}-120.05}{S/\sqrt{n}} \sim N(0,1)$ approximately Computation: $n = 50$, $\overline{t} = \frac{-66.4}{50} + 120.05 = 118.722$ $s^2 = \frac{1}{49}\left(1831.945 - \frac{(-66.4)^2}{50}\right) \approx 35.5871$ $p-\text{value} \approx 0.05773 = 0.0577$ (3s.f) Conclusion: Since $p-\text{value} = 0.0577 < 0.1$, H_0 is rejected at the 10% level of significance. So, there is sufficient evidence to conclude that the mean time taken to swim 100 m freestyle is less than 120.05 s.	If the variable used is t, define T as the random variable, and define μ as the population mean of T. Test statistic should have the value of $\mu = 120.05$ substituted – it is when H_0 is true. For final sentence: Remember to conclude that there is sufficient/insufficient evidence to conclude 'H_1' (**in context of the question**).

(ii) Explain why this test would be inappropriate if George had taken a random sample of 10 of his 100 m freestyle timings. [1]

The time taken for George to swim the 100 m freestyle is not known to be normally distributed. If a sample of 10 of his 100 m freestyle timings is taken, <u>Central Limit Theorem cannot be applied</u> to approximate <u>sample mean time taken,</u> $\overline{T}$, for George to swim the 100 m freestyle to a normal distribution. Hence the test would not be appropriate.	

(iii) Explain, in the context of the question, the meaning of a "10% significance level". [1]

A 10% significance level means that there is a probability of 0.1 that the test concludes that the mean time taken for George to swim 100 m freestyle is less than 120.05 seconds, when it is actually 120.05 seconds.	… that the test concludes 'H_1', when 'H_0 is actually true' (**in context**).

(b) The random variable X has distribution $\mathrm{N}\left(\mu,\ \sigma^2\right)$.

A random sample of n observations of X is taken, where n is sufficiently large. The mean and variance of this sample is k and 9 respectively.

(i) A test at the 1% level of significance level indicates that the null hypothesis $\mu = 25$ is rejected in favour of the alternative hypothesis $\mu \neq 25$. Find, in terms of n, the range of values of k, giving non exact answers correct to 4 decimal places. [3]

$\mathrm{H}_0:\ \mu = 25$ $\mathrm{H}_1:\ \mu \neq 25$ Level of significance: 1% Test Statistic: when H_0 is true $Z = \dfrac{\overline{X} - 25}{S/\sqrt{n}} \sim \mathrm{N}(0,1)$ approximately Rejection region: $z \le -2.57583$ or $z \ge 2.57583$ Computation: $\overline{x} = k$, $s^2 = \dfrac{n}{n-1} \times 9$ $z-\text{calculated} = \dfrac{k-25}{\frac{s}{\sqrt{n}}} = \dfrac{k-25}{\frac{\frac{3\sqrt{n}}{\sqrt{n-1}}}{\sqrt{n}}} = \dfrac{k-25}{\frac{3}{\sqrt{n-1}}}$ Conclusion: H_0 is rejected at 1% significance level $\Rightarrow \dfrac{k-25}{\frac{3}{\sqrt{n-1}}} \le -2.57583$ or $\dfrac{k-25}{\frac{3}{\sqrt{n-1}}} \ge 2.57583$ $\Rightarrow k \le 25 - \dfrac{7.7275}{\sqrt{n-1}}$ or $k \ge 25 + \dfrac{7.7275}{\sqrt{n-1}}$	

(ii) Hence state the conclusion of the hypothesis test in the case where $k = 24$ and $n = 42$. [1]

When $n = 42$, $\mathrm{H}_0:\ \mu = 25$ is rejected in favour of $\mathrm{H}_1:\ \mu \neq 25$ when $k \le 23.793$ or $k \ge 26.207$. Since $k = 24$ does not satisfy the inequalities, we do not reject H_0 at 1% level of significant and conclude that there is insufficient evidence to suggest that $\mu \neq 25$.	